RADON IMPLICIT FIELD TRANSFORM: FULLY RADAR-NATIVE NOVEL-VIEW SYNTHESIS AND 3D RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

We introduce the Radon Implicit Field Transform (RIFT), a fully radar-native framework for complex-valued novel-view synthesis and 3D reconstruction that combines an adaptive point-scatterer representation with a radar forward model. Both initialization and optimization use only complex radar measurements and calibrated sensor poses, without visual or LiDAR-derived priors, initialization, or supervision. RIFT represents the scene with point scatterers at adaptive spatial resolutions, with learnable positions and direction-dependent complex reflectance parameterized by spherical harmonics. We incorporate the generalized Radon transform as a forward model and optimize point-scatterer positions and reflectance by matching predicted and measured complex radar signals. During optimization, signal-loss gradients guide local increases in spatial resolution and spherical-harmonic expansion degree; these capacity updates leave the current signal predictions unchanged, except for the coarse-to-fine densification we use on lower-resolution measured radar data. We evaluate complex-valued novel-view synthesis by comparing forward-model predictions at unseen sensor poses with held-out complex measurements, and evaluate 3D reconstruction by comparing geometry extracted from the learned point scatterers with reference scene geometry. On our simulated dataset, RIFT achieves a mean held-out complex relative mean-squared error of 0.3909% and a mean symmetric squared Chamfer distance of $4.813 \times 10^{-5}\,\mathrm{m}^2$; on a real-world dataset, it reduces the symmetric squared Chamfer distance by up to 95.12% relative to the strongest baseline, reaching $5.536 \times 10^{-2}\,\mathrm{m}^2$.

1 INTRODUCTION

Research in Computer Vision has established novel-view synthesis (NVS) and 3D reconstruction from posed optical observations, from neural radiance fields to explicit voxel and Gaussian scene representations Mildenhall et al. (2020); Fridovich-Keil et al. (2022); Kerbl et al. (2023). These approaches, however, inherit the constraints of optical sensing: cameras depend on illumination, and both cameras and LiDAR rely on optical visibility. Radar instead actively probes a scene with radio-frequency (RF) waves and can remain informative in conditions where optical sensing is degraded, motivating radar-native reconstruction for settings in which optical sensing is unreliable Borts et al. (2024); Takawale & Roy (2025a;b); Kung et al. (2025).

Radar scene learning raises two complementary questions: can a learned field predict the complex frequency response at an unseen sensor pose (novel-view synthesis), and does its spatial support agree with the scene geometry (3D reconstruction)? These are the forward and inverse uses of the same representation. Accurate signal prediction does not by itself guarantee accurate geometry, nor does geometric accuracy guarantee faithful complex responses. Existing methods address only one side of this pair: SpINR and SpINRv2 reconstruct from complex measurements without evaluating complex-valued novel views Takawale & Roy (2025a;b), while Radar Fields, GeRaF, RadarSplat, and synthetic-aperture radar (SAR) renderers synthesize power, magnitude, or formed images rather than the underlying complex frequency response Borts et al. (2024); Lu et al. (2025); Kung et al. (2025); Li et al. (2025); Zhang et al. (2026b). Related complex-valued sonar reconstruction provides a cross-modal precedent Vengurlekar et al. (2026), but, to our knowledge, RIFT is the first radar-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

native learned scene representation evaluated for both complex-valued held-out-view prediction and 3D reconstruction from the same field.

This capability also has a radar signal-processing motivation: predicting complex frequency responses at unmeasured viewpoints offers a learned alternative to repeated physics-based simulation. Such responses are typically generated by full-wave or ray-tracing simulators, whose repeated forward solves are computationally expensive and require supplied geometry, materials, antennas, and propagation assumptions Ansys, Inc. (2023; 2026); Hoydis et al. (2023). A scene-specific learned field can amortize novel-view queries after fitting measured responses.

We introduce the *Radon Implicit Field Transform* (RIFT), a radar-native scene representation trained exclusively from complex radar measurements, sampled frequencies, and calibrated sensor poses, without visual- or LiDAR-derived priors, initialization, supervision, or optimization inputs. RIFT represents the scene using complex spherical-harmonic coefficients on point scatterers anchored to the cells of a lattice whose resolution is adaptively refined. A radar forward model based on the generalized Radon transform (GRT) Nolan & Cheney (2002) maps this learned representation to held-out complex responses, while the same coefficients define a radar-visible scattering-energy volume for 3D reconstruction. RIFT accelerates the forward model with a nonuniform fast Fourier transform (NUFFT) Dutt & Rokhlin (1993); Greengard & Lee (2004). Thus a single radar-native field supports both outputs, with signal and geometry fidelity evaluated separately.

Our contributions are:

- We formulate and evaluate, to our knowledge, the first radar-native learned field for complex-valued held-out-view synthesis and 3D reconstruction without geometric supervision or visual- or LiDAR-derived inputs.
- A complex scattering representation with adaptive spatial resolution that refines spatial support and increases the spherical-harmonic expansion degree while preserving the pre-refinement prediction at each capacity update; for lower-resolution measured SAR, a coarse-to-fine densification replaces the spatial refinement.
- We validate RIFT on both a controlled synthetic radar dataset and the publicly available GOTCHA Volumetric SAR dataset Casteel et al. (2007), with separate signal and geometry evaluations and matched observable-specific baselines.

2 Related Work

NVS and 3D Reconstruction in Vision. Neural radiance fields, explicit voxel representations, and 3D Gaussian Splatting learn scene representations that support both novel-view synthesis and geometric reconstruction Mildenhall et al. (2020); Fridovich-Keil et al. (2022); Kerbl et al. (2023). Related work has extended neural rendering to other sensing modalities; for example, LiDAR-NeRF and Neural LiDAR Fields model posed range measurements Tao et al. (2024); Huang et al. (2023). RIFT shares the goal of recovering a pose-queryable scene representation from which both measurements and geometry can be obtained, but operates on complex radar measurements.

Radar Scene Representations and Novel-View Synthesis. Recent work has adapted neural scene representations to radar measurements. Radar Fields learns geometry and reflectance from raw frequency-space frequency-modulated continuous-wave (FMCW) measurements and synthesizes radar observations at novel viewpoints Borts et al. (2024). DART models range–Doppler radar observations using a physics-informed neural field Huang et al. (2024), while RadarSplat represents automotive scenes with Gaussian primitives for radar-image synthesis and occupancy reconstruction Kung et al. (2025). GeRaF learns signed-distance, reflectance, and signal-power fields for RF geometry reconstruction and matched-filter image synthesis Lu et al. (2025), and RF4D extends radar neural fields to dynamic outdoor scenes through temporal modeling and scene flow Zhang et al. (2026a). Related SAR and inverse SAR (ISAR) methods synthesize formed radar images rather than the underlying complex frequency response Ehret et al. (2024); Lei et al. (2024); Liu et al. (2023); Li et al. (2025); Zhang et al. (2026b); Liu et al. (2026). NeuRadar instead generates processed automotive radar point clouds, jointly with camera and LiDAR outputs Rafidashti et al. (2025), while mmWEAVER conditions mmWave signal generation on visual and textual inputs Monjur & Nirjon

(2026). These methods establish radar novel-view synthesis and neural scene representation, but differ from RIFT in either the rendered observable, the input modalities, or both.

Reconstruction from Complex Radar Measurements. SpINR and SpINRv2 reconstruct volumes directly from complex FMCW frequency-domain measurements but do not evaluate complex-valued novel views Takawale & Roy (2025a;b); Sugavanam–Ertin reconstruct geometry from complex radar measurements Sugavanam & Ertin (2026), and matched filtering and backprojection are the classical non-learned counterparts Ulander et al. (2003). Other RF methods rely on different representations or learned priors: RFconstruct fuses radar point clouds before shape reconstruction Hussein et al. (2025), Wave-Former completes occluded geometry with a learned shape prior Dodds et al. (2026), RISE combines multipath processing with a simulation-trained prior Zhou et al. (2026), and Rascene reconstructs scenes from mmWave communication signals Song et al. (2026). In synthetic-aperture sonar, Reed et al. learn neural volumes from complex measurements Reed et al. (2023), and SH-SAS predicts complex spherical-harmonic coefficients with a hash-encoded neural field Vengurlekar et al. (2026); RIFT instead stores such coefficients on explicit point scatterers with adaptive spatial resolution.

3 METHODS

RIFT models a static scene as a field represented by point scatterers with direction-dependent complex responses. The same field is used in two ways (Fig. 1): a propagation operator renders the native complex radar response at a requested viewpoint, while its spatial scattering energy provides a 3D reconstruction.

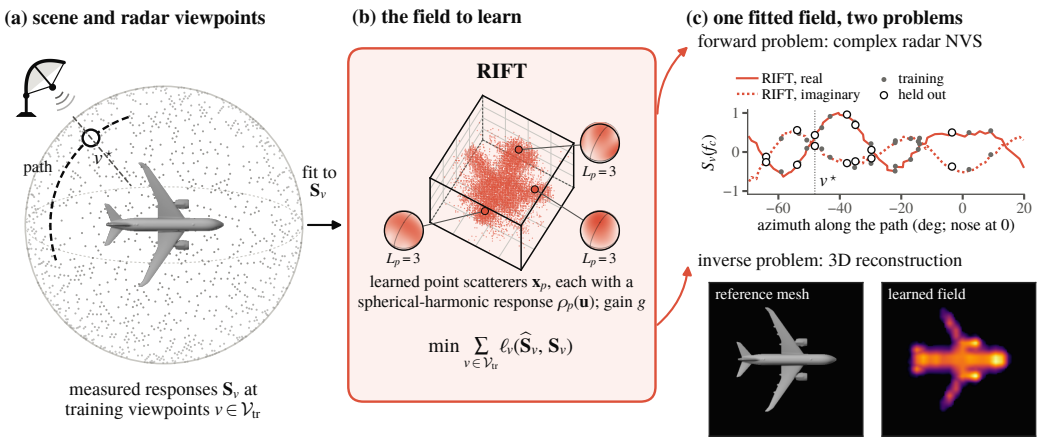

Figure 1: RIFT solves a forward and an inverse radar problem with one fitted field. (a) A scene (the B787 of our simulated dataset) is observed by a radar from viewpoints on a sphere (dots: training viewpoints $\mathcal{V}_{\text{tr}}$; not to scale); the dashed arc is a path of viewpoints through the held-out viewpoint $v^\star$. (b) RIFT fits point scatterers at positions $\mathbf{x}_p$, each with a spherical-harmonic response $\rho_p(\mathbf{u})$ of adaptive degree L_p, and a global gain g, so that the rendered responses $\widehat{\mathbf{S}}_v$ (Eq. 1) match the measured $\mathbf{S}_v$ (data term of Eq. 3); insets show $|\rho_p(\mathbf{u})|$ for three point scatterers. (c) The same field answers both problems. Forward problem, complex radar NVS: the real (solid) and imaginary (dotted) parts of the 10 GHz response along the path, rendered by RIFT and compared with measured training (dots) and validation (circles) viewpoints; $v^\star$ is a validation viewpoint with median complex relative mean-squared error (RelMSE). Inverse problem, 3D reconstruction: the reference mesh and the maximum-intensity projection of the learned field.

3.1 NEAR-FIELD-CAPABLE COMPLEX FORWARD MODEL

We use the standard scalar GRT phase model under the Born, or single-scattering, approximation Nolan & Cheney (2002). Let v and i index viewpoint and sampled frequency, and let p index an active point at $\mathbf{x}_p$. The quantity R_{vp} is the exact propagation-path length from the calibrated

transmitter, through point p, to the calibrated receiver (one bistatic pair on the RIFT dataset and one monostatic antenna on GOTCHA). RIFT predicts the complex sample, up to a known unit-modulus acquisition-reference factor (Appendix A), as

$$\widehat{S}_{v,i} = g \sum_{p \in \mathcal{A}} \frac{\rho_p(\mathbf{u}_v)}{(4\pi)^2 (R_{vp}^2 + \epsilon_R)} \exp(-\mathrm{i} k_i R_{vp}) \,, \tag{1}$$

Here $\rho_p(\mathbf{u}_v) \in \mathbb{C}$ is the learned point response at the viewpoint-level array-center aspect direction $\mathbf{u}_v$, $\mathcal{A}$ is the active point set, k_i is the wavenumber at f_i, $g \in \mathbb{C}$ is one global complex calibration factor, $\epsilon_R > 0$ stabilizes the denominator, and $\mathrm{i}^2 = -1$. Because the sampled frequencies are uniformly spaced, the phase factors into a carrier at a reference frequency and a term linear in the frequency index. Each point therefore acts as a complex source, with weight $g\rho_p(\mathbf{u}_v)e^{-\mathrm{i}k_{\mathrm{ref}}R_{vp}}/[(4\pi)^2(R_{vp}^2 + \epsilon_R)]$, located at its path length R_{vp} on a one-dimensional range axis. The response at all N_f frequencies is the Fourier transform of these nonuniformly placed sources, which a one-dimensional NUFFT evaluates in $O(J|\mathcal{A}| + M \log M)$ instead of $O(|\mathcal{A}|N_f)$ operations for J gridding taps and an oversampled grid of length M (Appendix B).

3.2 Adaptive Direction-Dependent Point Field

For each active point, RIFT stores an anchor $\mathbf{a}_p$ and a half-width h_p (the center and half-size of its lattice cell), a learned raw offset $\boldsymbol{\delta}_p$, and complex spherical-harmonic coefficients $c_{p\ell m}$. Its bounded location and aspect-dependent response are

$$\mathbf{x}_p = \mathrm{clip}_\Omega\{\mathbf{a}_p + h_p \tanh(\boldsymbol{\delta}_p)\} \,, \qquad \rho_p(\mathbf{u}) = \sum_{\ell=0}^{L_p} \sum_{m=-\ell}^{\ell} c_{p\ell m} Y_{\ell m}(\mathbf{u}). \tag{2}$$

Here Ω is the fixed scene support, with dataset-specific bounds given in Appendix A, and $Y_{\ell m}$ are real, orthonormal spherical harmonics. The degree-zero term represents isotropic scattering, while higher degrees capture aspect-dependent returns. Direction dependence also represents occlusion implicitly: a point scatterer that is hidden from some directions can learn a small response there, within the angular resolution of its spherical-harmonic degree, without explicit ray casting. Initial points begin at $L_p = 0$ and may unlock complete bands up to a prescribed maximum degree.

At viewpoint v, we evaluate Eq. 2 at the aspect direction $\mathbf{u}_v$ toward the calibrated array center (Appendix C); positions and coefficients feed the range operator directly, with no voxel rasterization.

3.3 Radar-Only Analysis-by-Synthesis Optimization

For the training viewpoints $\mathcal{V}_{\mathrm{tr}}$, we jointly optimize point offsets, spherical-harmonic coefficients, and one global complex gain g, shared across all viewpoints and frequencies, using complex residual mean-squared error and the regularized objective:

$$\mathcal{L}_{\mathrm{data}} = \frac{1}{|\mathcal{V}_{\mathrm{tr}}|} \sum_{v \in \mathcal{V}_{\mathrm{tr}}} \ell_v\Big(\widehat{\mathbf{S}}_v, \mathbf{S}_v\Big) \,, \qquad \mathcal{L} = \mathcal{L}_{\mathrm{data}} + \lambda_{\mathrm{sp}}\mathcal{R}_{\mathrm{sp}} + \lambda_{\mathrm{SH}}\mathcal{R}_{\mathrm{SH}}. \tag{3}$$

Here ℓ_v is the mean squared complex residual in the declared measurement domain, $\mathcal{R}_{\mathrm{sp}}$ is a point-group sparsity penalty, and $\mathcal{R}_{\mathrm{SH}}$ is an angular smoothness penalty; neither imposes a surface, occupancy, signed-distance, or mesh prior (Appendix D). Only radar measurements, sampled frequencies, and calibrated sensor poses supervise RIFT; no geometric target is used. We initialize a regular lattice of cell-center anchors over Ω, with zero offsets and degree zero, then initialize the isotropic coefficients with scaled matched-filter backprojection Ulander et al. (2003)—the adjoint of the fitting operator applied to a training subset—and then optimize by backpropagation through the renderer. We use AdamW to optimize the field and gain; Appendix D specifies the acquisition adaptations and training settings.

3.4 Training Loop and Adaptive Refinement

RIFT alternates gradient steps with occasional capacity updates (Algorithm 1). Between updates, AdamW moves each point within its lattice cell through its offset $\boldsymbol{\delta}_p$ and fits its coefficients; the point

Algorithm 1 RIFT training.

Place anchors $\mathbf{a}_p$ on a regular lattice over Ω; set $\boldsymbol{\delta}_p \leftarrow \mathbf{0}$ and $L_p \leftarrow 0$
Initialize c_{p00} by least-squares-scaled backprojection of 100 training views; warm-start g
for epoch $= 1, \ldots, E$ **do**
  for each training view v (a (pass, sector) unit on GOTCHA) **do**
    Render $\widehat{\mathbf{S}}_v$ (Eq. 1); take an AdamW step on $\mathcal{L}$ (Eq. 3) in $\{\boldsymbol{\delta}_p, c_{p\ell m}\}$ and g
    Accumulate spatial and angular evidence scores from $\nabla \mathcal{L}_{\text{data}}$
  end for
  if epoch is a refinement epoch **then** ▷ Appendix D, RIFT-dataset adaptive controller
    Unlock the next SH band of the top angular-score points (zero coefficients)
    Split the top spatial-score points into eight octants (RIFT dataset)
    Reset the evidence scores
  end if
  if epoch is a densification epoch (GOTCHA) **then**
    Replace each retained point by eight half-pitch children; lower the learning rate
  end if
  Keep the checkpoint with the lowest validation loss
end for
return point field $\{\mathbf{x}_p, \rho_p\}$ and gain g

set changes only at refinement events every ten epochs. During training, each point accumulates two evidence scores from the data-loss gradients. Its spatial score is the average magnitude of the gradient with respect to its position, scaled by its cell size; we use this sensitivity as a heuristic for allocating additional spatial capacity. Its angular score is the root-mean-square gradient of its next, still-locked spherical-harmonic band, probed every 16th view. At an event, the points with the highest angular scores (the top $1/16$) unlock their next band with zero coefficients, and the points with the highest spatial scores (the top $1/512$) are split: each keeps its position and coefficients and gains seven zero-coefficient siblings in the other octants of its cell, at half the cell size (each cell is split at most once). All new coefficients are zero, so an event leaves the current prediction unchanged and only adds capacity for the following gradient steps. The final scatterer positions are thus the lattice, refined where the data demand it, plus the learned offsets.

GOTCHA's 640 MHz bandwidth and narrow elevation span resolve the scene far more coarsely, so there the split is replaced by coarse-to-fine densification after epochs 4, 10, and 16: points with negligible energy are removed, and each remaining point is replaced by eight half-pitch children whose coefficients are interpolated trilinearly from the current field. Because this changes the prediction slightly, the learning rate is then reduced (Appendix D; Fig. 4).

3.5 Complex NVS and 3D Reconstruction

At a held-out calibrated viewpoint $v^\star$, RIFT performs no further fitting: it evaluates Eq. 2 at the new aspect direction, recomputes the exact propagation paths, and renders Eq. 1 with the same field and gain, giving the native complex frequency response $\widehat{\mathbf{S}}_{v^\star}$; magnitude, power, and formed images are deterministic downstream observables. Scene geometry is read from the same coefficients. Orthonormality of the real spherical harmonics gives each point a direction-integrated scattering energy

$$E_p = \int_{\mathbb{S}^2} |\rho_p(\mathbf{u})|^2 \, d\mathbf{u} = \sum_{\ell=0}^{L_p} \sum_{m=-\ell}^{\ell} |c_{p\ell m}|^2, \tag{4}$$

which, for the post-training readout only, we splat onto a fixed grid with cloud-in-cell deposition, take the square root of, min–max normalize over the volume, and threshold at a predeclared level (Appendix A).

4 Experiment Design

Datasets. We study the joint NVS–reconstruction task using controlled simulation and measured radar data. We introduce the *RIFT dataset*, generated with the Ansys AVxcelerate RSS ray tracer Ansys, Inc. (2026) for six object meshes: B787, A320, X-59, Fire truck, Race car, and Loader.

It provides complex frequency responses over spherical viewpoints (from a 16×16 transmitter–receiver archive, of which we use one bistatic pair), calibrated sensor geometry, and registered reference meshes for separate signal and geometry evaluation. *GOTCHA* provides measured airborne monostatic circular-SAR observations of a parking-lot scene containing vehicles and calibration targets Ertin et al. (2007). Appendix E specifies the archive metadata, selected acquisitions, and training/validation/test splits.

Metrics. For methods that predict the native complex response, the primary NVS metric is the global complex RelMSE over the held-out test views (on GOTCHA, the validation units): the total energy of the complex prediction error across all held-out views, frequencies, transmitters, and receivers, divided by the total energy of the measured complex responses. The exact expression is given in Appendix A. The main common power column is per-channel matched-range power mapped to normalized dB with training-only statistics and scored over the whole range region of interest (ROI). A complex prediction can therefore receive both signal scores: Complex RelMSE on the native response and common range-power RelMSE after this deterministic conversion. Method-native observables, such as matched-filter amplitude and elevation-integrated 2D power, are reported separately (Appendix G). Geometry is extracted with one fixed-threshold protocol and measured by symmetric squared Chamfer distance, maximum Hausdorff distance, 95th-percentile Hausdorff distance (HD95), solid IoU, and F-score. Maximum Hausdorff and HD95 are reported in parallel.

Implementation details. Appendix D gives the dataset-specific objectives, initialization, optimizer settings, refinement schedules, and training budgets. Test responses and reference geometry do not enter training or checkpoint selection.

Baselines. Each method contributes one checkpoint per scene, selected without test geometry, to both its signal and geometry cells. The seven-method comparison on the RIFT dataset includes RIFT, SpINR-style, matched filtering/backprojection, Sugavanam–Ertin (SE), Radar Fields, GeRaF, and RadarSplat Takawale & Roy (2025a); Ulander et al. (2003); Sugavanam & Ertin (2026); Borts et al. (2024); Lu et al. (2025); Kung et al. (2025); on GOTCHA we compare RIFT, SpINR-style, and backprojection. We did not locate official implementations of SpINRv1, GeRaF 1.0, or SE, so we implemented our own versions: a SpINR-style model that follows the SpINRv1 paper, a GeRaF 1.0 version that reuses available GeRaF release components, and an equation-based SE. Appendix F documents these implementations and our choices where the original descriptions were underspecified.

5 Results and Discussion

5.1 Results and Interpretation

Table 1 reports each method's unweighted mean over the six objects of the RIFT dataset. RIFT has the best value in every column. The same checkpoint supplies every reported output in a row. Per-object values and views appear in Appendix G.1.

Tables 5 and 6 give the per-object values behind every aggregate. Fig. 2 plots these values with their six-object means, and Figs. 8–10 show the corresponding 3D reconstructions in front, side, and plan views.

Radar-presentation mismatch. GeRaF, Radar Fields, and RadarSplat underperform in Table 1; we attribute this mainly to a mismatch between the radar presentation these methods were built for and the problem posed here, which our adaptations do not remove. RIFT solves the SAR and ISAR problems jointly, within one representation: each object is observed from 10,000 viewpoints on a sphere (Appendix E), so one field must explain apertures formed by sensor motion and by target rotation alike and must resolve range, azimuth, and elevation together. GeRaF, Radar Fields, and RadarSplat were instead designed mainly for azimuthal views, with native observables organized in range and azimuth. Converting our measurements into each method's native observable (Appendix F) lets these methods train, but in a problem setting that is not native to them. In addition, Radar Fields, RadarSplat, and GeRaF are trained on matched-filtered power or amplitude, which discards the phase of the measurements, so they cannot use the phase information that RIFT fits.

Table 1: Aggregate results on the RIFT dataset: each entry is the unweighted mean over the six objects, with geometry from the common fixed-$t = 0.20$ readout on the 48^3 evaluator lattice. The signal columns are reserved-test metrics in percent: native complex RelMSE and whole-ROI normalized-dB range-power RelMSE. N/A denotes a structurally unavailable observable. §: RadarSplat is scored on the 6 of 13 range bins its image covers and is not ranked. SE appears twice with the same Stage-1 signal prediction: the published two-stage method, whose geometry is its Stage-2 surface, and Stage 1 only, whose geometry is its Stage-1 scattering field. Bold marks the best populated value per column.

Method	Complex RelMSE ↓	Common range-power RelMSE ↓	Chamfer (10^{-4} m^2) ↓	Max. Hausdorff (mm) ↓	HD95 (mm) ↓	IoU ↑	F-score ↑
RIFT (Ours)	**0.3909%**	**0.5299%**	**0.4813**	**16.71**	**11.95**	**0.2604**	**0.8270**
SpINR-style (arXiv '25)	81.69%	9.223%	0.8179	24.58	12.89	0.1649	0.6620
Backprojection (IEEE TAES '03)	N/A	N/A	1.898	62.47	23.36	0.1566	0.5432
SE (arXiv '26)	46.12%	6.853%	97.94	193.7	151.3	0.006667	0.02406
SE, Stage 1 only	46.12%	6.853%	2.455	116.1	14.39	0.2257	0.7311
Radar Fields (SIGGRAPH '24)	N/A	32.09%	237.7	224.5	178.4	2.544×10^{-6}	2.575×10^{-5}
GeRaF (NeurIPS '25)	124.6%	33.13%	91.03	192.3	151.1	0.01402	0.04832
RadarSplat (ICCV '25)	N/A	67.07%§	67.06	178.6	133.2	0.009130	0.04497

Our comparison therefore evaluates these methods as adapted to our setting; it does not separate the individual contributions of these differences.

GOTCHA vehicles. Table 2 gives the same aggregate over the three measured GOTCHA vehicles (Toyota Camry, Nissan Sentra, and Hyundai Santa Fe), using the HH polarization (horizontal transmit, horizontal receive); Tables 7 and 8 list the per-vehicle values, and Fig. 3 plots the geometry values with their three-vehicle means. RIFT leads on every GOTCHA column except mean IoU, where backprojection is best.

Table 2: Aggregate results across the three measured GOTCHA vehicles (HH). Each entry is the unweighted mean over the three vehicles of the per-vehicle values in Tables 7 and 8, with the common fixed-$t = 0.20$ geometry readout on the 48^3 lattice over each vehicle's 6 m region cube. The signal columns are the pooled full-native complex RelMSE in percent over the train and validation units. N/A denotes a structurally unavailable observable. Bold marks the best populated value per column.

Method	Train complex RelMSE ↓	Validation complex RelMSE ↓	Chamfer (m^2) ↓	Max. Hausdorff (m) ↓	HD95 (m) ↓	IoU ↑	F-score ↑
RIFT (Ours)	**41.18%**	**54.33%**	**0.06254**	**0.9175**	**0.4156**	0.08962	**0.5513**
SpINR-style (arXiv '25)	107.1%	128.5%	2.025	3.560	2.604	0.06905	0.2209
Backprojection (IEEE TAES '03)	N/A	N/A	1.117	3.271	2.653	**0.09510**	0.4416

Reading the joint results. Native-complex NVS and 3D reconstruction probe different properties of the same learned field. Low complex RelMSE requires the model to preserve both amplitude and phase across frequencies at a withheld viewpoint, but it does not require every explanatory coefficient to lie on a physical surface. Conversely, matched-filter focusing or magnitude-domain training can recover useful spatial support without preserving the native measurement. The complex RelMSE, range-power RelMSE, and geometry columns must therefore be read jointly rather than collapsed into one notion of reconstruction quality. Requiring one checkpoint to supply both outputs prevents post-hoc selection of separate signal and geometry specialists.

Evidence across radar configurations. The two datasets provide complementary rather than interchangeable contexts. GOTCHA supplies physically measured, platform-resolved monostatic phase history, together with calibration error, noise, and system effects absent from the controlled simulation. The RIFT dataset supplies controlled Ansys RSS bistatic responses with broad bandwidth and exact registered geometry. RIFT's consistent lead across both dataset tables (every column except GOTCHA's mean IoU) supports robustness of the representation and optimization procedure to these tested measured-monostatic and controlled-bistatic configurations.

Interpreting geometric performance. The five geometry metrics expose different errors. Chamfer measures average surface agreement, Hausdorff exposes the worst outlier, and HD95 distin-

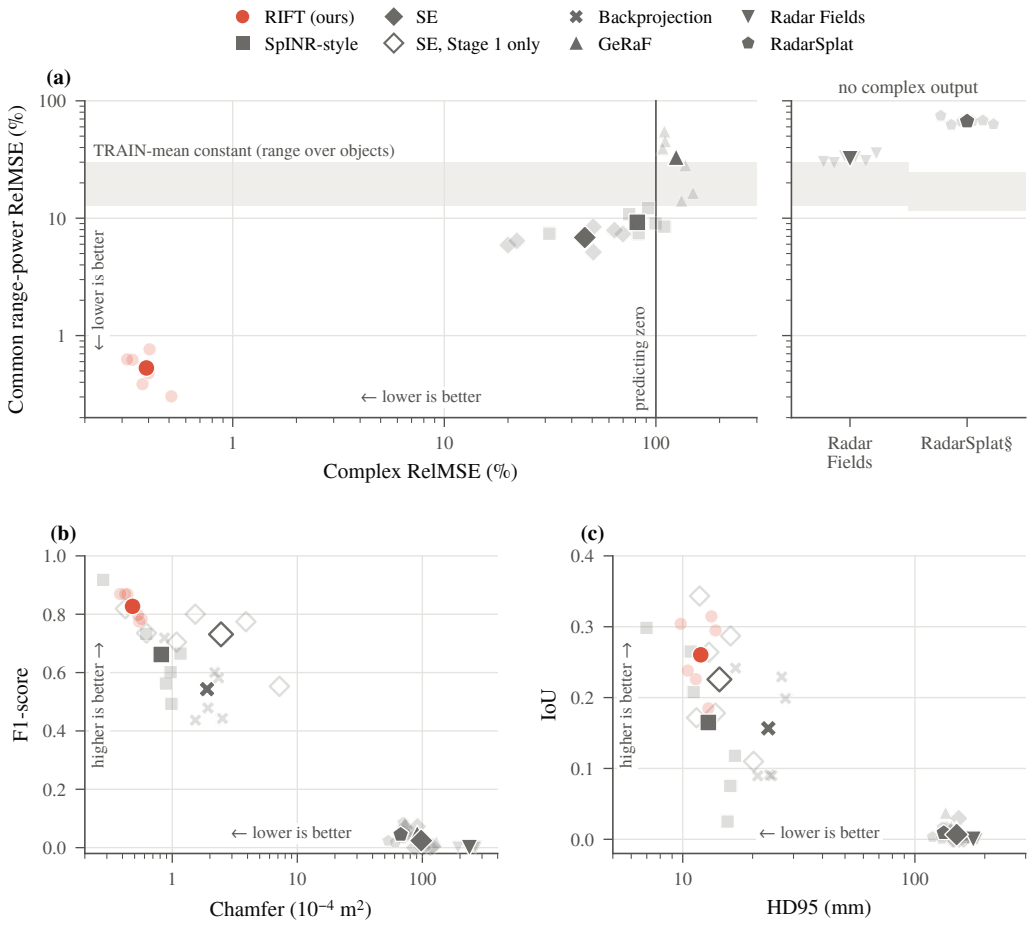

Figure 2: Per-object results (faint markers) and six-object means (solid markers, as in Table 1) on the RIFT dataset; RIFT is red-orange. (a) Complex RelMSE against common matched-range power RelMSE on the reserved-test views, in percent on logarithmic axes. The vertical line marks predicting zero (100% Complex RelMSE), and the shaded band spans the per-object train-mean constant floors (12.76–29.68%; 11.59–24.73% on RadarSplat's 6 bins). Radar Fields and RadarSplat produce no complex response and sit in the right strip; the two SE rows share one Stage-1 signal prediction (filled diamonds). (b) Chamfer distance against F1-score and (c) HD95 against IoU; hollow diamonds are SE with Stage 1 only.

guishes persistent boundary error from a small number of extreme points; IoU and F-score emphasize volumetric overlap and tolerance-scale coverage. Low Chamfer paired with poor Hausdorff therefore suggests mostly accurate support with isolated artifacts, whereas favorable distance metrics but weak IoU can indicate an incomplete or overly thin reconstruction.

5.2 Limitations and Future Work

Calibration and failure modes. RIFT's global complex gain absorbs one scene-wide scale and phase offset but cannot correct viewpoint-, frequency-, transmitter-, or receiver-dependent calibration errors. Residual phase-reference error can therefore dominate GOTCHA RelMSE even when formed imagery appears plausible, while incorrect registration can dominate its geometry metrics. Other failures are endpoint-specific: excess angular capacity can fit training views but degrade held-out phase; weak or specularly redirected surfaces can disappear from the energy readout; sidelobes and omitted scattering paths can create off-surface support; and an SDF- or Gaussian-based baseline can fail its surface or numerical-validity gate even after optimization terminates.

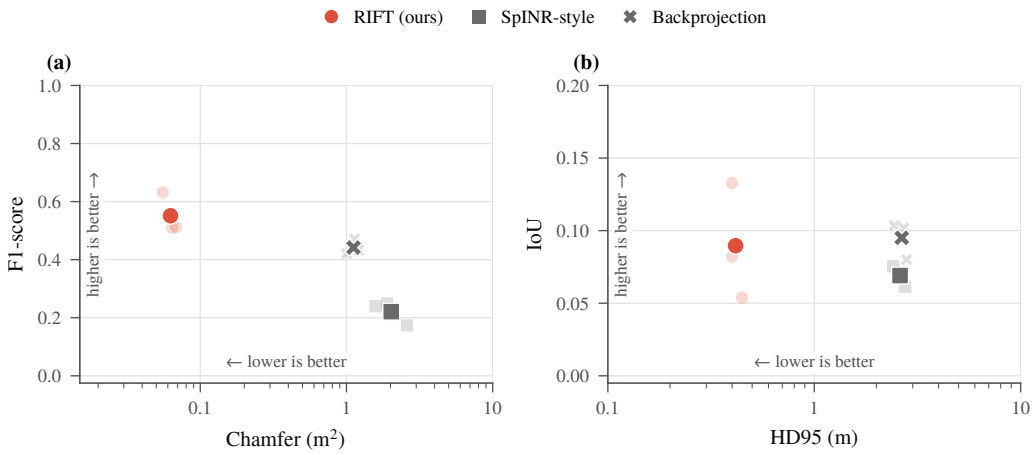

Figure 3: Per-vehicle results and three-vehicle means on GOTCHA (HH): (a) Chamfer distance against F1-score and (b) HD95 against IoU (Table 8), under that table's protocol (fixed threshold $t = 0.20$ on the 48^3 lattice over each vehicle's 6 m region cube, F1 at $\tau = 0.125$ m, solid IoU on 0.1 m cells, ground slab excluded). Each faint marker is one vehicle and method, and each method's solid marker is its unweighted mean over the three vehicles (the value in Table 2); RIFT is red-orange.

Modeling scope. RIFT's scalar single-scattering renderer omits explicit waveform and antenna-pattern responses, polarization, occlusion (which the direction-dependent responses represent only implicitly), diffraction, multiple scattering, dispersive materials, motion, and Doppler, while the RIFT dataset is simulated by one ray tracer from six object meshes scaled to 0.10 m.

Future work. The forward operator uses double-precision phase arithmetic, because single-precision accumulation does not reach the accuracy tolerance of the NUFFT evaluation (Appendix B). Scene reconstruction algorithms that tolerate quantized, reduced-precision arithmetic would make the entire model more lightweight and are a viable direction for future work. RIFT's spatial refinement splits each lattice cell at most once, and on GOTCHA it is replaced by uniform coarse-to-fine densification. Deeper, multi-level refinement driven by the spatial evidence score, on both simulated and measured data, could concentrate points on the scattering surfaces and localize them more finely. The enabled spherical-harmonic degree also varies across the scene (Appendix C); relating it to scene structure, such as flat surfaces, edges, and corners, is a further direction.

6 Conclusion

RIFT learns a direction-dependent complex point field from complex radar measurements, sampled frequencies, and calibrated sensor poses, without visual- or LiDAR-derived inputs. The same field predicts complex responses at unseen viewpoints and yields 3D scattering support. Across six simulated objects, RIFT achieves the best mean value on every reported signal and geometry metric. On three GOTCHA vehicles, it achieves the best mean complex validation RelMSE and the best mean value on every geometry metric except IoU, for which backprojection performs best. These results demonstrate the value of evaluating measurement fidelity and geometric fidelity jointly.

RECOMMENDED AND REQUIRED ICLR AUTHOR STATEMENTS

ETHICS STATEMENT

This work involves no human subjects and no personally identifiable data. The RIFT dataset is simulated with a commercial ray tracer from 3D object meshes, and the measured data are the publicly available GOTCHA Volumetric SAR data set (Casteel et al., 2007). Upon publication, we will publicly release the RIFT dataset and our code, including our adaptations of the baseline models to support fair comparisons, for all parts permitted by the licenses of the third-party meshes and code.

AI USAGE STATEMENT

We used AI in the three ways declared at registration. *To aid or polish writing:* starting from our own drafts, we used AI to improve the style and flow of the paper; no citation was generated by AI. *For retrieval and discovery (e.g., finding related work):* we used AI agents to search for related work, which surfaced baseline models and related works previously unknown to us; we read every retrieved baseline and work, and all citations were composed by hand. *For research ideation or execution:* we used AI to streamline our handwritten code, implementing our pseudocode and hand-derived math equations, and to adapt the baseline models to our datasets; every implementation of handwritten pseudocode or math formulae and every adaptation was checked by hand and tested numerically (for example, our NUFFT implementation matches direct evaluation of the forward model, Appendix B; the baseline adaptations and their checks are documented in Appendix F). We did **NOT** use AI to form the research idea or the mathematical formulation of this work.

REPRODUCIBILITY STATEMENT

Algorithm 1 summarizes RIFT's training loop. Appendix A specifies the forward model, scene supports, geometry readout, and evaluation metrics; Appendix B the NUFFT evaluation and its settings; Appendix C the aspect direction and angular capacity; and Appendix D the data terms, regularization, initialization, optimizer settings, adaptive controller, and GOTCHA densification. Appendix E gives the dataset acquisitions, splits, seeds, and compute. Appendix F documents each baseline implementation, its settings, and its spatial budget (Table 3). Each method is evaluated with one run per scene, using the seeds in Appendix E and the checkpoint-selection criteria in the notes of Tables 5 and 7. Evaluation across six simulated objects and three GOTCHA vehicles assesses performance across scenes; variability across random seeds is not evaluated.

REFERENCES

Ansys, Inc. *Ansys Electronics Desktop, Release 2023 R1*, 2023. URL https://www.ansys.com/products/electronics.

Ansys, Inc. *Ansys AVxcelerate Sensors*, 2026. URL https://www.ansys.com/products/av-simulation/ansys-avxcelerate-sensors. Accessed: 2026-08-11.

David Borts, Erich Liang, Tim Broedermann, Andrea Ramazzina, Stefanie Walz, Edoardo Palladin, Jipeng Sun, David Brueggemann, Christos Sakaridis, Luc Van Gool, Mario Bijelic, and Felix Heide. Radar fields: Frequency-space neural scene representations for fmcw radar. In *ACM SIGGRAPH 2024 Conference Papers*, SIGGRAPH '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400705250. doi: 10.1145/3641519.3657510. URL https://doi.org/10.1145/3641519.3657510.

Curtis H. Casteel, Jr., LeRoy A. Gorham, Michael J. Minardi, Steven M. Scarborough, Kiranmai D. Naidu, and Uttam K. Majumder. A challenge problem for 2d/3d imaging of targets from a volumetric data set in an urban environment. In *Algorithms for Synthetic Aperture Radar Imagery XIV*, volume 6568, pp. 65680D. SPIE, 2007. doi: 10.1117/12.731457.

Laura Dodds, Maisy Lam, Waleed Akbar, Yibo Cheng, and Fadel Adib. Wave-Former: Through-occlusion 3d reconstruction via wireless shape completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21713–21724, June 2026. URL https://arxiv.org/abs/2511.14152.

A. Dutt and V. Rokhlin. Fast fourier transforms for nonequispaced data. *SIAM Journal on Scientific Computing*, 14:1368–1393, 1993. doi: 10.1137/0914081.

Thibaud Ehret, Roger Marí, Dawa Derksen, Nicolas Gasnier, and Gabriele Facciolo. Radar fields: An extension of radiance fields to SAR. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 564–574, June 2024.

Emre Ertin, Christian D. Austin, Samir Sharma, Randolph L. Moses, and Lee C. Potter. GOTCHA experience report: Three-dimensional SAR imaging with complete circular apertures. In *Algorithms for Synthetic Aperture Radar Imagery XIV*, volume 6568, pp. 656802. SPIE, 2007. doi: 10.1117/12.723245.

Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *CVPR*, 2022.

Leslie Greengard and June-Yub Lee. Accelerating the nonuniform fast fourier transform. *SIAM Review*, 46(3): 443–454, 2004. doi: 10.1137/S003614450343200X.

Jakob Hoydis, Fayçal Aït Aoudia, Sebastian Cammerer, Merlin Nimier-David, Nikolaus Binder, Guillermo Marcus, and Alexander Keller. Sionna RT: Differentiable ray tracing for radio propagation modeling. In *2023 IEEE Globecom Workshops*, 2023. URL https://arxiv.org/abs/2303.11103.

Shengyu Huang, Zan Gojcic, Zian Wang, Francis Williams, Yoni Kasten, Sanja Fidler, Konrad Schindler, and Or Litany. Neural LiDAR fields for novel view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 18236–18246, October 2023.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24118–24129, June 2024.

Samah Hussein, Junfeng Guan, Swathi Narashiman, Saurabh Gupta, and Haitham Hassanieh. 3d object reconstruction with mmwave radars. *arXiv preprint arXiv:2504.12348*, 2025. URL https://arxiv.org/abs/2504.12348.

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

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*, pp. 27596–27606, October 2025.

Zhengxin Lei, Feng Xu, Jiangtao Wei, Feng Cai, Feng Wang, and Ya-Qiu Jin. SAR-NeRF: Neural radiance fields for synthetic aperture radar multi-view representation. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024. doi: 10.1109/TGRS.2024.3452505.

Aobo Li, Zhengxin Lei, Jiangtao Wei, and Feng Xu. SAR-GS: Gaussian splatting based SAR images rendering and target reconstruction. *arXiv preprint arXiv:2506.21633*, 2025. URL https://arxiv.org/abs/2506.21633.

Afei Liu, Shuanghui Zhang, Chi Zhang, Shuaifeng Zhi, and Xiang Li. RaNeRF: Neural 3d reconstruction of space targets from ISAR image sequences. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023. doi: 10.1109/TGRS.2023.3298067.

Zinan Liu, Shengren Niu, Xiaolan Qiu, Lingxiao Peng, Yuying Shang, Lihua Zhong, and Chibiao Ding. A differentiable method for novel view SAR image generation via 3d gaussian splatting. *ISPRS Journal of Photogrammetry and Remote Sensing*, 231:167–195, 2026. doi: 10.1016/j.isprsjprs.2025.10.034.

Lone Wolf. Nissan Sentra 2005. 3D Warehouse model, 2012. URL https://3dwarehouse.sketchup.com/model/e4a15cca51107dd035da01f298003d56/Nissan-Sentra-2005. License: 3D Warehouse General Model License. Accessed: 2026-09-24.

Jiachen Lu, Hailan Shanbhag, and Haitham Al Hassanieh. GeRaF: Neural geometry reconstruction from radio frequency signals. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://papers.nips.cc/paper_files/paper/2025/hash/8766fbc68e1ed1cdef712ce273e0a363-Abstract-Conference.html.

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

Mahathir Monjur and Shahriar Nirjon. mmWEAVER: Environment-specific mmwave signal synthesis from a photo and activity description. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1875–1884, March 2026.

Nieve5677. Toyota Camry (Mk4)(XV20) 1997. Sketchfab 3D model, 2025. URL https://sketchfab.com/3d-models/toyota-camry-mk4xv20-1997-196829bf4cbc4b8cb26c50964f35ff27. License: CC BY 4.0. Accessed: 2026-09-22.

Clifford J Nolan and Margaret Cheney. Synthetic aperture inversion. *Inverse Problems*, 18(1):221, jan 2002. doi: 10.1088/0266-5611/18/1/315. URL https://dx.doi.org/10.1088/0266-5611/18/1/315.

Mahan Rafidashti, Ji Lan, Maryam Fatemi, Junsheng Fu, Lars Hammarstrand, and Lennart Svensson. NeuRadar: Neural radiance fields for automotive radar point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2513–2523, June 2025.

Albert Reed, Juhyeon Kim, Thomas Blanford, Adithya Pediredla, Daniel Brown, and Suren Jayasuriya. Neural volumetric reconstruction for coherent synthetic aperture sonar. *ACM Trans. Graph.*, 42(4), jul 2023. ISSN 0730-0301. doi: 10.1145/3592141. URL https://doi.org/10.1145/3592141.

Kunzhe Song, Geo Jie Zhou, Xiaoming Liu, and Huacheng Zeng. Rascene: High-fidelity 3d scene imaging with mmwave communication signals. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 36454–36463, June 2026.

Nithin Sugavanam and Emre Ertin. Neural implicit representations for 3d synthetic aperture radar imaging. *arXiv preprint arXiv:2602.17556*, 2026. URL https://arxiv.org/abs/2602.17556.

Harshvardhan Takawale and Nirupam Roy. SpINR: Neural volumetric reconstruction for FMCW radars. *arXiv preprint arXiv:2503.23313*, 2025a. URL https://arxiv.org/abs/2503.23313.

Harshvardhan Takawale and Nirupam Roy. SpINRv2: Implicit neural representation for passband FMCW radars. *arXiv preprint arXiv:2506.08163*, 2025b. URL https://arxiv.org/abs/2506.08163.

Tang Tao, Longfei Gao, Guangrun Wang, Yixing Lao, Peng Chen, Hengshuang Zhao, Dayang Hao, Xiaodan Liang, Mathieu Salzmann, and Kaicheng Yu. LiDAR-NeRF: Novel LiDAR view synthesis via neural radiance fields. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 390–398, 2024. doi: 10.1145/3664647.3681482.

teenlin3. 2004 Hyundai Santa- Fe. Sketchfab 3D model, 2026. URL https://sketchfab.com/3d-models/2004-hyundai-santa-fe-b33c872e79b24508a1b9213e3537651c. License: CC BY 4.0. Accessed: 2026-09-24.

L.M.H. Ulander, H. Hellsten, and G. Stenstrom. Synthetic-aperture radar processing using fast factorized back-projection. *IEEE Transactions on Aerospace and Electronic Systems*, 39(3):760–776, 2003. doi: 10.1109/TAES.2003.1238734.

Omkar Shailendra Vengurlekar, Adithya Pediredla, and Suren Jayasuriya. SH-SAS: An implicit neural representation for complex spherical-harmonic scattering fields for 3d synthetic aperture sonar. In *2026 International Conference on 3D Vision*, pp. 967–976, 2026. doi: 10.1109/3DV69130.2026.00097.

Jiarui Zhang, Zhihao Li, Chong Wang, and Bihan Wen. RF4D: Neural radar fields for novel view synthesis in outdoor dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15387–15397, June 2026a.

Zixuan Zhang, Yunkai Deng, Wei Xiang, Heng Zhang, and Fengli Xue. SAR-3DGS: 3d gaussian splatting for synthetic aperture radar novel view synthesis and target reconstruction. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pp. 1–20, 2026b. doi: 10.1109/JSTARS.2026.3673407.

Kaichen Zhou, Laura Dodds, Sayed Saad Afzal, and Fadel Adib. RISE: Single static radar-based indoor scene understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 32194–32205, June 2026.

A Additional Forward-Model and Evaluation Details

Scene support. The support in Eq. 2 is $\Omega = [-0.15, 0.15]^3$ m for the RIFT dataset. For each GOTCHA vehicle, Ω is the box $[-3, 3] \times [-1.5, 1.5] \times [-1.25, 1.25]$ m in a vehicle-aligned local frame (x along the vehicle, y left, z up) centered on the vehicle's region of interest. The local origins are $(20.72, -19.23, 0.77)$ m for the Toyota Camry, $(22.83, -28.88, 0.77)$ m for the Nissan Sentra, and $(29.21, -19.81, 0.77)$ m for the Hyundai Santa Fe in the native GOTCHA frame.

Propagation geometry. The formulas below expand the geometric definitions used in Sec. 3.1. All experiments use one transmitter–receiver pair per viewpoint, so pair indices are omitted. For calibrated transmitter and receiver positions $\mathbf{p}_v^{\mathrm{tx}}$ and $\mathbf{p}_v^{\mathrm{rx}}$ (one antenna for GOTCHA's monostatic pulses) and active point positions $\{\mathbf{x}_p\}_{p\in\mathcal{A}}$, the two path legs, their bistatic sum, and the free-space wavenumber are

$$\begin{aligned} R_{vp}^{\mathrm{tx}} &= \left\|\mathbf{x}_p - \mathbf{p}_v^{\mathrm{tx}}\right\|_2, \\ R_{vp}^{\mathrm{rx}} &= \left\|\mathbf{x}_p - \mathbf{p}_v^{\mathrm{rx}}\right\|_2, \\ R_{vp} &= R_{vp}^{\mathrm{tx}} + R_{vp}^{\mathrm{rx}}, \qquad k_i = \frac{2\pi f_i}{c_0}, \end{aligned} \tag{A.1}$$

where c_0 is the propagation speed. Equation A.1 uses the exact antenna-to-point distances, so each point contributes a spherical wavefront along its exact transmitter–point–receiver path, with no plane-wave, far-field, or array phase-center approximation. The GRT forward model of Eq. 1 is therefore valid in the near field.

RIFT uses the sum-range-squared geometric-spreading convention

$$A_{vp} = \frac{1}{(4\pi)^2\left(R_{vp}^2 + \epsilon_R\right)}, \tag{A.2}$$

where $\epsilon_R = 10^{-9}$ is the fixed numerical-stability constant. This is the deterministic amplitude weight used in Eq. 1. The point coefficients already carry scattering weight, so no voxel-volume factor is applied.

GOTCHA phase reference. For a monostatic pulse at $\mathbf{a}_q$, let $r_{qp} = \|\mathbf{x}_p - \mathbf{a}_q\|_2$ and let r_q^{ref} be the reference range supplied with the data. The GOTCHA responses are referenced as

$$e^{-\mathrm{i}4\pi f_{q,i}(r_{qp} - r_q^{\mathrm{ref}})/c_0} = D_{q,i} e^{-\mathrm{i}4\pi f_{q,i} r_{qp}/c_0}, \qquad D_{q,i} = e^{+\mathrm{i}4\pi f_{q,i} r_q^{\mathrm{ref}}/c_0}. \tag{A.3}$$

$D_{q,i}$ is a known unit-magnitude factor from the data's range referencing; it varies with pulse and frequency and contains no learned parameter. The shared spreading law uses the physical path $2r_{qp}$, not the reference-adjusted path. The forward operator is evaluated with the NUFFT of Appendix B at each pulse's selected frequencies.

RIFT geometry readout. The geometry readout deposits each point's energy E_p (Eq. 4) onto a 48^3 cell-center grid by cloud-in-cell deposition, without upsampling (6.25 mm pitch over the 0.30 m cube on the RIFT dataset; 0.125 m over each vehicle's 6 m cube on GOTCHA). For deposited energy V_j, define $z_j = (\sqrt{V_j} - \min_k \sqrt{V_k})/(\max_k \sqrt{V_k} - \min_k \sqrt{V_k} + 10^{-30})$ and retain the cell centers with $z_j > 0.20$. The raster is used only for geometry evaluation, and the min–max normalization makes it independent of how the overall scale is split between the gain g and the coefficients.

Geometry metrics. Let $\mathcal{P}$ be the retained lattice centers, $\mathcal{G}$ the reference-surface samples, and $d(\mathbf{a}, \mathcal{B}) = \min_{\mathbf{b}\in\mathcal{B}} \|\mathbf{a} - \mathbf{b}\|_2$. The symmetric squared Chamfer distance is $|\mathcal{P}|^{-1} \sum_{\mathbf{p}\in\mathcal{P}} d(\mathbf{p}, \mathcal{G})^2 + |\mathcal{G}|^{-1} \sum_{\mathbf{q}\in\mathcal{G}} d(\mathbf{q}, \mathcal{P})^2$. The maximum Hausdorff distance is the larger of the two directed maxima of d, and HD95 is the larger of the two directed 95th percentiles. Precision is the fraction of $\mathcal{P}$ within τ of $\mathcal{G}$, recall is the fraction of $\mathcal{G}$ within τ of $\mathcal{P}$, and F1 is their harmonic mean. Solid IoU compares the grid cells occupied by $\mathcal{P}$ with those occupied by samples of the solid voxelized reference mesh. **Because each retained lattice center occupies a single IoU cell and the IoU cells (5 mm on the RIFT dataset, 0.1 m on GOTCHA) are smaller than the lattice pitch (6.25 mm and 0.125 m), even a complete readout occupies at most about $(5/6.25)^3 \approx 0.51$ of the reference solid's cells, so solid IoU is capped near 0.5 for every method.**

Relative signal error. For a set of held-out views $\mathcal{V}_{\text{role}}$ (validation or test), the complex RelMSE is

$$\text{RelMSE} = \frac{\sum_{v\in\mathcal{V}_{\text{role}}} \|\widehat{\mathbf{S}}_v - \mathbf{S}_v\|_F^2}{\sum_{v\in\mathcal{V}_{\text{role}}} \|\mathbf{S}_v\|_F^2}. \tag{A.4}$$

The sums aggregate all frequencies before normalization. The range-power RelMSE uses the same ratio after prediction and measurement are both converted to matched-range power in normalized dB (normalized with training statistics), summing over the ROI range bins of every view.

B Range-Factorized NUFFT Evaluation

The factorization is derived below for a uniform frequency grid; on GOTCHA the same operator is evaluated at each pulse's exact selected frequencies, with the known reference-phase factor given in Appendix A.

Evaluated directly, Eq. 1 requires $O(|\mathcal{A}|N_f)$ complex exponentials per viewpoint, one per point and frequency. The sum over points must remain, but the frequency axis has exploitable structure: the sampled frequencies are uniformly spaced, so the propagation phase separates into a frequency-independent carrier and a term linear in the frequency index. Writing $f_i = f_{\text{ref}} + (i - i_{\text{ref}})\Delta f$ with reference bin $i_{\text{ref}} = \lfloor N_f/2 \rfloor$ and $k_{\text{ref}} = 2\pi f_{\text{ref}}/c_0$,

$$\begin{aligned} \exp(\mathrm{i}s_\phi k_i R_{vp}) = \exp(\mathrm{i}s_\phi k_{\text{ref}} R_{vp}) \\ \exp(\mathrm{i}\xi_{vp}\,(i - i_{\text{ref}}))\,, \\ \xi_{vp} = s_\phi \frac{2\pi\Delta f}{c_0} R_{vp}, \qquad s_\phi = -1. \end{aligned} \tag{B.1}$$

For one fixed viewpoint, Eq. 1 therefore reduces to

$$\widehat{S}[i] = g \sum_{p\in\mathcal{A}} B_p \exp(\mathrm{i}\xi_p\,(i - i_{\text{ref}}))\,, \qquad B_p = A_{vp}\,\rho_p(\mathbf{u}_v)\exp(\mathrm{i}s_\phi k_{\text{ref}} R_{vp})\,, \tag{B.2}$$

where $\xi_p = \xi_{vp}$ and $R_p = R_{vp}$ abbreviate the fixed-viewpoint quantities. Equation B.2 is exactly a one-dimensional type-1 (nonuniform-to-uniform) discrete Fourier sum. In NUFFT terminology, a *source* is the weighted scene contribution (ξ_p, B_p); it is not the radar transmitter. Its location ξ_p is an arbitrary real number set by the exact path length, while the outputs are the N_f uniform frequency modes. Because the ξ_p are not lattice points, a plain fast Fourier transform (FFT) cannot be applied directly. We instead use a nonuniform fast Fourier transform (NUFFT) (Dutt & Rokhlin, 1993; Greengard & Lee, 2004), computed as Gaussian gridding followed by one oversampled inverse FFT per viewpoint.

The evaluation has three steps. First, with dimensionless oversampling factor σ we set $M = 2^{\lceil \log_2(\sigma N_f)\rceil}$. Each range-derived phase slope is converted from radians to the dimensionless fractional grid coordinate

$$u_p = \frac{M\xi_p}{2\pi} \bmod M = \left(s_\phi M \frac{\Delta f}{c_0} R_p\right) \bmod M. \tag{B.3}$$

Although the scene point is three-dimensional, u_p is scalar because, for a fixed viewpoint, frequency-dependent phase depends on position only through the scalar two-leg range R_p; gridding therefore places every point on a single range axis at its path length. A value such as $u_p = 137.4$ lies between auxiliary bins 137 and 138. We distribute its complex amplitude B_p over the J nearby integer bins with Gaussian kernel

$$\varphi(u_p - \widetilde{m}) = \exp\left(-\frac{(u_p - \widetilde{m})^2}{4\tau}\right), \qquad \widetilde{m} \in \mathcal{N}_J(u_p) = \left\{\lfloor u_p \rfloor - \tfrac{J}{2} + 1, \ldots, \lfloor u_p \rfloor + \tfrac{J}{2}\right\}, \tag{B.4}$$

where J is the number of retained Gaussian taps and τ controls their width. Because the Gaussian has infinite mathematical support, restricting it to J bins is the approximation that makes gridding finite. Writing $\widetilde{m}$ for an unwrapped neighbour and $m = \widetilde{m} \bmod M$ for its stored circular-grid index, the accumulation for one fixed viewpoint v is

$$G_v[m] \;\mathrel{+}=\; B_p\,\varphi(u_p - \widetilde{m}), \qquad \widetilde{m} \in \mathcal{N}_J(u_p), \tag{B.5}$$

starting from $G_v[m] = 0$. Contributions from different points that reach the same auxiliary bin add as complex numbers.

Second, the positive-exponent convention in Eq. B.2 gives the apodized mode sequence

$$Y_v[\ell] = M \ \mathrm{IFFT}(G_v)\,[\ell \bmod M]. \tag{B.6}$$

Gaussian spreading convolves the nonuniform sources on the auxiliary grid, so Eq. B.6 is multiplied by the Fourier response of that kernel. We remove this numerical taper by *deapodization*. Let G_0 be the grid obtained by spreading one unit source at $u = 0$ with the same truncated kernel, and define

$$d_i = M \ \mathrm{IFFT}(G_0)[(i - i_{\mathrm{ref}}) \bmod M], \qquad \widehat{S}[i] = g\,\frac{Y_v[i - i_{\mathrm{ref}}]}{d_i}. \tag{B.7}$$

Thus d_i is a deterministic correction computed from (M, J, τ). Computing d_i numerically accounts exactly for the J-tap truncation and circular indexing for a source on a grid node; for sources between nodes, the small residual is part of the error measured below. Centering the retained modes on i_{ref} keeps all desired outputs away from the oversampled grid's Nyquist edge, where division by a small kernel response would amplify error.

The cost becomes $O(J|\mathcal{A}|)$ for gridding plus $O(M \log M)$ for the transforms, so the per-scatterer frequency work falls from N_f to J. We use $\sigma = 2$, $J = 20$, $M = 2048$ for the 600-frequency RIFT-dataset responses, and $\tau = 0.05J + 0.08$. With these settings and double-precision phase arithmetic, the rendered response matches direct evaluation of Eq. 1 to a relative error below 10^{-9} in our numerical tests; the factorized operator is therefore an accelerated evaluation of the same physical model rather than a different model. Single-precision phase accumulation does not reach this tolerance, so ranges, phases, and NUFFT arithmetic use float64/complex128, computed from point and antenna positions stored in float32; the learned coefficients and raw point offsets remain float32. Scene reconstruction algorithms that tolerate quantized, reduced-precision phase arithmetic would reduce memory and compute cost and make the entire model more lightweight; they are a viable direction for future work (Section 5.2).

Only the frequency axis is factorized. The exact Euclidean ranges of Eq. A.1, the per-path spreading of Eq. A.2, and the aspect-dependent coefficients are all carried inside B_p and are evaluated per point, so no far-field, plane-wave, phase-center, or common-range approximation is introduced. Only the frequency dependence has to factor through R_{vp}, which leaves the amplitude term free to be any function of the two leg lengths.

Gradients propagate through B_p and through the smooth Gaussian tap weights of Eq. B.4; the only omitted derivative is that of the piecewise-constant J-bin support selection, whose jumps occur at the kernel tail. The same factorization transposes, so the adjoint used for the backprojection warm start of Sec. 3.3 scatters the residual into the mode bins, applies one forward FFT per viewpoint, and gathers with the same Gaussian taps. The factorization needs the complete uniform frequency grid: when an objective uses a subset of frequency bins, the full grid is rendered and the requested bins are gathered afterwards. The same operator is used during optimization and at held-out viewpoints, so training and novel-view rendering share one numerical path.

C Aspect Direction and Adaptive Angular Capacity

Section 3.2 evaluates the scattering field using the calibrated array-center aspect direction. Let $\mathbf{p}_v$ denote the array-center position relative to the scene origin. We define

$$\mathbf{u}_v = \frac{\mathbf{p}_v}{\|\mathbf{p}_v\|_2}. \tag{C.1}$$

Propagation itself uses the exact transmitter and receiver positions in Eq. 1. On the RIFT dataset, $\mathbf{p}_v$ is the center of the full source array; on GOTCHA it is the recorded antenna position of each monostatic pulse. The coefficients are shared across all frequencies, so frequency dependence enters only through the propagation phase; RIFT learns no dispersive material model.

The maximum spherical-harmonic degree is $L_{\max} = 3$ on the RIFT dataset and $L_{\max} = 4$ on GOTCHA. Every point of the initial lattice starts at $L_p = 0$, and spatial siblings inherit their parent's enabled degree. On the RIFT dataset, every 16th training view, with the probed views shifted

each epoch, the controller exposes the next locked band with zero coefficients and measures its RMS complex data-gradient, the angular score of Section 3.4; since these coefficients are zero, probing leaves the prediction unchanged. At each ten-epoch refinement event, the top $1/16$ of eligible points by angular score unlock one band; eligibility requires the exposure, maturity, and cooldown conditions of Appendix D. The enabled degree therefore varies by point and is determined from training-only gradient evidence, not from a ground-truth geometry target. GOTCHA uses the same probing and unlock rule at refinement events after every tenth epoch (Appendix D); points created by its densification start their maturity count when they are created.

At the validation-selected B787 checkpoint, the enabled-degree histogram is 85,705 points at degree zero, 2,541 at degree one, 2,549 at degree two, and 37,787 at degree three, for 128,582 active points total. These counts describe enabled bands, not nonzero coefficients. The enabled degree differs across points because the data demand different angular variation at different locations. Relating the spatial distribution of enabled degrees to the structure of the underlying scene, such as flat surfaces, edges, and corners, is a direction for future work (Section 5.2).

D Optimization Parameterization and Regularization

Dataset-specific data terms. For the RIFT dataset, the per-view term in Eq. 3 is

$$\ell_v = \frac{\|\widehat{\mathbf{S}}_v - \mathbf{S}_v\|_F^2}{N_f}, \tag{D.1}$$

using the selected acquisition in Appendix E and raw complex measurement units. For GOTCHA HH, a training view is a (pass, sector) unit with selected pulses $\mathcal{Q}_v$ (Appendix E). With N_q the number of selected frequencies of pulse q,

$$\ell_v = \frac{1}{|\mathcal{Q}_v|\,\sigma_{\mathrm{tr}}^2} \sum_{q\in\mathcal{Q}_v} \frac{\|\widehat{\mathbf{y}}_q - \mathbf{y}_q\|_2^2}{N_q}, \qquad \sigma_{\mathrm{tr}}^2 = \frac{\sum_{q\in\mathcal{Q}_{\mathrm{tr}}} \|\mathbf{y}_q\|_2^2}{\sum_{q\in\mathcal{Q}_{\mathrm{tr}}} N_q}. \tag{D.2}$$

The residual is the full native complex response of each pulse within our region of interest. The normalization uses training data only and is shared across views, not fitted per view. GOTCHA checkpoint selection uses the pooled complex validation RelMSE in the same domain; RIFT-dataset selection uses validation complex data loss. On both datasets ℓ_v compares real and imaginary components without a separate wrapped-phase loss.

Global gain and regularization. The single global complex calibration is parameterized by unconstrained $\alpha, \beta \in \mathbb{R}$ as

$$g = \exp(\alpha)\exp(\mathrm{i}\beta). \tag{D.3}$$

It is shared by all viewpoints and frequencies, so it absorbs only one scene-wide scale and phase, not view-dependent discrepancies.

Let N_0 be the fixed initial active count ($48^3 = 110{,}592$ on the RIFT dataset and $48 \times 24 \times 20 = 23{,}040$ on GOTCHA) and $\epsilon_c = 10^{-24}$. The two regularizers in Eq. 3 are

$$\begin{aligned} \mathcal{R}_{\mathrm{sp}} &= \frac{|g|}{N_0} \sum_{p\in\mathcal{A}} \left(\sqrt{\sum_{\ell,m} |c_{p\ell m}|^2 + \epsilon_c} - \sqrt{\epsilon_c} \right), \\ \mathcal{R}_{\mathrm{SH}} &= \frac{|g|^2}{N_0} \sum_{p\in\mathcal{A}} \sum_{\ell,m} \ell(\ell+1)|c_{p\ell m}|^2. \end{aligned} \tag{D.4}$$

The RIFT-dataset weights are $\lambda_{\mathrm{sp}} = 3 \times 10^{-7}$ and $\lambda_{\mathrm{SH}} = 10^{-9}$; GOTCHA's weight transfer is specified below. $\mathcal{R}_{\mathrm{sp}}$ promotes group sparsity over points, each group being one point's complete set of coefficients. $\mathcal{R}_{\mathrm{SH}}$ is rotation invariant, leaves the isotropic band ($\ell = 0$) unpenalized, and increasingly suppresses higher angular degrees. Normalization by fixed N_0, rather than the evolving active count, prevents the RIFT dataset's zero-valued siblings from diluting either prior; GOTCHA keeps N_0 and its transferred weights fixed through densification, whose children are interpolated rather than zero.

Under $c_{p\ell m} \mapsto \nu c_{p\ell m}$ and $g \mapsto g/\nu$ for nonzero $\nu \in \mathbb{C}$, $\mathcal{R}_{\mathrm{SH}}$ is exactly gauge invariant, while $\mathcal{R}_{\mathrm{sp}}$ is invariant as $\epsilon_c \to 0$ and approximately invariant when a group's energy dominates ϵ_c. RIFT uses no magnitude loss or warmup, view reweighting, gradient clipping, weight decay, or scheduled pruning; on GOTCHA, weak points are removed only at densification events.

RIFT-dataset initialization and optimizer. The initial 110,592 anchors are the centers of a 48^3 partition of the 0.30 m cube, with 6.25 mm pitch and 3.125 mm half-width. Raw offsets begin at zero and only degree zero is enabled. The first 100 training views initialize the isotropic coefficient by backprojection with the adjoint of the NUFFT operator (Appendix B). A complex least-squares scale fitted on the training views, including the Y_{00} factor, sets the coefficient scale; the global gain is then warm-started once on the training views and optimized jointly afterwards. The point set is capped at 262,144 point scatterers, each with at most 16 complex coefficients ($L_{\max} = 3$).

AdamW uses learning rate 0.003 for coefficients, raw positions, and gain, betas $(0.9, 0.999)$, $\epsilon = 10^{-8}$, and zero weight decay. CosineAnnealingWarmRestarts uses $T_0 = 10$, $T_{\mathrm{mult}} = 2$, and $\eta_{\min} = 10^{-6}$ with epoch-level stepping. There is one optimizer update per training view, or 2,400 per epoch in the selected RIFT-dataset protocol. The budget is 150 epochs (360,000 updates), with a fixed seed-42 view order; the warm-restart cycles last 10, 20, 40, and 80 epochs. We choose the checkpoint with minimum validation complex data loss; the same checkpoint supplies signal and geometry outputs.

RIFT-dataset adaptive controller. The adaptive controller proceeds as follows:

1. For every training-view update, accumulate the exposure-averaged norm of the cell-scaled world-coordinate sensitivity from the incremental data loss, before regularization. RIFT divides out the tanh coordinate Jacobian using a relative 10^{-6} floor.
2. Every 16th training view, with the probed views shifted each epoch, expose the next locked zero-coefficient SH band and accumulate its RMS complex data-gradient over probe exposures.
3. Every ten epochs, snapshot both scores before mutation. A spatial candidate needs an unsplit cell and at least 3,200 spatial exposures; an angular candidate needs a locked band and at least 200 angular exposures. Both need a positive finite score. A point is not eligible for an action at the event right after it received that action, and a point created at an event is not eligible at the next one.
4. Select the top $1/512$ of eligible spatial points and top $1/16$ of eligible angular points, rounding up and respecting the 262,144-point cap. Apply angular growth first, then spatial refinement, using the same pre-mutation score snapshot.
5. Angular growth enables one zero-initialized complete SH band and clears only its Adam moments. Spatial refinement keeps the parent's current position, coefficients, and coefficient moments in place; recenters its anchor; halves its half-width h_p; resets coordinate moments; and activates seven zero-coefficient siblings. All remain inside the original domain, and only one subdivision level is permitted.
6. Reset evidence statistics only after both decisions and actions. Because new angular coefficients and siblings are zero, each mutation preserves the current complex prediction in exact arithmetic while adding trainable capacity.

The full B787 run ends after 15 events, 3,203 spatial splits, and 92,248 angular unlocks with 133,013 points. The validation-selected state (epoch 121) follows 12 events and has 128,582 points (108,022/20,560 at levels 0/1) and 1,832,550 active real scene parameters, excluding gain, buffers, and optimizer state.

GOTCHA initialization and regularization transfer. On GOTCHA, RIFT begins with a $48 \times 24 \times 20$ lattice of cell-center anchors at 0.125 m pitch over our region of interest, zero offsets, and degree zero. It applies the adjoint of the native operator to the responses from the first 100 (pass, sector) units in the seed-42 order. A complex least-squares coefficient scale and a training-only gain warm start follow. A prediction-preserving transformation $c \mapsto ac$, $g \mapsto g/a$ sets the initial mean per-point coefficient norm to the B787 reference value 3.445×10^{-3}.

The regularizer forms and fixed-initial-count normalization remain unchanged. To account for differing data units and propagation scales, define initial moments $m_1 = |g| \operatorname{mean}_p \|c_p\|_2$ and $m_2 = |g|^2 \operatorname{mean}_p \|c_p\|_2^2$. The B787 reference fixes dimensionless strengths $\mu_1 = \lambda_{\mathrm{sp}} m_1 / \sigma_{\mathrm{tr}}^2 \simeq 0.115$ and $\mu_2 = \lambda_{\mathrm{SH}} m_2 / \sigma_{\mathrm{tr}}^2 \simeq 1.46 \times 10^{-6}$. Because the GOTCHA data term is already power-normalized, its weights are $\lambda_{\mathrm{sp}}^G = \mu_1 / m_1^G$ and $\lambda_{\mathrm{SH}}^G = \mu_2 / m_2^G$, determined at its own initialized scene. Priors are added once per update and excluded from refinement evidence.

GOTCHA optimizer and controller. AdamW uses one learning rate of 6×10^{-4} for coefficients, gain, and positions, betas $(0.9, 0.999)$, $\epsilon = 1.4909$ (the RIFT dataset's 10^{-8} divided by the RIFT dataset's training power, matching GOTCHA's power-normalized data term), and zero weight decay, with cosine warm restarts ($T_0 = 10$, $T_{\mathrm{mult}} = 2$, $\eta_{\min} = 10^{-6}$) stepped once per epoch. One update accumulates one (pass, sector) unit in a fixed seed-42 order, 2,410 updates per epoch, for a budget of 40 epochs.

Spatial capacity grows by coarse-to-fine densification after epochs 4, 10, and 16. Points whose coefficient norm is below 10^{-3} of the maximum are removed, at most the 131,072 strongest points are kept, and each is replaced by eight children at half its pitch (anchors offset by a quarter pitch), with coefficients interpolated trilinearly from the current lattice and scaled by $1/8$, zero offsets, and zeroed Adam moments. The lattice pitch refines from 0.125 m to 0.0156 m, up to 1,048,576 points. After each event all learning rates, with the cosine schedule's base and floor, are scaled down, never up, so that $(\eta/\epsilon)\,\lambda_{\max} \leq 1$, where η is the base learning rate and $\lambda_{\max}$ is the largest eigenvalue of one unit's data-term Hessian with respect to the coefficients, estimated by ten power iterations at the current gain on two fixed training units and maximized over them. The shared controller probes angular evidence every 16 updates, with the probed units shifted each epoch, and runs refinement events after every tenth epoch with angular fraction $1/16$, spatial fraction zero (densification replaces the spatial split), the RIFT dataset's exposure, cooldown, and maturity rules, and maximum SH degree four. Fig. 4 illustrates the GOTCHA acquisition, the initialization, and the densification.

E Dataset Acquisitions and Experimental Protocols

RIFT dataset archive. Each of the six objects has 10,000 Fibonacci-sphere viewpoints at 10 m standoff. The source acquisition has 16 transmitters and 16 receivers, 600 frequencies from 8.5 to 11.495 GHz in 5 MHz steps, and one chirp. Each view therefore stores a complex $16 \times 16 \times 600$ response. The nominal carrier and bandwidth are 10 GHz and 3 GHz. Source meshes are centered and scaled to a maximum dimension of 0.10 m; the reconstruction support is the larger cube specified in Appendix A. The stored transmitter and receiver element positions define the propagation paths. Reference meshes are used only for registered geometry evaluation.

Selected RIFT-dataset protocol. All methods use transmitter 0 and receiver 0 of the archive with all 600 frequencies, so each view's response is $1 \times 1 \times 600$. A seed-42 PCG64 permutation selects 2,400 training views (the first 2,400 of a 3,200-view training split), 1,000 validation views, and 1,000 reserved test views; the remaining 5,600 views are reserved for future validation by other researchers. Every method uses the same views and splits. Only selected training responses enter normalization, initialization, and target preparation; test responses remain sealed during training and checkpoint selection.

GOTCHA acquisition and selected protocol. GOTCHA records an airborne monostatic parking-lot scene near 9.6 GHz with nominal 640 MHz bandwidth. The selected protocol, identical for the Toyota Camry, Nissan Sentra, and Hyundai Santa Fe, uses HH from all eight elevation passes, each divided into 360 one-degree sectors. One scene field is shared across passes; the passes are not independent scene fits. The 55 sector identifiers of the seeded reserved-test role stay sealed. Every other (pass, sector) unit is used, and 10% of the pass-4 units, chosen with seed 42, form the validation role. This gives 2,410 training and 30 validation units, identical across the three regions of interest; validation units interpolate between neighboring passes rather than extrapolating in azimuth.

Each unit keeps every sixth native pulse, 20 pulses (13 in one short validation sector), giving 48,200 training and 593 validation pulses. The responses are restricted to our region of interest (Appendix A). Frequency selection retains source indices $0, 2, 4, \ldots$ plus the last endpoint, preserving 213–218 of the 424–434 native bins per pass, spanning approximately 9.288–9.910 GHz. Exact

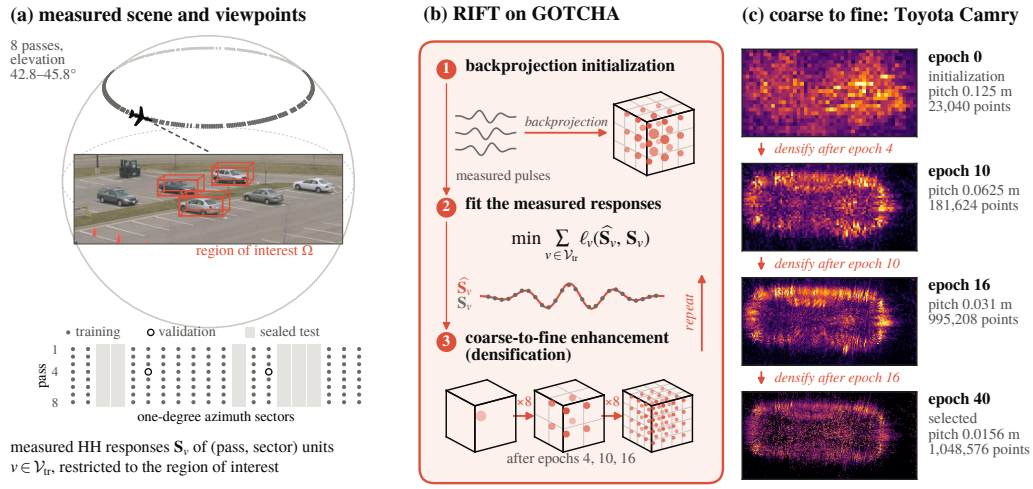

Figure 4: RIFT on GOTCHA. (a) A training view is a (pass, sector) unit: the HH responses of one one-degree azimuth sector of one of eight circular passes, restricted to our region of interest. The passes lie in a narrow elevation band, 42.8–45.8° over all pulses (sphere not to scale). Center: the parking lot (photo from Casteel et al. 2007, Fig. 2); the red-orange cubes, drawn schematically, mark the regions of interest Ω of the Toyota Camry, Nissan Sentra, and Hyundai Santa Fe. Inset: units in a 20° azimuth window; validation units (open) are 10% of the pass-4 units, and the 55 sealed test sectors are withheld on every pass. (b) Backprojection initialization of a coarse lattice from the first 100 (pass, sector) units in the seed-42 order, the fit of the measured responses (Eq. 3 with ℓ_v of Eq. D.2), and the coarse-to-fine enhancement (densification), after which the fit resumes (the GOTCHA paragraphs above). (c) Plan-view maximum-intensity projections (camera on $+z$, nose to the right, $|x| \leq 3$ m, $|y| \leq 1.5$ m) of the learned point energy from the one training run behind the Toyota Camry row, deposited at each stage's own lattice pitch: the backprojection initialization, the states just before the densifications after epochs 10 and 16, and the selected epoch-40 checkpoint. Brightness is the square root of the deposited energy over its 99.5th percentile in each panel (inferno, brighter = larger), not the min–max scale of the GOTCHA view figures.

selected frequencies, sensor geometry, and effective reference ranges are preserved; GOTCHA's HH autofocus correction is applied once when the data are loaded. There is no presumming. Fig. 5 shows the measured scene and the three regions of interest.

GOTCHA reference meshes. The GOTCHA data set supplies no 3D ground truth. The geometry reference of each car is a same-generation stand-in model from a public 3D repository, scaled separately along each axis to the published length, width without mirrors, and height of that generation; none is a scan of the car in GOTCHA. The stand-ins are a Toyota Camry XV20 (Nieve5677, 2025) scaled to $4.765 \times 1.785 \times 1.430$ m, a Nissan Sentra B15 (Lone Wolf, 2012) scaled to $4.508 \times 1.709 \times 1.410$ m, and a first-generation Hyundai Santa Fe (teenlin3, 2026) scaled to $4.500 \times 1.845 \times 1.730$ m, the height with roof rails.

Method budgets and checkpoint selection. The selected RIFT-dataset budgets are method-specific, not matched in updates, data passes, or compute time. RIFT's dataset-specific optimizer and controller are given in Appendix D.

On GOTCHA, RIFT, SpINR-style, and backprojection are run on each region of interest. RIFT uses 40 epochs of 2,410 single-unit updates (96,400 updates; Appendix D); SpINR-style has a 150-epoch budget over the 48,200 train pulses with 1,024 pulses per batch, or 48 updates per epoch; and backprojection is computed from the train units. RIFT and SpINR-style select the checkpoint with the lowest pooled full-native complex validation RelMSE. Training seed is 42 except for Radar Fields, which retains seed 0; split seed is 42 throughout.

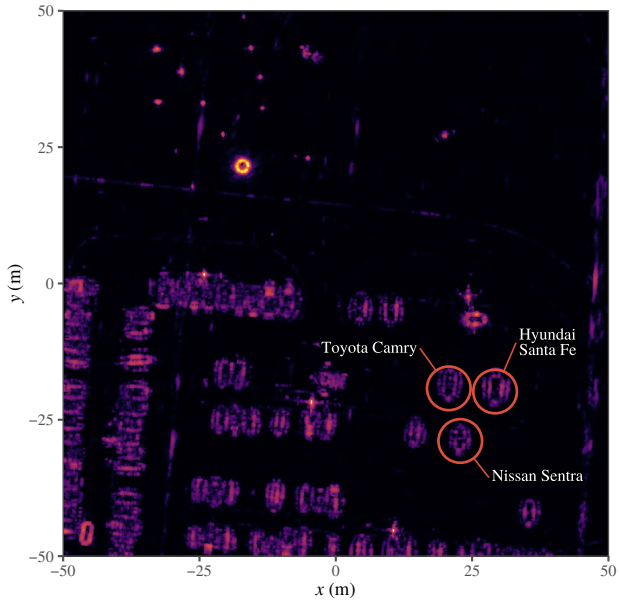

Figure 5: The measured GOTCHA scene: a 2D SAR image of the spotlighted scene center from one pass with 360° aperture, native frame (adapted from Casteel et al. 2007, Fig. 1; recolored to the inferno scale). Circles mark the Toyota Camry, Nissan Sentra, and Hyundai Santa Fe at the centers of their regions of interest (Appendix A).

Compute and backends. All experiments run on one Intel Data Center GPU Max 1100 per training run with PyTorch 2.12.1+xpu; SE Stage 1 uses CPU workers. Because tiny-cuda-nn and RadarSplat's rasterizer and fused-SSIM operations are CUDA-only, Radar Fields runs on our fp32 PyTorch reimplementation of tiny-cuda-nn, whose initialization and gradient reduction differ from the original backend, and RadarSplat runs on our PyTorch ports of its rasterizer and fused-SSIM operations, including their backward passes.

F Baseline Implementations and Adaptations

Source availability and implementation scope. We did not locate official implementations of SpINRv1, GeRaF 1.0, or Sugavanam–Ertin (SE), so we implemented them: SpINR-style follows the SpINRv1 paper with the local choices listed below, our GeRaF 1.0 version combines the released GeRaF stage-1 components with the settings of the v1 paper, and SE is an independent implementation of the paper's equations. Radar Fields and RadarSplat reuse their released model components, adapted to our acquisitions. Each paragraph below lists what we retain from the original method and what we change.

Spatial-complexity budget. We keep training-time spatial complexity comparable across methods: each uses a nominal scale of 48^3 samples, or about as many explicit primitives where its representation permits, with one configuration for all six RIFT-dataset scenes. Table 3 specifies what the budget controls for each representation. The one exception is SE's sparse stage, which uses 40^3 to limit computational cost. On GOTCHA, the larger region of interest gives a coarser physical sampling pitch.

SpINR-style. We retain a signed-real position-only field, product-distance spreading, selected-DFT-bin supervision, and magnitude squared error plus half-weighted complex squared error. The paper does not fully specify the network and optimizer; our choices are a six-layer width-840 Fourier-feature network, Adam, and gradient clipping. The selected budget uses 150 epochs and four RIFT-dataset views per update rather than 1,024; GOTCHA uses 1,024 pulses per update.

Table 3: Spatial budget of each method.

Method	Selected spatial budget
RIFT	48^3 initial points; adaptive growth up to 262,144 points
SpINR-style	48^3 midpoint integration samples of its continuous field
GeRaF	48^3 MF target lattice; continuous SDF/reflectivity networks
Radar Fields	Released hash-grid network; 48^3 geometry readout, not a parameter-count match
RadarSplat	112,000 Gaussians, versus 20,000 in the released configuration
Sugavanam–Ertin	40^3 sparse voxels per subaperture; continuous SDF and 48^3 surface readout

Its exact native point kernel retains reference-relative phase and physical inverse-distance-squared spreading.

GeRaF 1.0. We use the GeRaF-SENS release at commit `38266cb6`, retaining its stage-1 rendering/loss machinery. From the v1 paper we take the position-only learned reflectivity, scalar transmitted amplitude, SDF network dimensions, and squared MF-magnitude loss. This differs from the release example's fixed-power network, extra SDF features, and Charbonnier loss. Where the historical v1 settings are unavailable, we retain release defaults or example values: log transmitted amplitude zero, variance parameter 0.3, gradient-penalty weight 0.1, current/accumulated mask thresholds 0.04/0.15, and the absolute cosine learning-rate floor 5×10^{-4}. The executable released runner holds the model step at zero; we retain that behavior, so its step-dependent annealing and sharpness-unfreezing hooks do not advance. The SDF learning rate rises toward the absolute cosine floor. We retain the source positional encoding rather than substituting the paper's differently scaled expression.

Our acquisition geometry, frequencies, and phase references replace the original radar calibration. The selected MF lattice is 48^3 rather than the example's 601^3, and RIFT-dataset single-pair acquisition uses one signal bank. Validation selects the MF-magnitude checkpoint instead of the release's final iterate. Native targets use unit declared transmit power without a fitted radiometric initialization; the release's ADC-specific preprocessing does not apply to our data and is not used.

Sugavanam–Ertin (SE). We retain constrained complex sparse Fourier inversion per subaperture, magnitude aggregation, PCA normals, and the two-stage SDF construction. The paper does not identify a complete numerical solver/optimization recipe. We use SPGL1 basis-pursuit denoising with package defaults and a local residual budget of 1% of each subaperture's observed energy. Accepted solver exits provide the Stage-1 handoff. The SDF uses metric coordinates and nine Gaussian Fourier bands with standard deviation 2 cycles/m, concatenated with raw coordinates. Cloud extraction uses 15% of peak aggregated magnitude. SDF training uses Adam at 10^{-4}, equal weights for the six stated loss terms, and 2,048 samples per set. We interpret zero-set stopping as $|f| \leq 10^{-4}$ and evaluate the off-surface term on its stated background samples; undefined gradients or failed projections are rejected. The printed signed edge weighting is retained. We keep the paper's 0.3 m PCA radius, which is large relative to the 0.1 m RIFT-dataset objects. SE's signal metrics come from its Stage-1 sparse reconstruction, not from the final SDF.

SE initialization. A literal standard-Gaussian initialization assigns independent $\mathcal{N}(0, 1)$ weights and biases to the eight width-512 Softplus hidden layers and final tanh head. In our implementation this saturates the tanh output and eliminates usable gradients at initialization. We compared this recipe with $\mathcal{N}(0, \sigma^2)$ at $\sigma = 0.05$, holding the architecture, Fourier features, model seed, and sample locations fixed. The response-free probe uses seed 42, float32 on CPU, and 256 uniformly sampled points in the RIFT-dataset support. Table 4 reports the observed spatial gradients. With $\sigma = 1$, every sampled output rounded to $+1$ or -1, making the numerical tanh derivative $1 - f^2$ zero.

A separate 32-point backward probe of $\mathrm{mean}(f^2) + \mathrm{mean}[(\|\nabla_x f\|_2 - 1)^2]$ gave zero gradients for all 18 linear weight/bias tensors at $\sigma = 1$; at $\sigma = 0.05$, all 18 had finite nonzero gradients. No optimizer step or radar response was used. We therefore use $\sigma = 0.05$. The adapted initial values remain positive at all sampled points.

Table 4: SE initialization diagnostic. Counts refer to the 256 sampled points of the RIFT-dataset support; these are initialization checks, not reconstruction results.

Support	σ	Saturated outputs	Nonzero $\nabla_x f$	Mean $\|\nabla_x f\|_2$
RIFT dataset	1	256/256	0/256	0
RIFT dataset	0.05	0/256	256/256	0.115865

Radar Fields. Radar Fields retains the released network and core losses, with native matched-range power targets, an ROI-derived aperture and unit antenna response, omitted road/height priors and pose refinement, and a fixed-zero exterior with a finite-part KL treatment.

RadarSplat. RadarSplat retains the released-model initialization, optimizer groups, successful-path loss and fixed-step schedule, but uses transformed matched-filter power images, training-direction occupancy donors, adapted multipath transfer and sensor-filter inputs, and the Gaussian budget above. Its physical Gaussian scale and occupancy thresholds depend on the adapted image grid/normalization. Radar Fields and RadarSplat both run on our PyTorch ports described in Appendix E.

G Additional Dataset Results and Visualizations

G.1 RIFT dataset: per-object results and views

The RIFT dataset covers six objects: B787, A320, X-59, Fire truck, Race car, and Loader. Tables 5 and 6 and the three views in Figs. 8–10 share one method order (RIFT (Ours), SpINR-style, Radar Fields, Backprojection, GeRaF, RadarSplat, Sugavanam–Ertin, and Sugavanam–Ertin with Stage 1 only) and the same six object columns, so every table cell maps to the same object–method slot in all three views. The two Sugavanam–Ertin rows share the Stage-1 signal prediction, which Table 5 lists once, and differ in geometry: the published two-stage method reports its Stage-2 surface, while Stage 1 only reports its Stage-1 scattering field. The six-object means of these values are the entries of Table 1. Every entry follows the selected RIFT-dataset protocol of Appendix E: one checkpoint per method and object, selected without test responses or reference geometry, supplies the signal scores, the geometry scores, and all three views, and signal scores use the 1,000 reserved-test views of each object. Fig. 6 maps the per-view complex RelMSE of the four methods with a complex prediction over the train and validation views of each object, and Fig. 7 maps the per-view common matched-range power RelMSE of the six methods with a range-power prediction.

View protocol. Each method panel is the maximum-intensity projection, along the camera axis, of the method's declared native 3D support at its selected checkpoint: the common 48^3 evaluator-lattice readout over the reconstruction cube, densified $4\times$ by Plenoxel-style trilinear interpolation to 192^3, converted to magnitude (the square root of interpolated energy for energy readouts), min–max normalized per method, and shown above the fixed threshold $t = 0.20$ that also defines the geometry metrics (black below t). The native readouts are RIFT's point-SH energy deposited by conservative cloud-in-cell weights, the SpINR-style signed field σ (energy σ^2), GeRaF's NeuS logistic SDF density, the Radar Fields grid readout, the backprojection's matched-filter image energy $|A|^2$, RadarSplat's Gaussian occupancy-union support, and the Sugavanam–Ertin Stage-2 zero surface deposited onto the lattice, and for Sugavanam–Ertin with Stage 1 only its Stage-1 scattering field (Table 6, note ¶). Cameras are orthographic and axis-aligned in each object's registered scene frame. Because the asset frames of A320 and X-59 are oriented differently from those of the other four objects, the camera axes are chosen per object so that each figure shows the same look for every object: a front view (Fig. 8), a side view (Fig. 9), and a plan view (Fig. 10); the captions list the cameras. Within an object, every panel, including the reference mesh, shows the same ± 7.5 cm crop about the scene center with the same orientation, and no panel is aligned independently. The top row of each figure shows the ground truth: the registered reference mesh against which Table 6 is scored, drawn depth-shaded in neutral gray above the method panels of the same object. All method panels share one colormap and one color bar, but because brightness is normalized per method, equal colors do not mean equal physical magnitude across methods. N/A marks a structurally unavailable output. These views show radar-visible scattering support, not complete occupancy or a watertight shape; independent scene fits test consistency across configurations, not zero-shot cross-scene generalization.

Table 5: Per-object signal results on the 1,000 reserved-test views of each object, in the object–method slots of Figs. 8–10. Complex RelMSE is scored on the native complex response; common matched-range power is the whole-ROI normalized-dB observable of Table 1. The same selected checkpoint per method supplies Table 6. Bold marks the best populated value per object and block; the RadarSplat range-power entry uses a different bin set and is not ranked.

Method	B787	A320	X-59	Fire truck	Race car	Loader
Native complex RelMSE ↓[†]						
RIFT (Ours)[a]	**0.3361%**	**0.3168%**	**0.4045%**	**0.3746%**	**0.4003%**	**0.5130%**
SpINR-style[b]	99.63%	109.6%	31.36%	91.91%	74.76%	82.93%
Radar Fields	N/A	N/A	N/A	N/A	N/A	N/A
Backprojection	N/A	N/A	N/A	N/A	N/A	N/A
GeRaF[c]	137.9%	110.5%	109.8%	107.7%	149.8%	132.2%
RadarSplat	N/A	N/A	N/A	N/A	N/A	N/A
Sugavanam–Ertin[d]	19.98%	69.80%	63.74%	50.57%	22.00%	50.62%
Common matched-range power RelMSE ↓[‡]						
RIFT (Ours)[a]	**0.6216%**	**0.6275%**	**0.7638%**	**0.3857%**	**0.4770%**	**0.3038%**
SpINR-style[b]	9.032%	8.476%	7.374%	12.18%	10.82%	7.448%
Radar Fields[e]	30.08%	29.44%	33.60%	33.18%	30.61%	35.60%
Backprojection	N/A	N/A	N/A	N/A	N/A	N/A
GeRaF[c]	28.32%	45.43%	55.03%	39.39%	16.46%	14.14%
RadarSplat[f]	74.69%[§]	62.58%[§]	66.69%[§]	67.21%[§]	68.01%[§]	63.23%[§]
Sugavanam–Ertin[d]	5.904%	7.338%	7.903%	8.428%	6.409%	5.138%
Train-mean constant	24.84%	25.50%	29.68%	22.10%	20.29%	12.76%
on RadarSplat bins[§]	20.28%	21.15%	24.73%	18.09%	16.95%	11.59%

N/A denotes a structurally unavailable observable. These entries are single-object reserved-test scores; their six-object means are in Table 1.

[†] Predicting zero everywhere scores 100% complex RelMSE by construction.
[‡] Matched-range power is mapped to the release's normalized-dB intensity with train-only statistics (train peak power, 60 dB span, clipped to $[0, 1]$) and scored over all 13 ROI range bins of every view; predicting zero power scores 100%. The train-mean constant rows predict the mean train intensity at every bin, and every value in this block is read against them.
[a] RIFT: checkpoint at minimum validation complex data loss (epochs 111–132 of 150); its range power is computed from its own predicted complex response.
[b] SpINR-style: production fit on the 2,400 train views (150-epoch budget), checkpoint at minimum complex validation RelMSE (epoch 10; X-59: epoch 5). A complex RelMSE above 100% is worse than predicting zero.
[c] GeRaF: trained on matched-filter amplitude, not the complex signal; checkpoint at minimum pooled validation MF-magnitude error (steps 23,000–47,000 of 50,000).
[d] Sugavanam–Ertin: signal from the Stage-1 per-subaperture SPGL1 solution (72 subapertures); its geometry in Table 6 is the Stage-2 surface.
[e] Radar Fields: produces power only; checkpoint at minimum validation power RelMSE (step 240).
[f] RadarSplat: released log-power intensity mapping, final iterate (2,000 updates). [§] Its range–azimuth power image covers 6 of the 13 ROI range bins of each view, so this entry is scored on those bins only (6,000 of 13,000 elements per object) and is read against the train-mean row on the same bins, not ranked against whole-ROI entries.

B787 native reduced-observable diagnostics (separate domains).

Method	Native observable	Validation	Reserved TEST
GeRaF	Matched-filter amplitude RelMSE	65.23%	70.14%
RadarSplat	Elevation-integrated 2-D power RelMSE	73.17%	72.10%

These native observables are not relabeled as the common matched-range power domain above.

Table 6: Per-object 3D reconstruction results in the object–method slots of Figs. 8–10. Fixed threshold $t = 0.20$ on the common 48^3 evaluator lattice after per-method min–max normalization. Chamfer, maximum Hausdorff, and HD95 are measured between the supra-threshold lattice centers and 20,000 surface samples of the registered mesh; solid IoU uses 5 mm cells; F1 uses tolerance $\tau = 6.25$ mm (one lattice pitch). Bold marks the best populated value per object and metric.

Method	B787	A320	X-59	Fire truck	Race car	Loader
Symmetric squared Chamfer distance (10^{-4} m^2) ↓						
RIFT (Ours)	**0.4385**	**0.4203**	**0.5445**	**0.5685**	0.3836	**0.5321**
SpINR-style	0.9704	1.164	0.9832	0.8898	**0.2802**	0.6195
Radar Fields	268.7	235.8	238.7	229.2	261.4	192.6
Backprojection	1.930	2.519	1.533	2.351	0.8669	2.185
GeRaF	77.90	82.31	106.1	72.49	129.3	78.08
RadarSplat	69.30	53.26	60.80	70.08	78.60	70.32
Sugavanam–Ertin	86.55	72.73	115.8	120.8	100.4	91.38
Sugavanam–Ertin, Stage 1 only[¶]	0.6223	3.882	7.194	1.529	0.4214	1.082
Maximum Hausdorff distance (mm) ↓						
RIFT (Ours)	**16.04**	**16.81**	**14.88**	**19.41**	14.43	**18.71**
SpINR-style	18.00	28.30	21.02	40.73	**12.70**	26.71
Radar Fields	234.4	233.0	230.9	211.7	218.7	218.2
Backprojection	50.88	86.16	30.24	101.5	26.25	79.76
GeRaF	183.7	183.1	207.4	167.6	218.3	193.3
RadarSplat	178.9	175.6	175.9	178.1	188.6	174.6
Sugavanam–Ertin	183.8	177.8	187.5	206.6	211.6	195.0
Sugavanam–Ertin, Stage 1 only[¶]	22.85	183.6	183.6	176.6	16.37	113.7
HD95 (mm) ↓						
RIFT (Ours)	**11.38**	**10.51**	**12.89**	13.85	9.804	13.28
SpINR-style	16.02	16.79	15.60	**11.15**	**6.979**	**10.83**
Radar Fields	184.1	185.3	184.1	170.1	178.0	169.0
Backprojection	23.64	24.19	21.03	27.70	16.90	26.70
GeRaF	148.1	146.4	163.9	135.7	168.9	143.5
RadarSplat	131.5	119.1	130.6	149.7	136.7	131.4
Sugavanam–Ertin	145.5	142.7	155.1	162.3	147.7	154.5
Sugavanam–Ertin, Stage 1 only[¶]	11.46	13.82	20.19	12.96	11.82	16.06
Solid intersection over union (IoU) ↑						
RIFT (Ours)	**0.2260**	**0.2381**	**0.1847**	**0.2948**	0.3040	**0.3146**
SpINR-style	0.07536	0.1178	0.02500	0.2078	0.2983	0.2649
Radar Fields	–	–	–	–	–	1.526×10^{-5}
Backprojection	0.09102	0.08965	0.08951	0.1988	0.2416	0.2292
GeRaF	0.01133	0.007040	0.003314	0.03756	0.006633	0.01821
RadarSplat	0.004235	0.003245	0.0007809	0.01301	0.01677	0.01674
Sugavanam–Ertin	–	0.01041	–	–	–	0.02959
Sugavanam–Ertin, Stage 1 only[¶]	0.1715	0.1780	0.1100	0.2641	**0.3433**	0.2871
F1-score ↑						
RIFT (Ours)	**0.8686**	**0.8687**	**0.7749**	0.7837	0.8692	**0.7968**
SpINR-style	0.6014	0.6651	0.4929	0.5624	**0.9177**	0.7327
Radar Fields	–	–	–	–	–	0.0001545
Backprojection	0.4788	0.4428	0.4368	0.5818	0.7188	0.6004
GeRaF	0.06881	0.04311	0.02260	0.08721	0.01961	0.04859
RadarSplat	0.03155	0.02351	0.01834	0.08546	0.05391	0.05704
Sugavanam–Ertin	–	0.06858	–	0.003417	–	0.07239
Sugavanam–Ertin, Stage 1 only[¶]	0.7353	0.7746	0.5529	**0.8003**	0.8189	0.7045

The same selected checkpoint supplies these entries, the signal entries in Table 5, and the three views. N/A cells are not ranked. A dash (–) marks an IoU or F1 of exactly zero, meaning that in that run the method produced no successful 3D reconstruction of the object: for IoU, its supra-threshold support has no solid-cell overlap with the reference; for F1, no supra-threshold lattice center lies within τ of the reference surface. SpINR-style is the production fit of Table 5; Sugavanam–Ertin uses its Stage-2 zero surface; GeRaF uses its SDF density; Radar Fields uses its grid readout; and RadarSplat uses its Gaussian occupancy-union support. Backprojection is the matched-filter backprojection of the 2,400 train views of each object (phase-only matched filter over all 600 frequencies of the view's single transmit–receive pair), accumulated directly on the evaluator lattice; its energy $|A|^2$ is read out like the other energy readouts. It uses no validation or test responses and has no signal prediction. [¶] Sugavanam–Ertin with Stage 1 only: the same Stage-1 solution as the row above, whose geometry is read from its Stage-1 scattering field instead of the Stage-2 surface, under the same fixed readout. The field is the magnitude sum $\sum_m |S_m|$ over its 72 sub-apertures (the aggregate on which the published method thresholds its Stage-2 point cloud), on its own cell-centred 40^3 grid over the same cube, squared to energy and trilinearly resampled onto the evaluator lattice.

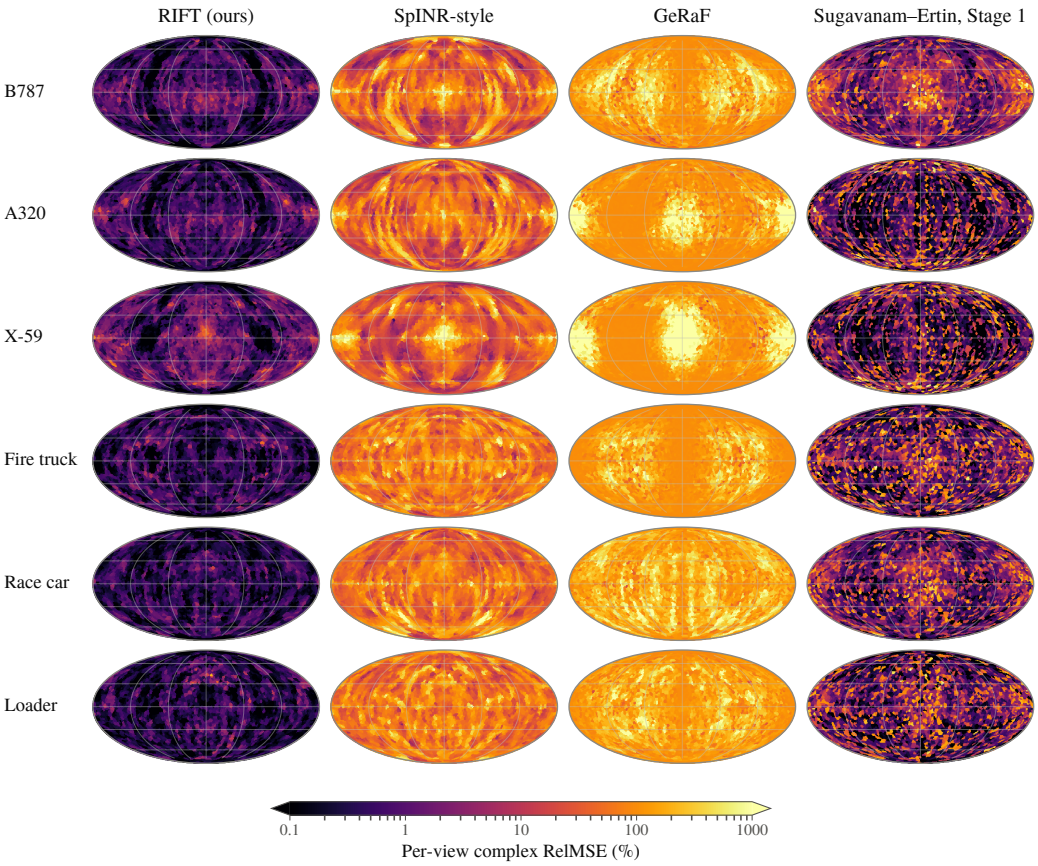

Figure 6: Per-view complex RelMSE on the view sphere of each object in the RIFT dataset, for the four methods with a complex prediction, at the checkpoints of Tables 1 and 5. Each of the 2,400 train and 1,000 validation views of an object is a radar viewpoint on a sphere of radius 10 m around it; its value is $\sum_f |\hat{S}(f) - S(f)|^2 / \sum_f |S(f)|^2$ over the 600 frequencies of its response, in percent. Each map is a Mollweide equal-area projection of the whole sphere seen from outside: longitude is azimuth about the object's up axis, measured from its nose at the map center, and latitude is elevation; grid lines are every 60° in longitude and 30° in latitude. Every point takes the value of its nearest train or validation view; reserved-test views are not evaluated. One logarithmic inferno scale is shared by all panels (darker is better; values outside 0.1–1000% take the end colors). The Sugavanam–Ertin Stage-1 solver fits each subaperture to a residual budget of 1% of its observed energy (Appendix F), so its pooled train RelMSE is 1.00% on every object while its pooled validation RelMSE is 18–69% across the six objects, which gives its maps their speckle.

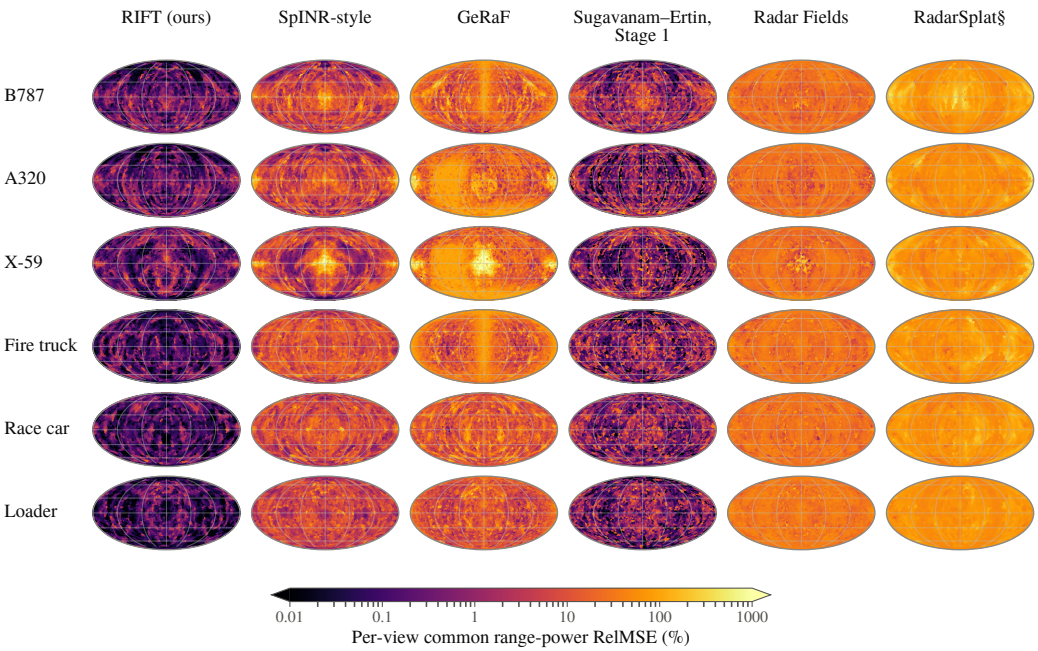

Figure 7: Per-view common matched-range power RelMSE on the view sphere of each object in the RIFT dataset, for the six methods with a range-power prediction, at the checkpoints of Tables 1 and 5; layout, projection, grid, and nearest-view fill are those of Fig. 6. The value of each train and validation view is the RelMSE of its normalized-dB range-power profile over its 13 ROI range bins, as in the second block of Table 5 (train-only peak power, 60 dB span); RIFT, SpINR-style, GeRaF, and Sugavanam–Ertin are scored through the matched-range power of their complex responses. §: RadarSplat is scored on the ROI bins inside its image's range crop. One logarithmic inferno scale is shared by all panels (darker is better; values outside 0.01–1000% take the end colors); its floor is a decade below that of Fig. 6 because 43% of RIFT's views score below 0.1%. Sugavanam–Ertin's speckle has the cause given there.

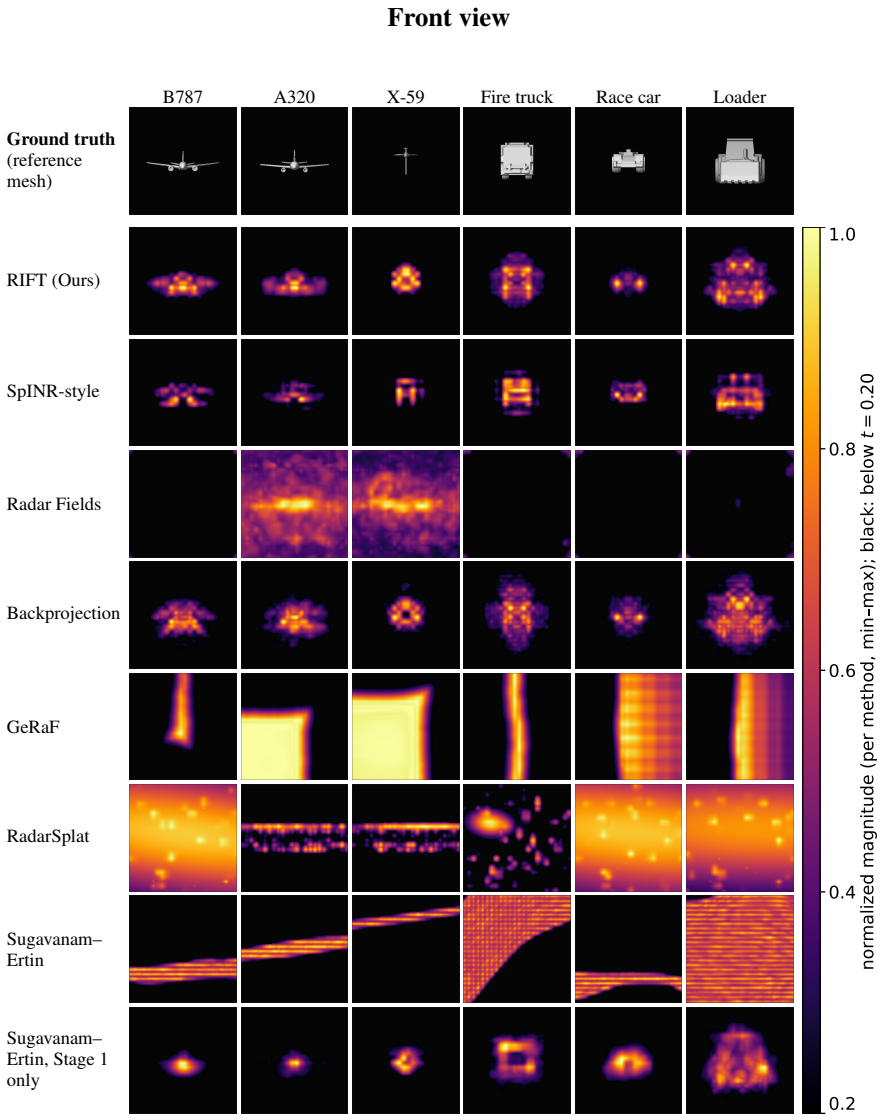

Figure 8: RIFT dataset, front view: orthographic cameras on the $-x$ axis for A320 and X-59 (image right $-y$, up $+z$) and on the $+z$ axis for B787, Fire truck, Race car, and Loader (image right $+x$, up $+y$), whose asset frames are oriented differently. Columns are objects. The top row shows each object's ground truth, its registered reference mesh, from the same camera, crop, and orientation as the method panels below it; it is reference only, never a method output. Each row below it is one method; each panel is the maximum-intensity projection of the method's native 3D support above the fixed threshold $t = 0.20$, at the checkpoint that supplies its entries in Tables 5 and 6. Color: per-method min–max normalized magnitude on one shared scale (black below t); brightness is not comparable across methods.

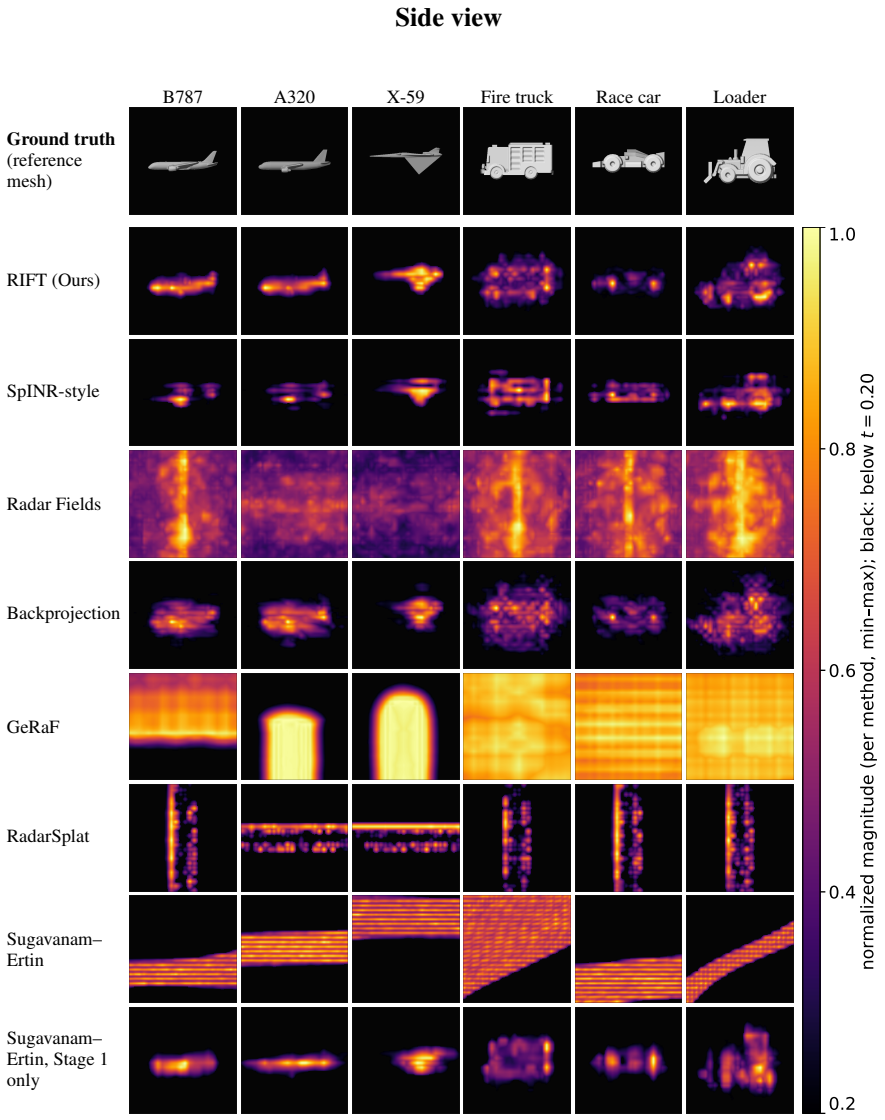

Figure 9: RIFT dataset, side view: cameras on $-y$ for A320 and X-59 (image right $+x$, up $+z$) and on $+x$, rotated 90° counter-clockwise, for the other four objects (image right $-z$, up $+y$), under the view protocol and layout of Fig. 8: object columns, with the ground-truth reference mesh from the same camera in the top row and one method per row below it.

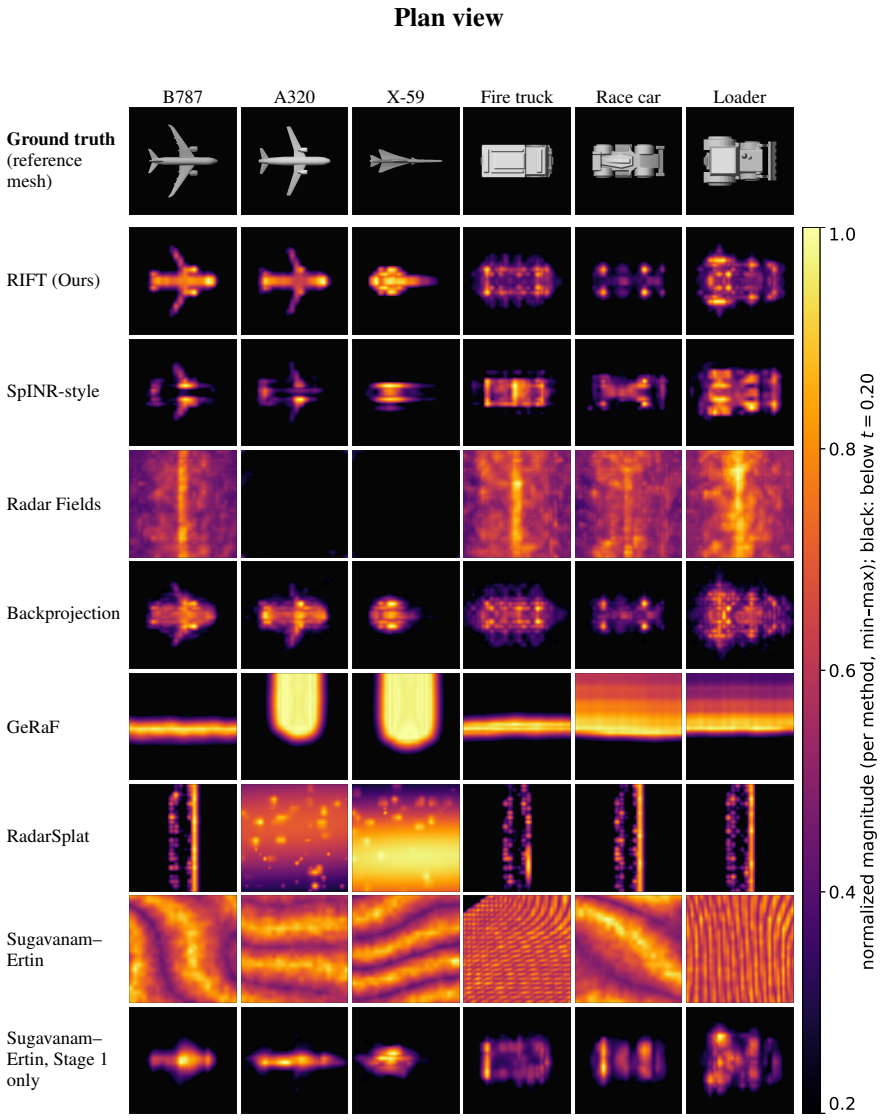

Figure 10: RIFT dataset, plan view: cameras on $-z$ for A320 and X-59 (image right $-x$, up $+y$) and on $+y$ for the other four objects (image right $+z$, up $+x$, so that every object lies nose-right), under the view protocol and layout of Fig. 8: object columns, with the ground-truth reference mesh from the same camera in the top row and one method per row below it.

GOTCHA visualization. Tables 7 and 8 report the per-vehicle GOTCHA train and validation error and the 3D reconstruction metrics for the Toyota Camry, Nissan Sentra, and Hyundai Santa Fe, and Figs. 11–13 show front, side, and plan views of the same checkpoints under the view protocol of Figs. 8–10.

Table 7: Per-vehicle signal results on GOTCHA (HH), in the vehicle–method slots of Figs. 11–13. Pooled full-native complex RelMSE in percent over every train and validation (pass, sector) unit of each vehicle's full split (2,410 train and 30 validation units); predicting zero scores 100%. Validation holds out pass-4 units between neighbouring passes and also selects the checkpoint. Each learned row uses its checkpoint with the lowest validation RelMSE within its training budget, 40 epochs for RIFT (Camry: RIFT at epoch 40 of 40 epochs and SpINR-style at epoch 5 of 150 epochs; Sentra and Santa Fe: RIFT at epochs 24 and 25 of 40 and SpINR-style at epoch 5 of 150), which also supplies Table 8. Backprojection has no signal prediction. Bold marks the best populated value per vehicle and block.

Method	Toyota Camry	Nissan Sentra	Hyundai Santa Fe
Train full-native complex RelMSE ↓			
RIFT (Ours)	**41.19%**	**47.34%**	**35.00%**
SpINR-style	108.7%	107.3%	105.4%
Backprojection	N/A	N/A	N/A
Validation full-native complex RelMSE ↓			
RIFT (Ours)	**51.41%**	**57.95%**	**53.62%**
SpINR-style	127.3%	122.1%	136.0%
Backprojection	N/A	N/A	N/A

Table 8: Per-vehicle 3D reconstruction results on GOTCHA (HH), in the vehicle–method slots of Figs. 11–13, from the checkpoints of Table 7. Geometry follows Table 6 at the vehicle scale: fixed threshold $t = 0.20$ on the 48^3 lattice over each vehicle's 6 m region cube, with F1 at $\tau = 0.125$ m (one lattice pitch) and solid IoU on 0.1 m cells. Each reference is a same-generation stand-in mesh, not a scan of the vehicle (Camry XV20 (Nieve5677, 2025), Sentra B15 (Lone Wolf, 2012), Santa Fe first generation (teenlin3, 2026)), registered to the measured data by one rigid transform shared by every method: the pose that maximizes the normalized magnitude of the model-free train backprojection on the mesh surface, with the height kept at the calibrated ground (Camry: heading $-3.0°$ and horizontal shift $(-0.425, +0.175)$ m within $\pm 6°$; Sentra: $+3.25°$ and $(-0.125, -0.100)$ m; Santa Fe: $-6.00°$ and $(-0.025, -0.200)$ m, both within $\pm 15°$ because their start is a radar-derived placement). This placement is a train-only radar locate estimate; applied to the Camry, it lies 0.28 m and $4°$ from the Camry's survey-based placement. For the Sentra and Santa Fe, the stand-in's front–back orientation is also chosen by this objective. The references have no ground, so lattice points and reference samples within one lattice pitch above the ground plane or below it, where car–ground multipath lies, are excluded on both sides. Against the survey-placed Camry mesh without this exclusion, every method scores lower (RIFT F1 0.475). Bold marks the best populated value per vehicle and metric.

Method	Toyota Camry	Nissan Sentra	Hyundai Santa Fe
Symmetric squared Chamfer distance (m^2) ↓			
RIFT (Ours)	**0.05536**	**0.06411**	**0.06814**
SpINR-style	2.594	1.898	1.584
Backprojection	1.135	1.212	1.004
Maximum Hausdorff distance (m) ↓			
RIFT (Ours)	**1.336**	**0.7251**	**0.6913**
SpINR-style	3.893	3.538	3.248
Backprojection	3.460	3.295	3.059
HD95 (m) ↓			
RIFT (Ours)	**0.4001**	**0.3998**	**0.4469**
SpINR-style	2.758	2.644	2.412
Backprojection	2.715	2.801	2.443
Solid IoU ↑			
RIFT (Ours)	**0.1328**	**0.08218**	0.05384
SpINR-style	0.06138	0.07034	0.07543
Backprojection	0.1018	0.08002	**0.1035**
F-score ↑			
RIFT (Ours)	**0.6318**	**0.5098**	**0.5123**
SpINR-style	0.1736	0.2491	0.2400
Backprojection	0.4715	0.4331	0.4203

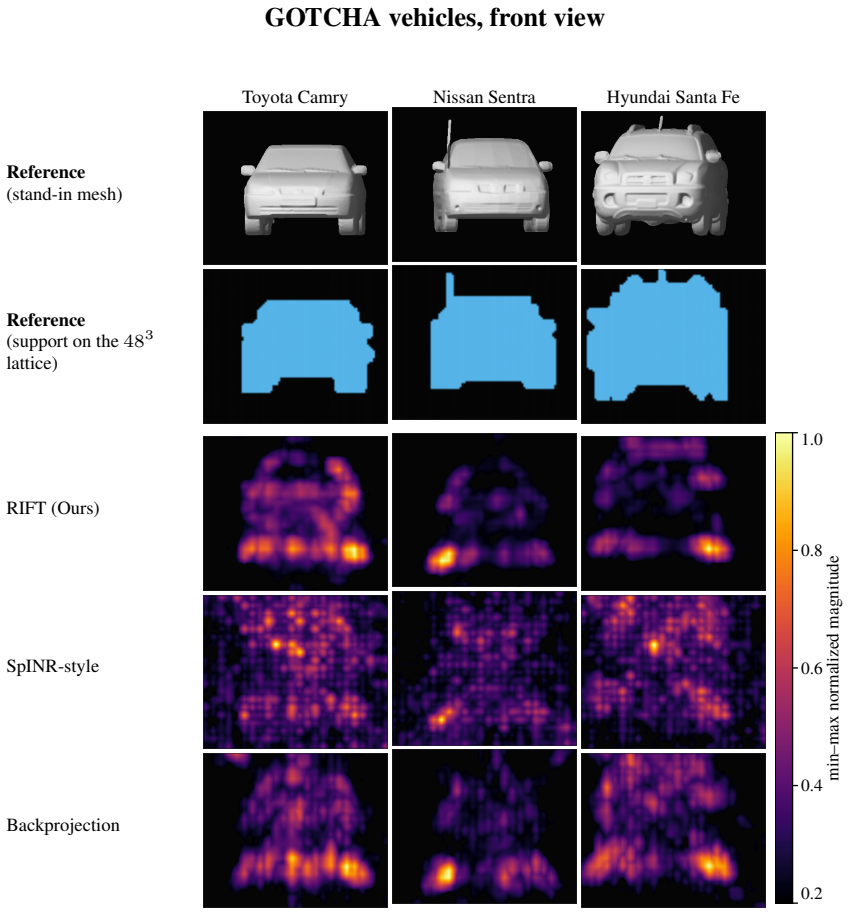

Figure 11: GOTCHA vehicles (HH), front view, one column per vehicle (Toyota Camry, Nissan Sentra, Hyundai Santa Fe): an orthographic camera on the $+x$ axis of each car-aligned scene frame (x along the car, nose at $+x$; image right $+y$, up $+z$), showing $|y| \leq 1.5$ m and $|z| \leq 1.25$ m. The layout is that of Fig. 8. The top two rows are the geometry reference, not ground truth: each vehicle's same-generation stand-in mesh (Camry XV20 (Nieve5677, 2025), Sentra B15 (Lone Wolf, 2012), Santa Fe first generation (teenlin3, 2026)) at its data-registered pose from the same camera and crop, and its support on the 48^3 evaluator lattice (12.5 cm pitch over the 6 m region cube): the cells that contain the reference surface, projected along the camera axis and drawn in one uniform color with no magnitude, since the reference carries no radar reflectivity. Each row below it is one method; each panel is the maximum-intensity projection of the method's 3D field on the same lattice above the fixed threshold $t = 0.20$, trilinearly upsampled $4\times$ for display. Every method uses the 2,410 train (pass, sector) units of the full split of each vehicle's box region. For the Camry, each learned row uses the validation-selected checkpoint that supplies its entries in Tables 7 and 8: RIFT at epoch 40 of 40 epochs and SpINR-style at epoch 5 of 150 epochs. Backprojection is $A^{\mathsf{H}}y$ over the measured train units. Color: per-method min–max normalized magnitude on one shared scale (black below t); brightness is not comparable across methods.

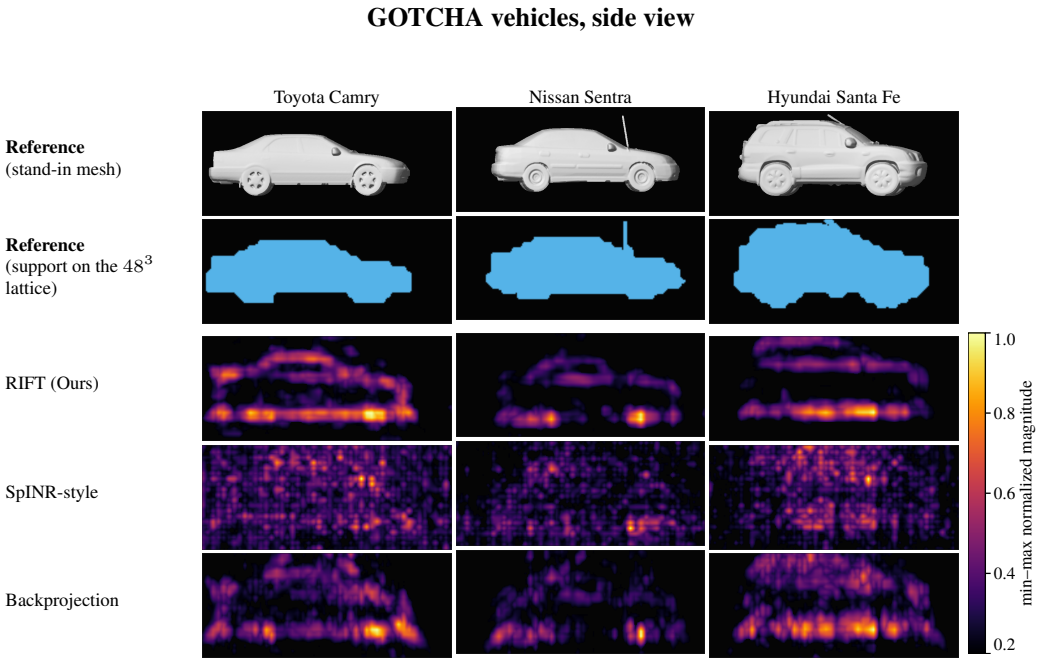

Figure 12: GOTCHA vehicles (HH), side view: a camera on $-y$ (image right $+x$, up $+z$, nose to the right), showing $|x| \leq 3$ m and $|z| \leq 1.25$ m, under the protocol and layout of Fig. 11: two reference rows (stand-in mesh and its lattice support) above one method per row, from the same checkpoints.

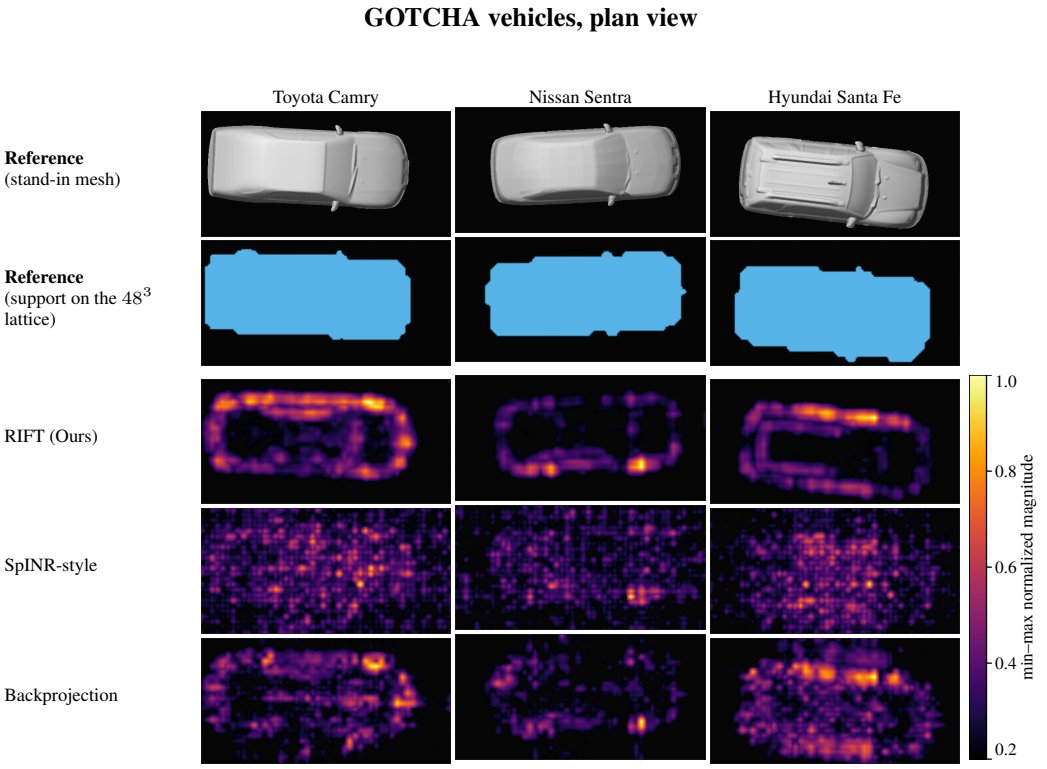

Figure 13: GOTCHA vehicles (HH), plan view: a camera on $+z$ (image right $+x$, up $+y$, nose to the right), showing $|x| \leq 3$ m and $|y| \leq 1.5$ m, under the protocol and layout of Fig. 11: two reference rows (stand-in mesh and its lattice support) above one method per row, from the same checkpoints.