Radar2Shape: 3D Shape Reconstruction from High-Frequency Radar using Multiresolution Signed Distance Functions

Neel Sortur[1]   Justin Goodwin[2]   Purvik Patel[1]   Luis Enrique Martinez Jr[2]
Tzofi Klinghoffer[3]   Rajmonda S. Caceres[2]   Robin Walters[1]

[1]Northeastern University   [2]MIT Lincoln Laboratory   [3]Massachusetts Institute of Technology
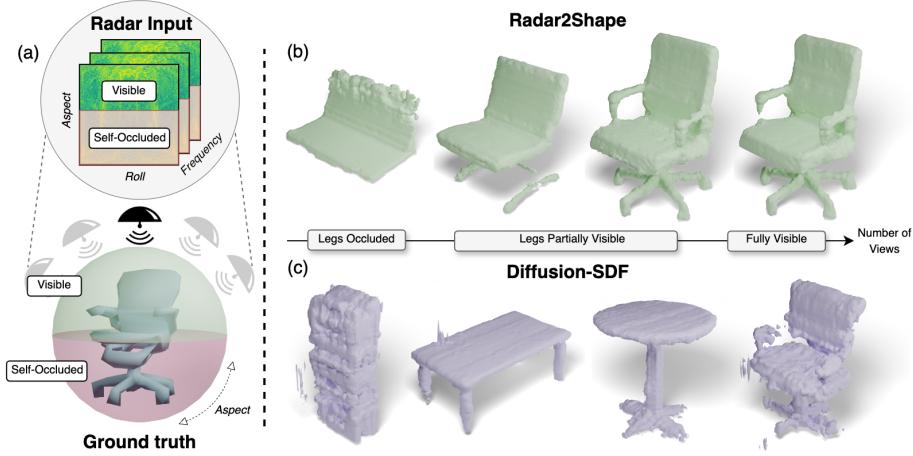
Figure 1. **Overview.** Radar2Shape solves the challenging task of 3D shape reconstruction from radar captured at limiting viewing angles. **(a)** Limited views cause self-occlusion, resulting in missing information in the measurement. **(b)** Our approach overcomes this ambiguity by using a data-driven diffusion prior with a novel coarse-to-fine refinement technique in signed distance function space. This method accurately generates occluded geometries based on partial radar measurements, leading to better performance than **(c)** existing domain-adapted methods that can fail with limited views and struggle even in full observability.

Abstract

Determining the shape of 3D objects from high-frequency radar signals is analytically complex but critical for commercial and aerospace applications. Previous deep learning methods have been applied to radar modeling; however, they often fail to represent arbitrary shapes or have difficulty with real-world radar signals which are collected over limited viewing angles. Existing methods in optical 3D reconstruction can generate arbitrary shapes from limited camera views, but struggle when they naively treat the radar signal as a camera view. In this work, we present Radar2Shape, a denoising diffusion model that handles a partially observable radar signal for 3D reconstruction by correlating its frequencies with multiresolution shape features. Our method consists of a two-stage approach: first, Radar2Shape learns a regularized latent space with hierarchical resolutions of shape features, and second, it diffuses into this latent space by conditioning on the frequencies of the radar signal in an analogous coarse-to-fine manner. We demonstrate that Radar2Shape can successfully reconstruct arbitrary 3D shapes even from partially-observed radar signals, and we show robust generalization to two different simulation methods and real-world data. Additionally, we release two synthetic benchmark datasets to encourage future research in the high-frequency radar domain so that models like Radar2Shape can safely be adapted into real-world radar systems.

1. Introduction

Radar is a reliable sensing mechanism in adverse light and weather conditions with wide-ranging applications such as robotics [3], autonomous driving [6], and remote sensing [5]. It operates by transmitting radio waves and analyzing the response, echoes that return after striking an object. Types of radar are typically distinguished by their wavelength – longer wavelengths struggle to detect small objects, like raindrops or geologic particles, while shorter wavelengths (high-frequency radar) can detect a variety of sizes, but may be more noisy. In both regimes, geometrically characterizing and reconstructing an object from its radar response still presents a challenging inverse learning task. At long ranges, radar signals are often noisy and provide poor resolution [26]. Furthermore, radar sensors often do not fully observe an object at all viewing angles. This results in *partial observability* in the radar response that introduces uncertainties in the reconstruction process. In this work, we tackle this difficult problem of object reconstruction from long-range, high-frequency radar responses that are partially observed.

Previous deep learning approaches to long-range high-frequency radar modeling have focused primarily on extracting high-level features for classification, segmentation, or pose estimation tasks [41, 42], which are still difficult open problems. However, many downstream tasks require full-shape reconstruction of an observed object, a higher dimensional and even more challenging problem. The computer vision community has separately developed models for 3D shape reconstruction – these models are typically conditioned on partial point clouds or multi-view images, and they often try to estimate camera intrinsic and extrinsic parameters. However, there are unique challenges when conditioning on a radar signal for 3D reconstruction instead of multi-view images. First, radar lacks analogous camera parameters to estimate because the observed shape does not correspond to simple geometric projections of the radar response. Second, the partial observability introduces high uncertainty. Much more of the object may be occluded from a radar's line of sight compared to a camera's single view. Additionally, the radar signal is spread across multiple frequencies, which can correspond to different resolutions of the object's geometry. Previous methods have not focused on the difficult problem of full-shape reconstruction from long-range high-frequency radar signals or taken advantage of individual frequencies in the radar response.

We propose Radar2Shape, a method that can reconstruct full 3D shapes from high-frequency radar responses by associating radar frequencies with shape resolutions. Our approach consists of two stages: 1) learning a multi-resolution, hierarchical latent space for 3D shapes, and 2) training a diffusion model to denoise in this space by conditioning on radar responses. The first stage uses a series

of encodings and a VAE to learn a regularized latent space of vectors defining signed distance functions (SDFs) [12]. Instead of representing a shape with a single latent vector, we separate the latent vector into components that represent shape features at multiple resolutions (e.g. the thin structures in a shape versus its overall structure). This representation is created by projecting multi-resolution point cloud features onto triplanes of various spatial resolutions. These are then processed separately and combined as input into an SDF network. The second stage uses a Transformer backbone to predict the denoised sequence of latent shape vectors, iteratively from coarsest to finest resolution, conditioned with attention on embeddings of the corresponding radar resolution. Additionally, we incorporate a domain-relevant 2D shape prior [41] and propose a more efficient version of our method for this lower dimensional shape space by 1) encoding the shape space as a projection of our 2D shape parameterization and 2) using a U-Net to jointly encode the radar response and predict denoised latent shape vectors. Overall, we make four primary contributions:

- We present Radar2Shape, a novel denoising diffusion model that reconstructs an object's 3D geometry from partially-observable, high-frequency radar observations.
- We show superior results compared to many 3D reconstruction models adapted to the radar domain and an existing competitive radar baseline.
- We demonstrate a general method for learning multiresolution signed distance functions of 3D geometries.
- We introduce the *Manifold40-PO* and *Manifold40-PO-SBR* benchmark datasets, the first public datasets of diverse meshes and simulated high-frequency radar responses for radar-based single object reconstruction.

2. Related Work

2.1. Diffusion Models

Diffusion models [20] have emerged as a powerful generative model applicable in many scientific domains ranging from bioinformatics [18] to climate science [4, 30]. Alternative methods for generative modeling include Generative Adversarial Networks (GANs) [51], but diffusion has been shown to outperform GANs [14, 48]. Flow Matching [33] has also recently emerged as an alternative, but diffusion's demonstrated versatility across domains and its robustness to noisy inputs [54] motivates its use in this work.

2.2. 3D Representations and Reconstruction

3D reconstruction is a long-standing task in computer graphics and computer vision, leading to the development of many deep learning methods that take as input 2D shape projections (e.g., images, radar) and reconstruct the 3D geometries. In many use cases, entire scenes, consisting of geometries, lighting, transparency, density, and textures, must

be modeled. Techniques like Gaussian Splatting [25] and Neural Radiance Fields [27, 38] (NeRF) excel at modeling these high dimensional structures, but they are optimized for rendering, and extracting meshes from these representations is not straightforward. Instead, much research has focused on triplane features, point clouds, meshes, voxel, or signed distance function (SDF) representations for meshes [1, 9, 12, 17, 23, 41, 46, 50, 55].

Among these representations, Deep SDFs [47] and subsequent improvements have gained popularity due to their efficiency and small memory footprint. Two-stage Diffusion [20] approaches have demonstrated success in generating these SDFs [12, 15, 57, 62]. Multiresolution hash encodings have improved performance of SDF-based representations by projecting the coordinates of query points to a higher dimensional, spatially-aware feature [40]. These models are typically conditioned on 2D images or partial point clouds, and some use Octree-based structures to generate hierarchical features. However, none of these works are conditioned on radar observations, and the Octree's hierarchical features require manual part segmentation of the geometries as training labels. In this work, we use SDFs and learn hierarchical features without part segmentation labels by utilizing the multiresolution hash encoding in a novel way.

2.3. Radar Modeling with Deep Learning

Many existing deep learning methods for radar modeling have used tools like NeRF and Gaussian splatting to extract geometric representations from autonomous driving data [7, 29], but they require per-scene optimization during inference. There are existing methods that do not require per-scene optimization [16, 39, 60], but these focus on reconstructing point-clouds, or use preprocessed radar data like NuScenes [8] point clouds as input. None of these works solve the difficult problem of full mesh reconstruction from unprocessed radar data, and they also do not focus on long-range radar signals, where signal interference between closely spaced objects becomes minimal and single-object reconstruction is feasible. Instead, this work focuses on *full mesh reconstruction* from *raw, long-range, high-frequency* radar signals without per-scene optimization.

Within this domain of high-frequency radar, many existing deep learning algorithms infer object class rather than full shape. They typically encode spatial information using 1D convolutional neural networks (CNNs) [36, 53] or recurrent neural networks (RNNs) [58]. Some of these approaches also apply attention to spatial encodings [45, 53, 58], increasing model performance. InvRT [41], a custom transformer model, was designed to encode both the spatial and temporal structure of the radar signature to reconstruct the shape of roll-symmetric objects. The few methods that focus on full single-shape reconstruction do so for human

body meshes [10, 59, 61], but rely on parametric body priors, like SMPL [34], to reconstruct meshes.

The radar-based reconstruction of 3D shapes with *arbitrary* topology remains a challenging task with high sensitivity across geometries in noisy, partially observable settings. In addition, there is a lack of diverse high-frequency radar datasets that can be used to train robust deep learning models for radar-based reconstruction. This gap is likely because existing real high-frequency radar data is heavily restricted by security and IP concerns, and capturing such data at high fidelity requires access to specialized equipment. In order to drive future research and reduce the barrier to entry for single-shape reconstruction methods from high-frequency radar, we introduce two large-scale datasets of *diverse* geometries and simulated radar responses. We also tackle full 3D shape reconstruction and evaluate Radar2Shape against noisy, partially observable, and real radar responses to observe robustness.

3. Background

In this section, we discuss two core background concepts for Radar2Shape: Denoising Diffusion Probabilistic models and the first-principles physics model that generates high-frequency radar signatures from 3D object scattering. More details for each topic are provided in Appendix A and B.

3.1. Denoising Diffusion Probabilistic Models

Denoising Diffusion Probabilistic Models (DDPMs) [20] are generative models that leverage a forward diffusion process and a reverse denoising process to generate samples. The forward process adds Gaussian noise to a clean data sample x_0 over T timesteps, creating noisy samples x_t. The reverse process aims to recover the clean data distribution by progressively denoising x_t. The training objective of DDPMs is to minimize the variational lower bound of the negative log-likelihood of the generated data to match the true data distribution, over all timesteps 1 to T.

3.2. High-Frequency Radar Simulation for Single Object Reconstruction

The techniques for modeling radar signatures of 3D objects depend on the relative size of the object l and the wavelength of the radar λ. Most commercial and defense-related applications use high-frequency radar waveforms, where the object size is much larger than the radar wavelength, and where multiple closely spaced objects can be resolved. Because of this, signal interference between objects is minimal and can be ignored [43]. Therefore, in this domain, *single* object reconstruction from radar is feasible. A single object's scattering response can often be reduced to a summation of discrete scattering centers by taking advantage of the Geometric Theory of Diffraction (GTD) [24]. This reduction allows the use of parametric, component-based, scat-
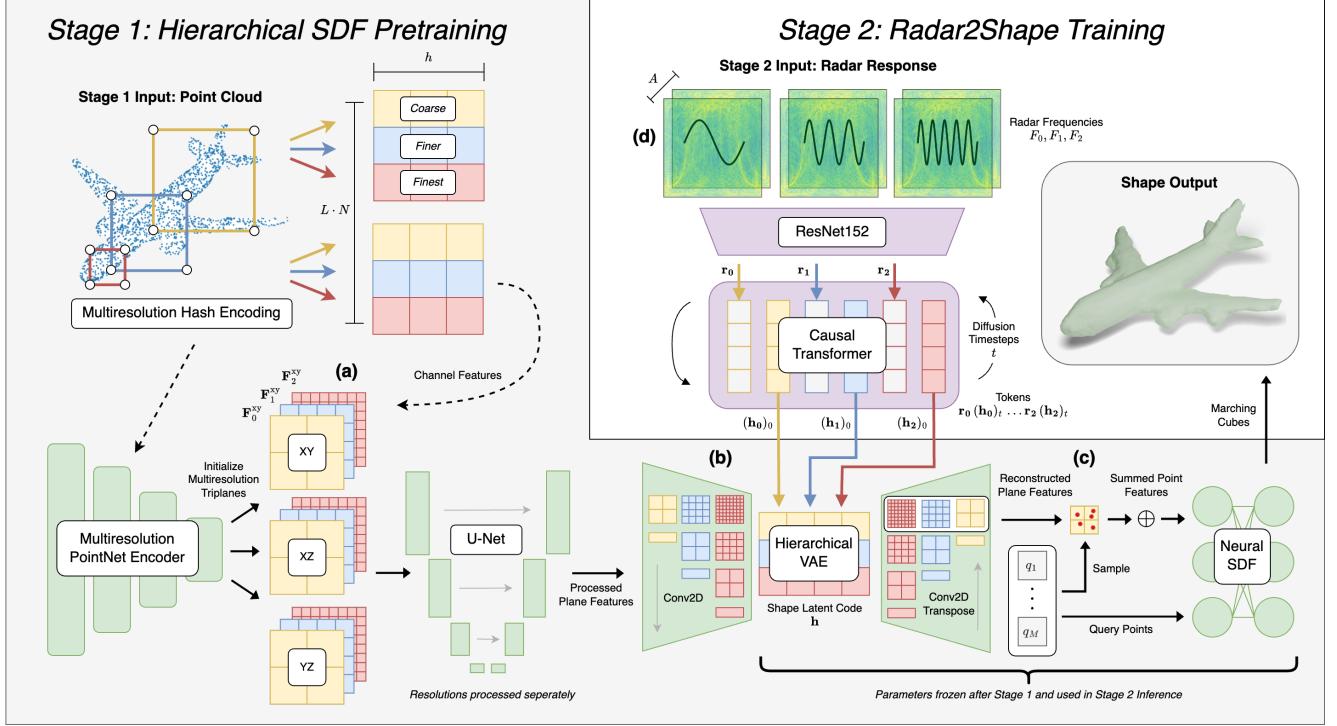
Figure 2. **Method.** Radar2Shape consists of two stages: 1) learning a multi-resolution, hierarchical latent space for 3D shapes, and 2) training a diffusion model to denoise in this space by conditioning on radar responses. In this figure, three hierarchical levels ($L = 3$) are shown. **(a)** In Stage 1, we learn per-point multiresolution features from a point cloud that are projected onto triplanes of L different grid resolutions. **(b)** A VAE then reconstructs each triplane independently to keep feature resolutions separate in its latent space. **(c)** Features are combined across resolutions to reconstruct the 3D geometry. **(d)** In Stage 2, a Transformer learns a sequence of L multiresolution radar embeddings from a radar response interleaved with the VAE's multiresolution latent shape features. This enable coarse-to-fine prediction in a conditional diffusion process. Green and purple modules represent parameters trained during Stage 1 and Stage 2, respectively.

tering models that reduce radar modeling to summing over responses of individual components. Examples of components are discrete points, spheres, rings, and triangles of a mesh, the latter being the focus of this paper.

4. Method

In this section, we present Radar2Shape, our method for generating 3D geometries from radar observations. Our approach relies on raw radar responses (Section 4.1). Radar2Shape consists of two stages: first, we learn a latent space for 3D geometries using a point cloud to SDF model (Section 4.2). Second, we train a diffusion model to denoise in this latent space by conditioning on radar responses, then produce an SDF by feeding the generated latent vector into our SDF decoder and running Marching Cubes [35] to obtain the mesh (Section 4.3). More details on each topic are in Appendix C and D.

4.1. Radar Signal Input

Figure 1 demonstrates our problem setting. The input to the learning task is a collection of radar responses at different viewing directions, and the output is the proposed observed geometry. Let u be viewing direction $u = (\sin\alpha\cos\phi, \sin\alpha\sin\phi, \cos\alpha)$, for $\alpha \in [0,\pi]$ and $\phi \in [0, 2\pi]$, representing the aspect and roll angles, respectively. The corresponding radar response $F(u, f)$ of an object is a sequence of real and imaginary scattering responses calculated using a linear set of frequencies $\{f_i\} \in [f_{\min}, f_{\max}]$, where the bandwidth of the signal is $B = f_{\min} - f_{\max}$. The input to Radar2Shape is the amplitude measurement, calculated by taking the magnitude of the real and imaginary components F and converted into decibel scale following $20 \times \log_{10}(|F|)$. This scale smooths out large fluctuations in signal strength and allows the input to represent a large range of values.

For general 3D geometries, we discretize α and ϕ into N_α and N_ϕ bins, respectively, such that $F \in \mathbb{R}^{N_\alpha \times N_\phi \times |\{f_i\}|}$. We also incorporate a roll-symmetric shape prior when comparing to the baseline method, InvRT, which considered only roll-symmetric shapes of the *Frusta* dataset [41]. In this case, the radar response F is identical across all roll angles ϕ, and it is sufficient to index F by the aspect angle α and the frequency f alone, such that $R \in \mathbb{R}^{N_\alpha \times |\{f_i\}|}$.

4.2. Stage 1: Hierarchical SDF Training

In the first stage denoted by the gray box in Figure 2, the SDF model is trained together with an encoder mapping object meshes $\mathcal{M}$ to hierarchical latent SDF codes $\mathbf{h} \in \mathcal{H}$, which represent the geometry of $\mathcal{M}$ at different resolutions. We draw inspiration from Diffusion SDF's [12] architecture that regularizes $\mathcal{H}$ for easier downstream diffusion training, but we disentangle the hierarchies in the regularized latent space which has not been done in previous work.

For a given number of resolution levels L, batch size B, points $\mathbf{p_i} \in \mathbf{P}$ on the surface of $\mathcal{M}$ where $1 \leq i \leq N$, we first embed each point into hierarchical feature space $\mathbf{f}_i^{(in)} = \text{MultiRes}(\mathbf{p_i})$, $\mathbf{f}_i^{(in)} \in \mathbb{R}^{L \times h}$ using the multiresolution hash encoding, and linearly project to the model dimension H such that $\mathbf{F}^{(in)} \in \mathbb{R}^{B \times N \times L \times H}$. Each level's features are independently processed with a series of ResNet blocks with local pooling, resulting in $\mathbf{F}^{(out)} \in \mathbb{R}^{B \times N \times L \times H}$ that contains per-point spatial context.

We then train a *Hierarchical VAE* (Figure 2.b.) to reconstruct triplanes generated from these features, while maintaining hierarchical levels. Each plane (XY, XZ, YZ) is initialized as a grid for each level $l \in L$ with spatial resolutions R increasing by powers of 2, from coarse to fine – for Radar2Shape, we use $R \in \{8, 16, 32, 64\}$. Each grid cell's value is the mean of features from points projecting orthographically into it:

$$(\mathbf{F}_l^\pi)_{u,v} = \frac{1}{|S_{u,v}^\pi|} \sum_{i \in S_{u,v}^\pi} (\mathbf{F}_l^{(out)})_i \qquad (1)$$

This creates a sparse feature grid for each level $\mathbf{F}_l^\pi \in \mathbb{R}^{B \times C \times R \times R}$ where the channel dimension of each level's grid uniquely corresponds to the features at that resolution from the multiresolution hash encoding (Figure 2.a.). A 2D U-Net then densifies each sparse grid independently, resulting in multiresolution triplane features $\{\mathbf{F}_l^{xy}, \mathbf{F}_l^{yz}, \mathbf{F}_l^{xz}\}$. The VAE independently reconstructs each resolution triplane feature (Figure 2.b.), maintaining the level dimension in its stochastic latent variables $\mu_\mathbf{h}, \sigma_\mathbf{h}^2 \in \mathbb{R}^{L \times Z}$ where Z is the latent shape dimension. We treat the all levels jointly as a distribution and apply the following KL-divergence loss: $\text{KL}(\mathcal{N}(\mu_\mathbf{h}, \sigma_\mathbf{h}^2) \| \mathcal{N}(0, 0.25))$, where $\mathbf{h} \sim \mathcal{N}(\mu_\mathbf{h}, \sigma_\mathbf{h}^2)$ with the reparameterization trick, $\mathbf{h} \in \mathbb{R}^{L \times Z}$, and $\mathbf{h}_l \in \mathbb{R}^Z$.

Query points $q_i \ldots q_M$ are then used to grid sample each triplane for all resolutions L, then summed across planes and resolutions to create a rich per-point multiresolution shape representation π_i. Each point's q_i coordinate position is concatenated with π_i which is input into a SDF MLP (Figure 2.c). The final objective becomes the sum of the L1 SDF prediction error and the KL-divergence loss, with no need for a traditional VAE reconstruction loss.

4.3. Stage 2: Radar-Conditional Generation

In the second stage, we jointly train a radar encoder $\Phi : F \mapsto \mathbf{r}$ and a denoising network $\Theta : (\mathbf{h}_l)_t, \mathbf{r}, t \mapsto \epsilon_t$ to predict denoised latent shape codes $(\mathbf{h}_l)_0$ in a coarse-to-fine manner along hierarchical levels. First, the radar response F defined in Section 4.1 is split into L linearly spaced blocks along the frequency dimension such that $F_j = \mathbb{R}^{N_\alpha \times N_\phi \times A}$ where $A = \frac{|\{f_i\}|}{L}$. We treat bin $j = 0 \ldots L$ as positions and add a sine-cosine positional encoding so Φ can distinguish inputs of different radar frequencies. We use a ResNet152 [19] for Φ due to its ability to efficiently extract information from signals with two spatial dimensions, which are aspect and roll in F. Each block is encoded as $\mathbf{r}_j = \Phi(F_j^{(in)})$. To encourage robustness in partially-observable scenarios, we randomly mask between 0% and 70% of the aspect and roll dimensions during training such that the unmasked regions remain continuous. We use a Transformer [52] to learn a sequence of interleaved low-to-high-frequency radar encodings and low-to-high resolution shape encodings (defined in Section 4.2) as $\mathbf{r}_0, \mathbf{h}_0, \ldots, \mathbf{r}_j, \mathbf{h}_l, \ldots, \mathbf{r}_L, \mathbf{h}_L$, and apply a lower-triangular causal attention mask. Therefore, Φ predicts the noise of a shape feature at resolution l *only* by attending to shape features at a coarser resolution and frequencies of the radar response $j \leq l$, enabling coarse-to-fine prediction.

The loss function is the mean squared error between the predicted noise and scheduled noise added to the odd tokens of the sequence. The denoising process is defined as:

$$(\mathbf{h})_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left((\mathbf{h})_t - \frac{\sqrt{1 - \alpha_t}}{\sqrt{1 - \bar{\alpha}_t}} \Theta((\mathbf{h})_t, \mathbf{r}, t) \right) \quad (2)$$

Incorporating Roll-Symmetric Shape Priors. For roll symmetric objects, we define a lower dimensional space of potential shapes. This space can be used to define a reduced parameter model and to output the radial profile, a sequence of coordinates $(\mathbf{r}, \mathbf{z})$ defining the outer bounds of a half cross-section. We encode this shape parameterization $\mathbf{h}_t$ with a single linear layer and ReLU to match the α dimension of R, then concatenate $\mathbf{h}_t$ with R along the frequency dimension. A 1D U-Net [49] then jointly encodes the noisy latent shape vector at timestep t with the radar response during downsampling, and learns the noise prediction for $\mathbf{h}_t$ during upsampling.

5. Experiments

In this section, we describe baselines, metrics, and the data generation process, then present results of Radar2Shape on three benchmark datasets and real radar data.

Baselines. We use three competitive image-to-shape models with tuned hyperparameters that span a variety of

Table 1. **Quantitative Results.** Test performance of models after training on the Manifold40-PO dataset, with partial observability models using training-time mask augmentations. Metrics are evaluated across 20 random heldout meshes in the zero noise setting. Test-time partial observability is applied using a randomly sampled mask on 70% of the signal, with monoconic having a fixed masked for consistent evaluation. Radar2Shape largely outperforms TMNET and LIST which struggle to learn and also Diffusion-SDF which is competitive in this domain. Radar2Shape has relatively strong zero-shot generalization to Manifold40-PO-SBR and the real monoconic radar response.

Dataset	Model	Full Observability			Partial Observability		
		CD ($\downarrow$)	IoU ($\uparrow$)	F-Score ($\uparrow$)	CD ($\downarrow$)	IoU ($\uparrow$)	F-Score ($\uparrow$)
Manifold40-PO	Radar2Shape	**64.47 $\pm$ 86.87**	**0.51 $\pm$ 0.19**	**0.22 $\pm$ 0.11**	**44.72 $\pm$ 58.54**	**0.59 $\pm$ 0.27**	**0.27 $\pm$ 0.18**
	Diffusion-SDF	508.92 $\pm$ 386.10	0.13 $\pm$ 0.11	0.05 $\pm$ 0.03	566.73 $\pm$ 321.76	0.10 $\pm$ 0.06	0.04 $\pm$ 0.02
	TMNet	3501.14 $\pm$ 391.20	0.01 $\pm$ 0.00	0.02 $\pm$ 0.02	9801.15 $\pm$ 3164.11	0.00 $\pm$ 0.00	0.01 $\pm$ 0.01
	LIST	73599.60 $\pm$ 10230.19	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	79936.75 $\pm$ 12058.06	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00
Manifold40-PO-SBR	Radar2Shape	**96.91 $\pm$ 85.96**	**0.44 $\pm$ 0.26**	**0.10 $\pm$ 0.05**	**121.14 $\pm$ 91.18**	**0.44 $\pm$ 0.25**	**0.10 $\pm$ 0.08**
	Diffusion-SDF	531.71 $\pm$ 322.96	0.12 $\pm$ 0.10	0.04 $\pm$ 0.01	456.92 $\pm$ 284.12	0.11 $\pm$ 0.10	0.05 $\pm$ 0.02
	TMNet	1235.35 $\pm$ 773.37	0.02 $\pm$ 0.02	0.02 $\pm$ 0.04	1060.80 $\pm$ 703.15	0.02 $\pm$ 0.02	0.03 $\pm$ 0.02
	LIST	20723.39 $\pm$ 2073.50	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	20728.32 $\pm$ 2080.41	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00
Monoconic	Radar2Shape	**8.403**	**0.811**	**0.139**	**40.647**	**0.681**	**0.104**
	Diffusion-SDF	40.510	0.712	0.077	504.227	0.191	0.047
	TMNet	229.233	0.041	0.028	916.115	0.014	0.012
	LIST	564.531	0.025	0.007	1189.351	0.021	0.016

Table 2. **Quantitative Results for Roll-symmetric Shapes.** Test performance of Radar2Shape and InvRT after training on the Frusta dataset, with both models using training-time mask and noise augmentations for their respective observability and noise level. Metrics are evaluated across 20 random heldout meshes under different observability and noise conditions (low =−80dB, medium =−60dB, and high =−40dB), with test-time partial observability applied as the same randomly sampled masks for up to 70% of the aspect. Radar2Shape outperforms InvRT across most metrics, notably with a larger performance gap in the difficult high-noise setting.

Noise	Model	Full Observability			Partial Observability		
		IoU-R ($\uparrow$)	IoU-S ($\uparrow$)	MATCH-S ($\downarrow$)	IoU-R ($\uparrow$)	IoU-S ($\uparrow$)	MATCH-S ($\downarrow$)
Low	Radar2Shape	0.67 $\pm$ 0.22	**0.73 $\pm$ 0.24**	**0.11 $\pm$ 0.09**	**0.62 $\pm$ 0.24**	**0.67 $\pm$ 0.25**	**0.12 $\pm$ 0.12**
	InvRT	**0.70 $\pm$ 0.13**	0.66 $\pm$ 0.20	0.16 $\pm$ 0.11	0.61 $\pm$ 0.25	0.66 $\pm$ 0.20	0.18 $\pm$ 0.21
Medium	Radar2Shape	**0.71 $\pm$ 0.18**	**0.76 $\pm$ 0.20**	**0.11 $\pm$ 0.11**	**0.66 $\pm$ 0.21**	**0.71 $\pm$ 0.24**	**0.12 $\pm$ 0.11**
	InvRT	0.70 $\pm$ 0.24	0.64 $\pm$ 0.18	0.18 $\pm$ 0.09	0.63 $\pm$ 0.23	0.66 $\pm$ 0.15	0.19 $\pm$ 0.12
High	Radar2Shape	**0.77 $\pm$ 0.16**	**0.79 $\pm$ 0.17**	**0.10 $\pm$ 0.10**	**0.70 $\pm$ 0.19**	**0.74 $\pm$ 0.21**	**0.14 $\pm$ 0.12**
	InvRT	0.70 $\pm$ 0.20	0.72 $\pm$ 0.13	0.26 $\pm$ 0.22	0.63 $\pm$ 0.23	0.67 $\pm$ 0.17	0.27 $\pm$ 0.20

reconstruction methods to benchmark the performance of Radar2Shape on 3D reconstruction. TMNet [44] iteratively transforms a topology to fit a target shape, and LIST [1] uses spatial transformers for a coarse and fine prediction. Diffusion-SDF [12] uses diffusion to predict SDFs, conceptually similar to our method. We adapt baselines to the radar problem by swapping their existing point cloud or image encoders for the same encoding network as Radar2Shape, a ResNet152, allowing equal comparison. We also choose these models for a fair comparison because they do not require camera estimation or pixel alignment, which would not make sense in the radar domain. For roll-symmetric geometries, we compare directly against the InvRT method, a transformer-based model that provides state-of-the-art performance on the roll-symmetric Frusta dataset [28, 41].

Metrics of Evaluation. We measure the ability of Radar2Shape to reconstruct general geometries by considering popular metrics for 3D mesh reconstruction [1, 11, 13, 31, 44]: Chamfer distance (CD), intersection over union (IoU), and F-Score (1%). For roll-symmetric shapes, the simplified $(\mathbf{r}, \mathbf{z})$ parametrization enables accuracy to be evaluated more extensively than the metrics used for general geometries: *IoU-S* measures the

quality of shape predictions using the 2-dimensional binary mask intersection-over-union (IoU). (2) *IoU-R* evaluates the ability of the predicted shapes to generate the ground truth radar phenomenology, and *MATCH-S* evaluates the accuracy between ground truth and predicted shape segments by matching pairs of $(\mathbf{r}, \mathbf{z})$. For further details on these metrics and their calculations, refer to Appendix E.

5.1. Dataset Generation

As discussed in Section 2.3, there is a lack of diverse large-scale high-frequency radar datasets that can be used to train robust deep learning models for radar-based single-object reconstruction. We use the previously studied Frusta dataset to train and evaluate Radar2Shape against a competitive radar baseline, and although these shapes are domain-relevant, they are limited to 2D, roll-symmetric geometries. Therefore, we introduce *Manifold40-PO*, the first publicly available, large-scale, high-frequency radar dataset which is generated from ModelNet40's [56] diverse set of over ten thousand unique real-world meshes. We rely on a widely accepted first-principles simulator using Physical Optics (PO) [2] to generate radar responses, and use the *Manifold40* [22] variant of ModelNet40 for it's advantageous simulation properties.

Coarse Feature Source

(a) (b) (c) (d)

(e)

(f)

No Arms, Legs from (f)
Body from (c)

(g)

Arms, Legs from (g)
Body from (d)

(h)

Thick legs from (h)
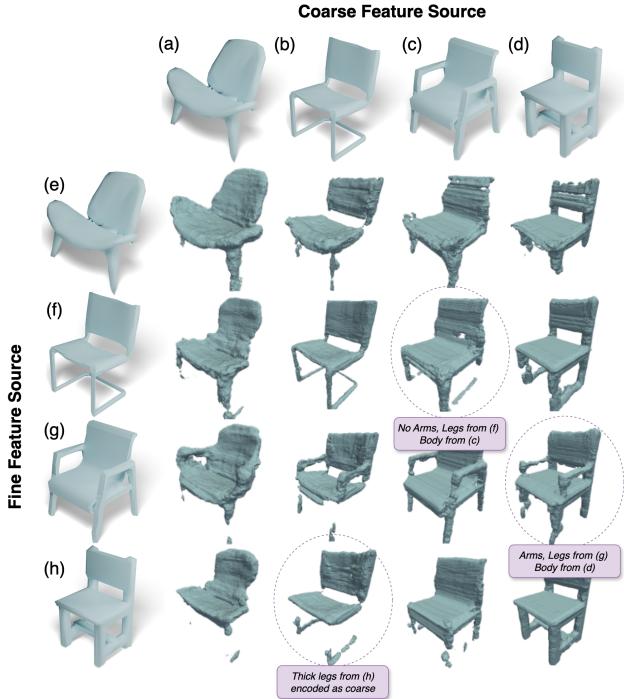encoded as coarse

Fine Feature Source

Figure 3. **Ablation.** Reconstruction of learned hierarchical latent codes with mixed coarse and fine features. For chairs, the model learns that the fine features correspond to arms and legs, because coarse features maintain the overall shape while the arms and legs are added or removed. This interpretability experiment demonstrates that our hierarchical SDF training method does indeed capture these coarse and fine features geometrically.

We also introduce a benchmark evaluation dateset with higher-fidelity effects, like multi-bounce interactions, using the Physical Optics and Shooting and Bouncing Rays (PO-SBR) algorithm [32], which we refer to as *Manifold40-PO-SBR*. This algorithm is computationally expensive, so we only generate approximately two thousand samples for fine-tuning and evaluation to show that Radar2Shape can generalize when trained with Manifold40-PO.

Real data in this radar domain, and equipment to record such data, is heavily restricted by cost, security and IP concerns. To our knowledge, there is no publicly available real high-frequency radar responses of single meshes. However, for this work, we are able to obtain real radar measurements across varying viewing angles of a monoconic object introduced in [37]. We evaluate Radar2Shape on these measurements after training on Manifold40-PO, and we make this data publicly available for future benchmarking. For further details on the monoconic object and simulation, see Appendix F and G.

5.2. Performance on SDF Reconstruction

Validating learned shape representations. To validate that Stage 1 learns to represent shape latent vectors at dif-

ferent shape resolutions, Figure 3 shows the meaning of these latent resolution levels from a geometric view. For chairs, the model learns that the fine features correspond to arms and legs, because the figure shows that coarse features maintain the overall shape while the arms and legs are added or removed. Note that in the bottom row, the legs of (h) are not substituted into the other shapes. This does not indicate failure, but is likely a mixing of geometric features across multiple granularity dimensions in the latent code, so unlike other samples, the legs of (h) may be encoded along with the shape's coarse feature. This is correct, and it can be expected because the legs appear thicker than other chairs, so they can be represented more coarsely. Stage 2 can then "piece together" these shapes part-by-part in a coarse-to-fine manner, without any ground truth segmentation labels.

Comparison to baselines. Radar2Shape's two-stage training approach greatly outperforms two competitive multi-view reconstruction methods, as shown in Figure 4. TMNet is unable to learn sphere deformations of sharp local features from the radar response, although it can generally learn the relative dimensions of the shape. LIST typically correlates query points with local image features, but since radar responses have different geometric dimensions than images, the model is unable to learn. This performance demonstrates the necessity for Radar2Shape, which geometrically leverages the frequencies of the radar response.

We further observe the benefit of Radar2Shape's coarse-to-fine refinement technique geometrically by comparing against the strongest baseline, Diffusion-SDF. Figure 4.b shows how Diffusion-SDF incorrectly reconstructs an airplane shape from the radar response of a chair. However, Diffusion-SDF still extracts the overall upward bending shape of the chair's arms and backrest, which it decodes as upward bending wings and a tail fin. Instead, Radar2Shape can first classify the shape correctly as a chair using the lowest frequency of the radar response and coarsest SDF resolution, then can focus on fine-grained features like the arms and backrest for an accurate reconstruction.

Figure 1 demonstrates the largest advantage of our method – in partial observability. Stage 1 of Diffusion-SDF learns a latent space which struggles to disentangle lower level shape features from the overarching structure, so the latent space for diffusion tends to represent entire objects. If Stage 2's encoder learns a feature that represents only part of an object, as is often the case in partially-observable radar responses, it may not be representable in the latent space but might be "nearest" to an entire object with similar features – this results in high-variance guesses (Figure 1.c.). Instead, Radar2Shape first learns a disentangled latent space with hierarchical features (Figure 3), allowing the radar features to be learned in accordance with the observed signal (Figure 1.b.). Table 1 reports reconstruction accuracy, where
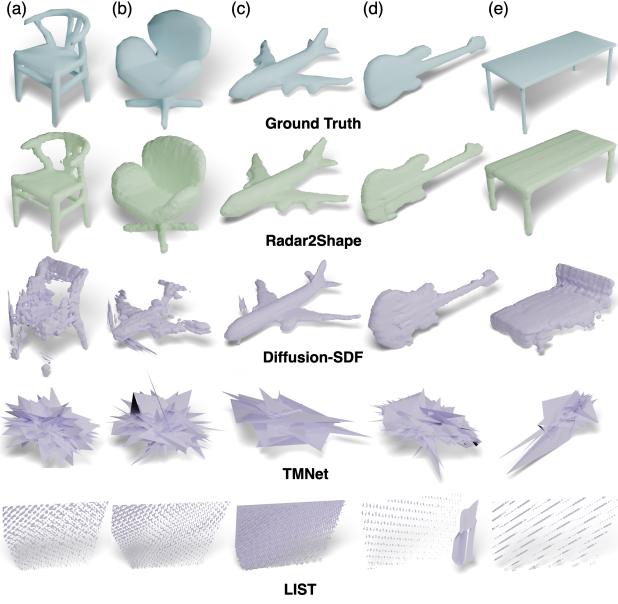
Figure 4. **Qualitative Results.** Comparison of select reconstructions from heldout fully-observed radar responses of Manifold40-PO. Radar2Shape consistently outperforms all baselines across a diverse set of meshes. TMNet and LIST exhibit mode-collapse, showing the difficulty of the radar-based 3D reconstruction problem when adapted to deterministic single/multi-view image-based reconstruction methods. Diffusion-SDF does the best among baselines, but often fails at reconstructing low-level features (shown with the chairs, table legs, and number of airplane engines).

Radar2Shape also largely outperforms baselines across aggregate reconstruction metrics.

Fine Tuning on Higher Fidelity Simulation. We show the zero-shot generalization results to Manifold40-PO-SBR in Table 1, but to extract additional performance, we also fine-tune on a portion of the generated data while maintaining heldout samples for evaluation. We use LoRA [21] fine-tuning on query and value attention projections, instead of full fine-tuning, to demonstrate that domain adaptation requires only lightweight changes to our model pretrained on the Manifold40-PO dataset. We observe a slight improvement in the fully observed setting, but modest improvement of about 0.07 IoU and F-Score in the partial setting, suggesting high-fidelity artifacts may be more important when shape information is sparse. Further discussion and quantitive metrics are in Appendix H.

5.3. Performance on Roll-symmetric Shapes

Table 2 compares the U-Net variation of Radar2Shape (Section 4.3) against InvRT across a variety of signal noise and observability settings. MATCH-S scores degrade for InvRT as noise increases, while Radar2Shape maintains performance. Since more noise creates an increasingly ill-posed problem, this performance gap demonstrates the advantage

Figure 5. **Qualitative Results on Real Data.** Reconstructions of a monoconic object from its *real* radar response, using Radar2Shape trained on Manifold40-PO. Radar2Shape predicts a wider tip, but is able to correctly predict the overall shape, base width, height, and angle near the base with low variance.

of diffusion as an inherently probabilistic model compared to a Transformer. Additionally, IOU-S provides an analogous metric to 3D IoU in Table 1, with a modest performance gap between Manifold40-PO and Frusta performance in partial and full observability. This demonstrates that incorporating a roll-symmetric shape prior indeed improves shape reconstructions. Appendix I contains analysis on common failure cases and distributional accuracy.

5.4. Application to Real Radar Data

To test zero-shot generalization properties of Radar2Shape, we consider real radar measurement data of a monoconic object introduced in [37]. Since this object is roll-symmetric, its measurements are taken only along the aspect dimension α. Although we could use the roll-symmetric variation of Radar2Shape, we choose to demonstrate the harder problem of full 3D reconstruction. Therefore, the input to Radar2Shape becomes the measurement repeated along roll angle ϕ. Figure 5 and Table 1 (Monoconic) show the results of single-shot generalization to this object, given that Radar2Shape is trained on Manifold40-PO objects. Radar2Shape struggles with the tip of the cone likely due to the real data being recorded at a different object scale, but even with this slight distribution shift, it exhibits the best performance compared to the other baselines and reconstructs elements like base width accurately.

6. Conclusion

We present Radar2Shape, a novel method that can reconstruct 3D shapes from radar responses by associating signal frequencies with multiresolution shape features. We empirically demonstrate that this method proves to be more accurate than previous work, especially in noisy and partially observable settings. This work also introduces a general method to learn multiresolution signed distance functions, and establishes two benchmark datasets consisting of diverse meshes and high-frequency radar responses to drive future research in high-frequency radar modeling.

Limitations and Future Work. There are some limitations to this work. We find that the hierarchical features learned in Stage 1 can be spatially bound (e.g. if pose

changes, the fine-grained representation of a chair's leg might change its shape), which future work could mitigate by using rotation invariance. Radar2Shape does not attempt to learn the scale of reconstructed objects, since Model-Net40 does not represent relative scale among objects correctly. Future work can also collect more diverse real-world data to evaluate performance, fine-tuning if necessary as we have demonstrated with Manifold40-PO-SBR.

Acknowledgments

References

[1] Mohammad Samiul Arshad and William J Beksi. List: learning implicitly from spatial transformers for single-view 3d reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9321–9330, 2023. 3, 6

[2] Constantine A Balanis. *Advanced engineering electromagnetics*. John Wiley & Sons, 2012. 6

[3] Dan Barnes and Ingmar Posner. Under the radar: Learning to predict robust keypoints for odometry estimation and metric localisation in radar. pages 9484–9490, 2020. 2

[4] Seth Bassetti, Brian Hutchinson, Claudia Tebaldi, and Ben Kravitz. Diffesm: Conditional emulation of temperature and precipitation in earth system models with 3d diffusion models. *Journal of Advances in Modeling Earth Systems*, 16(10): e2023MS004194, 2024. 2

[5] Kathleen M. Bergen, Daniel G. Brown, Eric J. Gustafson, and M. Craig Dobson. Radar remote sensing of habitat structure for biodiversity informatics. In *DG.O*. Digital Government Research Center, 2002. 2

[6] I. Bilik, O. Longman, S. Villeval, and J. Tabrikian. The rise of radar for autonomous vehicles: Signal processing solutions and future research directions. *IEEE Sig. Proc. Magazine*, 36(5):20–31, 2019. 2

[7] David Borts, Erich Liang, Tim Broedermann, Andrea Ramazzina, Stefanie Walz, Edoardo Palladin, Jipeng Sun, David Brueggemann, Christos Sakaridis, Luc Van Gool, et al. Radar fields: Frequency-space neural scene representations for fmcw radar. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–10, 2024. 3

[8] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020. 3

[9] Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. Efficient geometry-aware 3d generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16123–16133, 2022. 3

[10] Anjun Chen, Xiangyu Wang, Shaohao Zhu, Yanxu Li, Jiming Chen, and Qi Ye. mmbody benchmark: 3d body reconstruction dataset and analysis for millimeter wave radar. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 3501–3510, 2022. 3

[11] Yiwen Chen, Tong He, Di Huang, Weicai Ye, Sijin Chen, Jiaxiang Tang, Zhongang Cai, Lei Yang, Gang Yu, Guosheng Lin, and Chi Zhang. Meshanything: Artist-created mesh generation with autoregressive transformers. In *International Conference on Learning Representations (ICLR)*, 2025. Poster. 6

[12] Gene Chou, Yuval Bahat, and Felix Heide. Diffusion-sdf: Conditional generative modeling of signed distance functions. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2262–2272, 2023. 2, 3, 5, 6

[13] Christopher B. Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction. *CoRR*, abs/1604.00449, 2016. 6

[14] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 2

[15] Xinjie Gao, Bi'an Du, and Wei Hu. Hieroctfusion: Multi-scale octree-based 3d shape generation via part-whole-hierarchy message passing. *arXiv preprint arXiv:2508.11106*, 2025. 3

[16] Ruixu Geng, Yadong Li, Dongheng Zhang, Jincheng Wu, Yating Gao, Yang Hu, and Yan Chen. Dream-pcd: Deep reconstruction and enhancement of mmwave radar pointcloud. *IEEE Transactions on Image Processing*, 2024. 3

[17] Rohit Girdhar, David F. Fouhey, Mikel Rodriguez, and Abhinav Gupta. Learning a predictable and generative vector representation for objects. *CoRR*, abs/1603.08637, 2016. 3

[18] Zhiye Guo, Jian Liu, Yanli Wang, Mengrui Chen, Duolin Wang, Dong Xu, and Jianlin Cheng. Diffusion models in bioinformatics and computational biology. *Nature reviews bioengineering*, 2(2):136–154, 2024. 2

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 5

[20] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 2, 3

[21] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. 8

[22] Shi-Min Hu, Zheng-Ning Liu, Meng-Hao Guo, Jun-Xiong Cai, Jiahui Huang, Tai-Jiang Mu, and Ralph R Martin. Subdivision-based mesh convolution networks. *ACM Transactions on Graphics (TOG)*, 41(3):1–16, 2022. 6

[23] Abhishek Kar, Shubham Tulsiani, João Carreira, and Jitendra Malik. Category-specific object reconstruction from a single image. In *CVPR*, pages 1966–1974. IEEE Computer Society, 2015. 3

[24] Joseph B Keller. Geometrical theory of diffraction. *J. Opt. Soc. Am*, 52(2):1, 1962. 3

[25] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 3

[26] D. Kissinger. *Radar Fundamentals*, pages 9–19. Springer US, Boston, MA, 2012. 2

[27] Tzofi Klinghoffer, Xiaoyu Xiang, Siddharth Somasundaram, Yuchen Fan, Christian Richardt, Ramesh Raskar, and Rakesh Ranjan. Platonerf: 3d reconstruction in plato's cave via single-view two-bounce lidar. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14565–14574, 2024. 3

[28] Colin Kohler, Nathan Vaska, Ramya Muthukrishnan, Whangbong Choi, Jung Yeon Park, Justin Goodwin, Rajmonda Caceres, and Robin Walters. Symmetric models for radar response modeling. In *NeurIPS 2023 Workshop on Symmetry and Geometry in Neural Representations*, 2023. 6

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

[30] Lizao Li, Robert Carver, Ignacio Lopez-Gomez, Fei Sha, and John Anderson. Generative emulation of weather forecast ensembles with diffusion models. *Science Advances*, 10(13): eadk4489, 2024. 2

[31] Manyi Li and Hao Zhang. D2im-net: Learning detail disentangled implicit fields from single images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10246–10255, 2021. 6

[32] H. Ling, R.-C. Chou, and S.-W. Lee. Shooting and bouncing rays: calculating the rcs of an arbitrarily shaped cavity. *IEEE Transactions on Antennas and Propagation*, 37(2):194–205, 1989. 7

[33] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations (ICLR) 2023*, 2023. 2

[34] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. Smpl: A skinned multiperson linear model. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 851–866. 2023. 3

[35] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Seminal graphics: pioneering efforts that shaped the field*, pages 347–353. 1998. 4

[36] J. Lundén and V. Koivunen. Deep learning for HRRP-based target recognition in multistatic radar systems. In *IEEE Radar Conf.*, pages 1–6, 2016. 3

[37] J.T. Mayhan, M.L. Burrows, K.M. Cuomo, and J.E. Piou. High resolution 3d "snapshot" isar imaging and feature extraction. *IEEE Transactions on Aerospace and Electronic Systems*, 37(2):630–642, 2001. 7, 8

[38] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 3

[39] Ajay Narasimha Mopidevi, Kyle Harlow, and Christoffer Heckman. Rmap: Millimeter-wave radar mapping through volumetric upsampling. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1108–1115. IEEE, 2024. 3

[40] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022. 3

[41] Ramya Muthukrishnan, Justin Goodwin, Adam Kern, Nathan Vaska, and Rajmonda S Caceres. Invrt: Solving radar inverse problems with transformers. In *AAAI*, 2023. 2, 3, 4, 6

[42] Air Force Institute of Technology. Target pose estimation from radar data using adaptive networks. Technical report, Defense Technical Information Center, 1999. ADA361658. 2

[43] Alper Kürşat Öztürk. *Implementation of physical theory of diffraction for radar cross section calculations*. Bilkent Universitesi (Turkey), 2002. 3

[44] Junyi Pan, Xiaoguang Han, Weikai Chen, Jiapeng Tang, and Kui Jia. Deep mesh reconstruction from single rgb images via topology modification networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9964–9973, 2019. 6

[45] M. Pan, A. Liu, Y. Yu, P. Wang, J. Li, Y. Liu, S. Lv, and H. Zhu. Radar hrrp target recognition model based on a stacked CNN–Bi-RNN with attention mechanism. *IEEE Trans. on Geoscience and Remote Sensing*, 60:1–14, 2022. 3

[46] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation, 2019. cite arxiv:1901.05103. 3

[47] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019. 3

[48] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 2

[49] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention*, 19(3):934–941, 2015. 5

[50] Abhishek Sharma, Oliver Grau, and Mario Fritz. Vconv-dae: Deep volumetric shape learning without object labels. *CoRR*, abs/1604.03755, 2016. 3

[51] Thomas Truong and Svetlana Yanushkevich. Generative adversarial network for radar signal synthesis. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, 2019. 2

[52] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017. 5

[53] J. Wan, B. Chen, Y. Liu, Y. Yuan, H. Liu, and L. Jin. Recognizing the HRRP by combining CNN and BiRNN with attention mechanism. *IEEE Access*, 8:20828–20837, 2020. 3

[54] George Webber and Andrew J Reader. Diffusion models for medical image reconstruction. *BJR— Artificial Intelligence*, 1(1):ubae013, 2024. 2

[55] Jiajun Wu, Tianfan Xue, Joseph J. Lim, Yuandong Tian, Joshua B. Tenenbaum, Antonio Torralba, and William T. Freeman. Single image 3d interpreter network. *ArXiv*, abs/1604.08685, 2016. 3

[56] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1912–1920, 2015. 6

[57] Bojun Xiong, Si-Tong Wei, Xin-Yang Zheng, Yan-Pei Cao, Zhouhui Lian, and Peng-Shuai Wang. Octfusion: Octree-based diffusion models for 3d shape generation. In *Computer Graphics Forum*, page e70198. Wiley Online Library, 2025. 3

[58] B. Xu, B. Chen, J. Wan, H. Liu, and L. Jin. Target-aware recurrent attentional network for radar HRRP target recognition. *Sig. Proc.*, 155:268–280, 2019. 3

[59] Hongfei Xue, Yan Ju, Chenglin Miao, Yijiang Wang, Shiyang Wang, Aidong Zhang, and Lu Su. mmmesh: Towards 3d real-time dynamic human mesh construction using millimeter-wave. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, pages 269–282, 2021. 3

[60] Ruibin Zhang, Donglai Xue, Yuhan Wang, Ruixu Geng, and Fei Gao. Towards dense and accurate radar perception via efficient cross-modal diffusion model. *IEEE Robotics and Automation Letters*, 2024. 3

[61] Mingmin Zhao, Yingcheng Liu, Aniruddh Raghu, Tianhong Li, Hang Zhao, Antonio Torralba, and Dina Katabi. Through-wall human mesh recovery using radio signals. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10113–10122, 2019. 3

[62] Xin-Yang Zheng, Hao Pan, Peng-Shuai Wang, Xin Tong, Yang Liu, and Heung-Yeung Shum. Locally attentional sdf diffusion for controllable 3d shape generation. *ACM Transactions on Graphics (ToG)*, 42(4):1–13, 2023. 3