EQUIVARIANT DIFFUSION FOR THE INVERSE RADAR PROBLEM *

Anonymous authors
Paper under double-blind review

ABSTRACT

Reconstructing 3D geometries from their radar signal is a complex inverse problem, often involving unique domain expertise and manual steps. Although deep learning approaches have emerged to address the automation challenges of this problem, there are still significant performance gaps due to noisy radar signals and partial observability. In this work, we propose symmetry-aware modeling to reduce uncertainty over potential 3D shape distributions measured via radar signals. We present a radar-conditioned equivariant latent diffusion model that uses a two-stage training approach. The first stage learns shape representations in latent space and the second stage uses an equivariant diffusion model to denosie the latent representation, conditioned on radar signal to generate the 3D shape. We introduce an equivariant FiLM layer that enables conditioning of our diffusion model while ensuring rotational equivariance throughout the generation process. Finally, we ensure equivariant latent representations of the conditioning radar signal by using a spherical CNN model. We show that our model predicts 3D geometries consistent with the observed radar signatures. In addition, we demonstrate improved performance over other competitive baselines in reconstruction and sample diversity.

1 INTRODUCTION

Although radar has been used in diverse applications as a sensing modality, the problem of reconstructing the full 3D geometry of objects or scenes captured by radar signals is less explored and challenging. Often, radar signals are used to perform other inverse learning tasks, such as object classification, tracking, and task-specific feature extraction. In real-world settings, it is also common to have only partial radar observations that do not cover the full geometry of the objects, introducing significant uncertainty in the geometry reconstruction task.

Previous deep learning approaches have attempted to solve this problem (Muthukrishnan et al., 2023; Lundén & Koivunen, 2016; Wan et al., 2020), but significant performance gaps persist. Due to the lack of large datasets and the high computational complexity required for this task, many methods fall short. We address this challenge employing geometric deep learning, which utilizes equivariant neural networks (Cohen et al., 2018). Equivariant neural networks use symmetries present in the data to learn more robust features and have been shown to require less data and training time compared to non-equivariant models. Kohler et al. (2025) demonstrates the superiority of equivariant models for forward radar problems, where the goal is to predict radar signals from 3D geometry.

Due to the presence of noise in radar signals, it is difficult to solve this problem with deterministic models. Diffusion models are a promising avenue to capture this uncertainty, but most existing state-

*DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited. This material is based upon work supported by the Under Secretary of War for Research and Engineering under Air Force Contract No. FA8702-15-D-0001 or FA8702-25-D-B002. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of War for Research and Engineering. © 2026 Massachusetts Institute of Technology. Delivered to the U.S. Government with Unlimited Rights, as defined in DFARS Part 252.227-7013 or 7014 (Feb 2014). Notwithstanding any copyright notice, U.S. Government rights in this work are defined by DFARS 252.227-7013 or DFARS 252.227-7014 as detailed above. Use of this work other than as specifically authorized by the U.S. Government may violate any copyrights that exist in this work.

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.

of-the-art diffusion models for shape generation are conditioned on single-view images or partial point clouds rather than spherical radar signals (Chou et al., 2023; Zeng et al., 2022). Furthermore, they lack architectural components for conditioning using equivariant embeddings.

To address these challenges, we introduce a radar-conditioned equivariant latent diffusion model that can learn a probabilistic mapping of radar signatures to the corresponding 3D geometry. We make the following contributions:

- We introduce an equivariant FiLM layer for conditioning diffusion models on radar embeddings.
- We develop an SO(3) equivariant latent denoising diffusion model for 3D shape reconstruction from radar signals.
- We also show that our model outperforms other competitive baselines for F-score(1%), TMD and IOU metrics for Manifold40 Airplane dataset and have competitive results for frusta dataset.

2 Related Work

Equivariant Neural Networks. Equivariant neural networks are mathematically biased to respect symmetries in the input data (Cohen & Welling, 2016; Thomas et al., 2018; Fuchs et al., 2020), improving performance and efficiency in many domains. They have been used for 3D shape classification (Esteves et al., 2018; Cohen et al., 2018) and for predicting chemical properties of molecules from 3D atomic structures (Liao & Smidt, 2022; Liao et al., 2023). Deng et al. (2021) developed a framework for SO(3)-equivariant point cloud networks, later used to learn equivariant latent embeddings for 3D segmentation (Lei et al., 2023) and robotic diffusion policies (Yang et al., 2024). Kohler et al. (2023) use it to map 3D geometries to spherical signals to predict radar signals. These models outperform non-equivariant baselines, but none have been used to generate 3D shapes from radar signals.

3D Shape Reconstruction. Recent approaches to reconstructing 3D shapes from 2D images or partial point clouds largely focus on latent generative models and implicit representations. Zeng et al. (2022) combine a hierarchical VAE with two denoising diffusion models, while Tang et al. (2024) use an autoregressive auto-encoder to generate variable-length meshes from fixed-length latent codes, which are also used to train a diffusion model that generates meshes from 2D images or partial point clouds (Chou et al., 2023). Many of these works use Deep Signed Distance Functions (SDFs) (Park et al., 2019a), which map query points in $\mathbb{R}^3$ to their shortest distance to the surface, with the shape given by the zero level-set. Unlike these methods, we use an equivariant network to denoise the latent codes and an equivariant FiLM layer to condition the denoising process.

Radar Modeling. Prior deep learning works encode radar responses using 1D CNNs (Lundén & Koivunen, 2016; Wan et al., 2020), RNNs (Xu et al., 2019), or attention mechanisms (Pan et al., 2022; Wan et al., 2020; Xu et al., 2019), typically for target classification rather than full reconstruction. Muthukrishnan et al. (2023) use a transformer to encode the spatial and temporal aspects of the radar response and predict the shape profile of roll-symmetric objects. This strong shape prior improves performance but prevents reconstructing arbitrary shapes, a harder task when the radar response is noisy and partially observable. In contrast, we use a spherical CNN (Cohen et al., 2018) to extract features from radar signals.

3 Background

Problem setting We consider the problem of generating 3D objects from radar responses. We focus in particular on high-frequency radar waveforms, where the radar wavelength is much smaller than the object size. This type of radar is used in most commercial and defence-related applications. To calculate a radar response, we utilise Physical Optics (PO) (Balanis, 2012) to estimate how the waves scatter off a triangle of a mesh. Using the Geometric Theory of Diffraction (Keller, 1962), the entire mesh's scattering response from a viewing angle u can be reduced to a sum of individual triangle scattering responses. We use a *range profile*, a common view of radar, as input to our models.

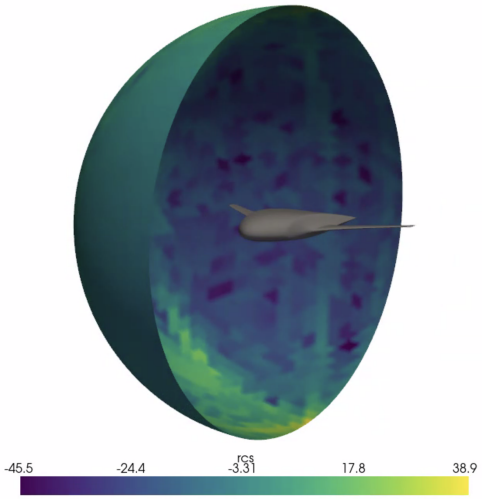

Figure 1: Illustration of the spherical radar signal of an airplane 3D shape for sensor viewing angles $[0, 180]$. Lighter, yellow colours represent higher amplitudes of the radar response.

This is calculated by taking the Discrete Fourier Transform of the mesh's scattering response across each frequency.

Figure 1 demonstrates the spherical nature of a radar signal (as a range profile), where the sphere is parametrised by the viewing angle u of a 3D geometry. In this work, we learn to reconstruct Signed Distance Functions (SDFs) (Park et al., 2019b) from such a spherical signal. SDFs are implicit representations that, for a given query point $X \in \mathbb{R}^3$, give us the distance of that query point from the surface of the shape. From a learned continuous SDF, we can reconstruct a shape by sampling several points in a bounding box and filtering out points that are close to the surface. We set the distance threshold to 0.01. Any query point with an SDF value less than the threshold is considered to be on the surface of the shape.

Equivariance. Consider a spherical radar signal $g\colon S^2 \to \mathbb{R}^d$, where S^2 is the 2-sphere and d is the number of range bins, and an inverse radar function f that maps g to a shape $P \subset \mathbb{R}^3$, i.e., $f(g) = P$. If g is rotated by $R \in \mathrm{SO}(3)$, the output shape is rotated by the same R, i.e., $f(Rg) = RP$. The inverse radar function is therefore equivariant with respect to SO(3), the group of all rotations about the origin in $\mathbb{R}^3$.

Equivariant Neural Networks. To respect this relationship between radar signals and output geometries, we use SO(3)-equivariant neural networks. A distinctive aspect of our work is that we combine two different equivariant architectures.

For our SDF model, we use the Vector Neurons framework (Deng et al., 2021) to generate SO(3)-equivariant latent codes for 3D point clouds. Vector Neurons extend the scalar neurons of traditional networks to 3D vectors and provide SO(3)-equivariant vector counterparts of standard neural network layers.

For our diffusion model, we build the denoising network with the `e3nn` library (Geiger et al., 2022), which projects spherical signals onto the irreducible representations (irreps) of SO(3) and provides operations that preserve equivariance in the irreps space. The irreps of SO(3) are indexed by non-negative integers $\ell = 0, 1, 2, \ldots$, where the ℓ-irrep has dimension $2\ell + 1$; $\ell = 0$ irreps are scalars and $\ell = 1$ irreps are vectors. Any representation of SO(3) decomposes into a direct sum of irreps.

Conditional Latent Diffusion Model. The denoising diffusion probabilistic model (DDPM) of Ho et al. (2020) uses a forward and a reverse diffusion process to generate new data samples from an

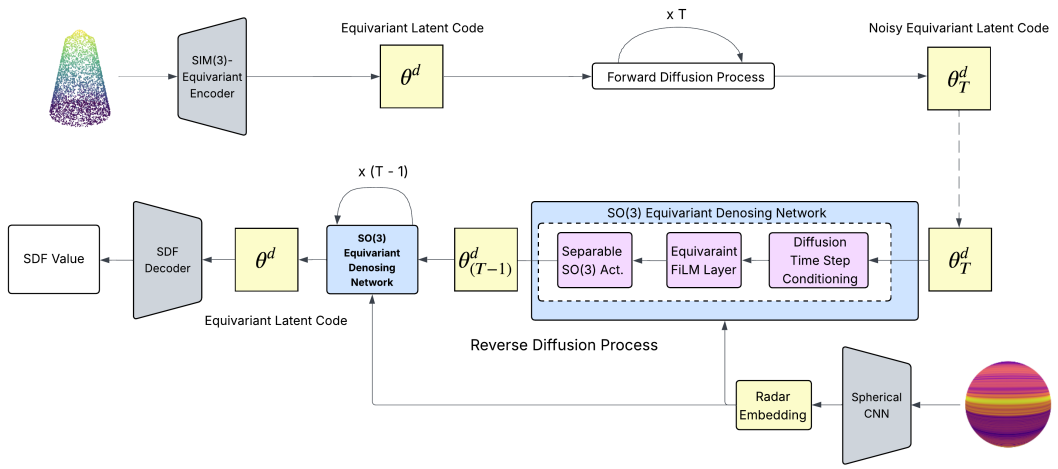

Figure 2: **A schematic representation of our method and its core components.** During training, an SO(3)-equivariant encoder produces point cloud latent representations. At each timestep t, an SO(3)-equivariant diffusion network predicts the denoised latent code, using radar spherical embeddings to condition the process via equivariant FiLM layers.

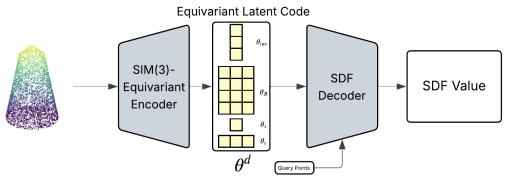

Figure 3: **SO(3)-equivariant SDF Model.** The input point clouds are mapped to an equivariant latent representation. The SDF decoder predicts SDF value for any query points using the equivariant latent code.

input distribution $q(x)$. During the forward process, we take a sampled data point x_0 and iteratively add Gaussian noise to it for T time steps, resulting in x_T. The forward diffusion process is formulated as

$$q(x_t \mid x_{t-1}) = \mathcal{N}\big(x_t;\, \sqrt{1-\beta_t}\, x_{t-1},\, \beta_t I\big), \tag{1}$$

where β_t is a hyperparameter.

During the reverse diffusion process, we do the opposite and iteratively denoise x_T for T time steps to obtain x_0. The reverse diffusion process is performed using a neural network model p_θ. The training objective of DDPMs is to minimize the variational lower bound of the negative log-likelihood of the generated data that matches the true data distribution, over all timesteps 1 to T. In conditional diffusion models, the input to p_θ also includes a condition to guide the denoising step.

4 METHOD

Our two-stage approach first trains an SO(3)-equivariant encoder-decoder to map point clouds to 3D shape latent codes. Subsequently, these codes were used to train a radar-conditioned latent diffusion model. The complete architecture is illustrated in Figure 2.

4.1 SO(3)-EQUIVARIANT SDF MODEL

Point Cloud Encoder. We use an SO(3)-equivariant encoder to map point clouds into latent shape encodings. More precisely, given a point cloud $P_{(N\times 3)}$, we generate a latent embedding $\theta = f(P) \in \mathbb{R}^d$, where d is the dimension of the latent embedding. The encoder f is a Vector

Neuron-based encoder (Deng et al., 2021; Lei et al., 2023). Vector Neuron is a simple yet effective architecture for mapping point clouds to equivariant latent space. It extends traditional 1D neurons in neural networks to 3D vector neurons and builds common neural network operators such as linear, non-linear and pooling layers to make equivariant models. The Vector Neuron encoder here transforms the input point cloud into a global embedding F, which is then mapped to four separate components, $(\theta_R, \theta_{inv}, \theta_c, \theta_s)$, where $\theta_R \in \mathbb{R}^{d\times 3}$ is the equivariant latent embedding of the input point cloud, $\theta_{inv} \in \mathbb{R}^{d\times 1}$ is the invariant latent embedding of the input point cloud, and $\theta_c \in \mathbb{R}^{1\times 3}$ and $\theta_s \in \mathbb{R}^{1\times 1}$ are the centroid and scale of the input point cloud, respectively. Here, θ_R and θ_c are vectors, while θ_{inv} and θ_s are scalar components.

SDF Decoder. For any given query point, $x \epsilon R^{3\times 1}$, we canonicalize x by subtracting θ_c and dividing by θ_s. The channel-wise inner product of the canonicalized x and θ_R is concatenated with θ_{inv} and passed through an MLP Ψ to predict the SDF value at x,

$$\mathrm{SDF}(x) = \Psi([\theta_{inv}, \langle \theta_R, \frac{x-\theta_c}{\theta_s} \rangle]).$$

4.2 SO(3)-EQUIVARIANT LATENT DIFFUSION MODEL

We use a denoising diffusion probabilistic model to generate latent embeddings, which are then used to generate 3D shapes by sampling SDF values. Unlike many diffusion models, which typically use a U-Net as the denoising network, we designed an SO(3)-equivariant MLP for denoising the latent embeddings. We also use an SO(3)-equivariant spherical CNN to generate the radar signal embeddings. Then, using our novel equivariant FiLM layers, we condition our denoising network on the radar embeddings. It is also worth noting that, although vector neuron architecture has not been previously used with e3nn-based networks, they can work together because both operate in the irreducible space of SO(3), the main distinction being that the vector neuron architecture is limited to the maximum frequency $\ell = 1$, while the e3nn-based architectures can theoretically extend to arbitrary maximum frequencies.

SO(3) Equivariant Denoising Network. The denoising network is created using the e3nn library (Geiger & Smidt, 2022). The latent embedding we are trying to denoise consists of scalars and vectors that are equivalent to the irreducible representations of degrees $\ell = 0$ (scalars) and $\ell = 1$ (vectors). First, the diffusion time step is encoded using a sinusoidal positional embedding and passed through an MLP to produce a time embedding, t_{embed}. The scalar feature t_{embed} is fused with all the intermediate feature representations using a fully connected tensor product as shown in Figure 4. To condition our model on the equivariant radar embedding, we created equivariant FiLM layers, which are described in the next section. The input is passed through the SO(3) equivariant linear, layer norm, and activation layers, before and after the FiLM layer. For activation functions, we used separable SO(3) activation functions introduced in the Equiformer architecture(Liao et al., 2023).

Equivariant FiLM Layer. Standard FiLM layers, first introduced in (Perez et al., 2018), use element-wise product and summation to modulate the scale and shift of conditional features. Such layers are used in label-conditioned diffusion models such as (Kumar et al., 2024). Using the standard FiLM layer directly will break the end-to-end equivariance properties of our model. To address this, we design a novel equivariant FiLM layer that allows us to interface between the vector neuron representation and e3nn through the use of SO(3) irreps.

As shown in Figure 5, our equivariant FiLM layer first transforms the input using three layers: an equivariant linear layer, a normalization layer, and an activation layer. The hidden features are then conditioned by scale and shift features. We get scale and shift by dividing the conditional feature across the channel dimension. In our case, given the radar embedding $R_L \in \mathbb{R}^{B\times C\times (L+1)^2}$, we get scale μ and shift $\sigma \in \mathbb{R}^{B\times (C/2)\times (L+1)^2}$. To maintain equivariance, we use a fully connected tensor product (Geiger et al., 2022) for scale and add shift operations across the Irreps dimensions (Cohen et al., 2018). We follow this with another equivariant linear layer and a normalization layer. We also add a skip connection (He et al., 2016). Given a hidden feature h, scale vector μ, and shift vector σ, we can express the equivaraint FiLM layer as,

$$\mathrm{EqFiLM}(h|\mu,\sigma) = (h \bigotimes \mu) \oplus \sigma. \tag{2}$$

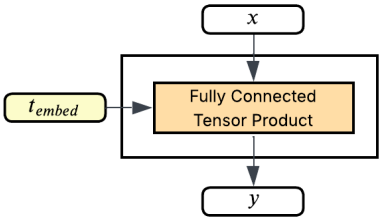

Figure 4: **Time step embedding layer**. Each time step embedding is encoded using a fully connected tensor product layer.

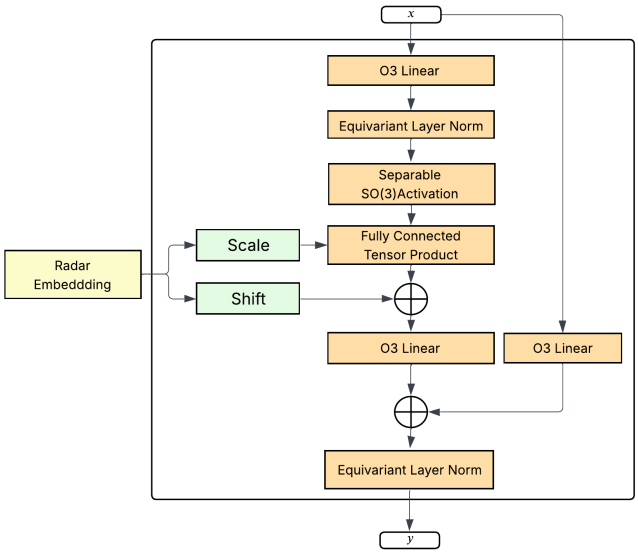

Figure 5: **Equivariant FiLM layer**. To condition diffusion models on equivariant features, we introduce an equivariant FiLM layer that uses a fully connected tensor product and channel-wise addition to preserve feature equivariance.

Proof of equivariance. We now show that the equivariant FiLM layer is SO(3)-equivariant.

Proposition 4.1. *For all* $R \in \mathrm{SO}(3)$*, the* EqFiLM *layer satisfies*

$$\mathrm{EqFiLM}\big(D(R)h \mid D(R)\mu,\, D(R)\sigma\big) = D(R)\ \mathrm{EqFiLM}(h \mid \mu, \sigma).$$

Proof. Let h, μ, and σ be the SO(3) irreps features in equation 2, and let $D(R)$ denote the Wigner D-matrix of $R \in \mathrm{SO}(3)$ acting on each feature. The fully connected tensor product is equivariant (Geiger et al., 2022, Eq. 1), i.e., $(D(R)h) \otimes (D(R)\mu) = D(R)(h \otimes \mu)$, and $D(R)$ acts linearly on $\oplus$. Therefore,

$$\begin{aligned}
\mathrm{EqFiLM}\big(D(R)h \mid D(R)\mu,\, D(R)\sigma\big) &= \big((D(R)h) \otimes (D(R)\mu)\big) \oplus D(R)\sigma \\
&= D(R)(h \otimes \mu) \oplus D(R)\sigma \\
&= D(R)\big((h \otimes \mu) \oplus \sigma\big) \\
&= D(R)\ \mathrm{EqFiLM}(h \mid \mu, \sigma). \qquad \square
\end{aligned}$$

SO(3) Equivariant Radar Embedding Network The radar embedding network takes in a spherical radar signal of resolution (ω, ϕ), which is expanded into a set of irreducible representations of SO(3) with maximum frequency of L using a truncated Fourier transformation $\mathbb{R}^{C\times\omega\times\phi} \rightarrow \mathbb{R}^{C\times(L+1)^2}$. The spherical Fourier transform is defined as follows (Esteves et al., 2018):

$$\hat{f}_m^\ell = \int_\omega \int_\phi f(\omega, \phi) Y_m^\ell(\omega, \phi)\, d\omega\, d\phi \tag{3}$$

where Y_m^l is the spherical harmonics of degree ℓ and order m. We perform the same fourier transformation for each input channel C to get irreps of SO(3) from a multi channel spherical signal.

We then perform SO(3) equivariant group convolutions Cohen et al. (2018) and separable SO(3) activation functions Liao et al. (2023) in multiple stages to progressively refine the representation. See Cohen et al. (2018) for more details on spherical CNNs. Kohler et al. (2025) demonstrate the benefit of using higher frequency spherical harmonics when working with radar signals. Higher frequency spherical harmonics help encode finer details and higher frequency components of the radar signal more accurately. Similar to their work, we gradually increase the maximum spherical harmonic degree and the channel dimensionality until they reach l_{radar} and d_{radar}, respectively.

5 EXPERIMENTS

In this section, we train our model on two different datasets and compare its performance across different metrics with other baseline methods. We also evaluate our model's performance under the partial observability setting.

5.1 DATASETS

Large-scale, open-source, real-world measurements of 3D objects are lacking for our setting of high-frequency, far-field radar sensing. As in other simulation-intensive engineering tasks, like inverse material design and aerodynamics, we use a radar simulator to generate controlled, large-scale training data: a physical optics simulator (Balanis, 2012), commonly used to model high-frequency, far-field radar signatures. We consider two 3D object benchmarks: Frusta (Kohler et al., 2023; Sortur et al., 2025a) and Manifold40 Airplane (Hu et al., 2022; Sortur et al., 2025b). The Frusta dataset consists of roll-symmetric meshes built by stacking a varying number of frusta of different radii and lengths, with flat plates or hemispheres on either side. Due to this symmetry, we generate radar responses along the non-symmetric axis over 360 orientations, $\theta \in [0, 2\pi]$. The Manifold40 Airplane dataset contains more complex, non-roll-symmetric objects, so we generate radar responses over both body axes in $[0, 2\pi]$, with 180 orientations per axis.

5.2 BASELINES

We compare against two baselines that isolate the role of equivariance in each architecture component and the role of the radar encoder.

Diffusion-SDF. Diffusion-SDF Chou et al. (2023) also uses two stages: a VAE encodes point clouds into latent codes, and a DALLE-2-style diffusion model Ramesh et al. (2022) generates shapes in this latent space. We replace its PointNet encoder for partial point clouds with a ResNet He et al. (2016) to condition on radar signals.

Non-equivariant Diffusion. This baseline keeps our SO(3)-equivariant SDF model but replaces the equivariant denoiser with a four-layer MLP (hidden dimensions $[2056, 4112, 8824, 2056]$), conditioned on transformer-encoded radar embeddings via cross-attention. This separates the contribution of equivariance in stage 1 from that in stage 2.

5.3 EQUIVARIANT DIFFUSION ARCHITECTURE

We set our latent embedding dimension $d = 256$. The SO(3) equivariant denoising network takes an input with 257 channels, which is first projected to 256 channels while expanding the Irreps maximum spherical harmonic degree from $L_{\max} = 1$ to $L_{\max} = 5$ using the equivariant linear and activation layers. The network then consists of three sequential blocks of time steps: conditioning layer, equivariant FiLM layers, and SO(3) activation layer, each with hidden dimensionality of 256 and maximum spherical harmonic degree $L_{max} = 5$. Finally, an equivariant linear and activation layer maps the representation back to match the original input channel dimension and L_{max}.

The SO(3) equivariant radar embedding network uses a 3 layer spherical CNN with hidden dimensions $[64, 64, 128]$ and $L_{\max}$ values $[3, 5, 5]$, giving us $d_{\text{radar}} = 128$ and $l_{\text{radar}} = 5$. We set the time embedding to $t_{embed} = 128$.

For our SDF model, we used the SO(3)-equivariant SDF model used by Lei et al. (2023) for their shape prior model.

5.4 MODEL TRAINING DETAILS

We train the SO(3)-Equivariant SDF model for 5000 epochs with a batch size of 16 and a learning rate of 1×10^{-4} with a decay of 0.3 at 1000, 2000, and 3500 epochs. After training, we select the best validation model to generate latent code for training the diffusion model. The SO(3) equivariant diffusion model was trained for 4600 epochs with a batch size of 8 for the frusta dataset and 20000

epochs with a batch size of 16 for the Manifold40 airplane dataset.The learning rate of 6×10^{-5} with a learning rate decay of 0.985 at every 400 training steps was used for both datasets.

Table 1: Performance results across various metrics on the Frusta and Manifold40 Airplane dataset. For each radar signal, we generate 10 meshes using the diffusion model. The metrics are averaged across all the test samples, $\pm$ indicates std. Error. **Bold** indicates the best performance across all the models for that metric.

		Full Observability			Partial Observability		
Dataset	Model	IOU (↑)	F-Score(1%) (↑)	TMD (↑)	IOU (↑)	F-Score(1%) (↑)	TMD (↑)
Frusta	Equiv. Diffusion	$\mathbf{0.6484} \pm 1\text{E}{-3}$	$\mathbf{0.4162} \pm 1\text{E}{-3}$	$\mathbf{0.063} \pm 1\text{E}{-3}$	$0.3804 \pm 1\text{E}{-3}$	$0.1222 \pm 5\text{E}{-4}$	$\mathbf{0.137} \pm 3\text{E}{-3}$
	Non-equiv. Diffusion	$0.6482 \pm 1\text{E}{-3}$	$0.3477 \pm 1\text{E}{-3}$	$0.040 \pm 1\text{E}{-3}$	$0.5102 \pm 1\text{E}{-3}$	$0.2061 \pm 1\text{E}{-3}$	$0.039 \pm 7\text{E}{-4}$
	Diffusion-SDF	$0.6481 \pm 1\text{E}{-3}$	$0.3852 \pm 1\text{E}{-3}$	$0.053 \pm 1\text{E}{-3}$	$\mathbf{0.5481} \pm 1\text{E}{-3}$	$\mathbf{0.2599} \pm 1\text{E}{-3}$	$0.068 \pm 1\text{E}{-3}$
Manifold40 Airplane	Equiv. Diffusion	$\mathbf{0.4055} \pm 6\text{E}{-3}$	$\mathbf{0.4437} \pm 7\text{E}{-3}$	$\mathbf{0.0348} \pm 6\text{E}{-2}$	$0.1405 \pm 4\text{E}{-3}$	$0.1182 \pm 3\text{E}{-3}$	$0.0129 \pm 6\text{E}{-3}$
	Non-equiv. Diffusion	$0.3038 \pm 4\text{E}{-3}$	$0.3279 \pm 4\text{E}{-3}$	$0.0253 \pm 7\text{E}{-4}$	$0.3038 \pm 4\text{E}{-3}$	$0.3279 \pm 4\text{E}{-3}$	$\mathbf{0.0245} \pm 6\text{E}{-4}$
	Diffusion-SDF	$0.3461 \pm 5\text{E}{-3}$	$0.3934 \pm 6\text{E}{-3}$	$0.0230 \pm 7\text{E}{-4}$	$\mathbf{0.3189} \pm 4\text{E}{-3}$	$\mathbf{0.3665} \pm 5\text{E}{-3}$	$0.0189 \pm 1\text{E}{-3}$

5.5 Metrics

We use standard metrics from the 3D reconstruction literature Arshad & Beksi (2023); Chou et al. (2023); Wu et al. (2020).

Chamfer Distance (CD) Fan et al. (2016) sums the nearest-neighbor distances between two point clouds in both directions.

Intersection over Union (IoU) Arshad & Beksi (2023) measures volumetric overlap as the ratio of the intersection to the union of the generated and ground truth meshes.

Total Mutual Difference (TMD) Chou et al. (2023) measures sample diversity. For k generated samples, we compute each sample's average CD to the other $k-1$ samples and sum these averages. Higher TMD indicates more diverse outputs, which matters most under partial observability, where many shapes are consistent with the input.

F-Score (1%) Wang et al. (2018) is the harmonic mean of precision (fraction of predicted points near a ground truth point) and recall (fraction of ground truth points near a predicted point), where points are considered close if they lie within a 1% distance threshold.

5.6 Shape Reconstruction

We evaluate our model on radar signals from the test set. For each signal, the diffusion model generates 10 latent codes, which we use to compute SDF values on a $128 \times 128 \times 128$ grid within a bounding box. We then extract meshes using marching cubes (Lorensen & Cline, 1998), as shown in Figure 6. On the Manifold40 Airplane dataset of non-symmetric objects, our model significantly improves on all metrics, outperforming the baselines in both reconstruction accuracy (IoU and F-Score (1%)) and output diversity (TMD). Figure 6 illustrates this by comparing the ground truth meshes with meshes generated by our model and the baselines. On the roll-symmetric Frusta dataset, our model outperforms all baselines on every metric when using fully observed radar signals. Notably, Diffusion-SDF was trained for about 20,000 epochs, compared with 4,600 and 3,799 epochs for Equivariant and Non-equivariant Diffusion, respectively, showing that the equivariant model achieves better results in far fewer epochs. Although Non-equivariant Diffusion uses a non-equivariant denoising network, it still relies on an equivariant model for shape reconstruction.

5.7 Shape Reconstruction From Partial Observability

We also evaluate our model on partially observed radar signals, obtained by masking 10–40% of each signal before generating latent codes with the diffusion model. Since the complete shape cannot be directly inferred from a partial signal, the generated shapes should reflect this uncertainty, which we measure with TMD. Table 1 summarizes the results of our method and the baselines.

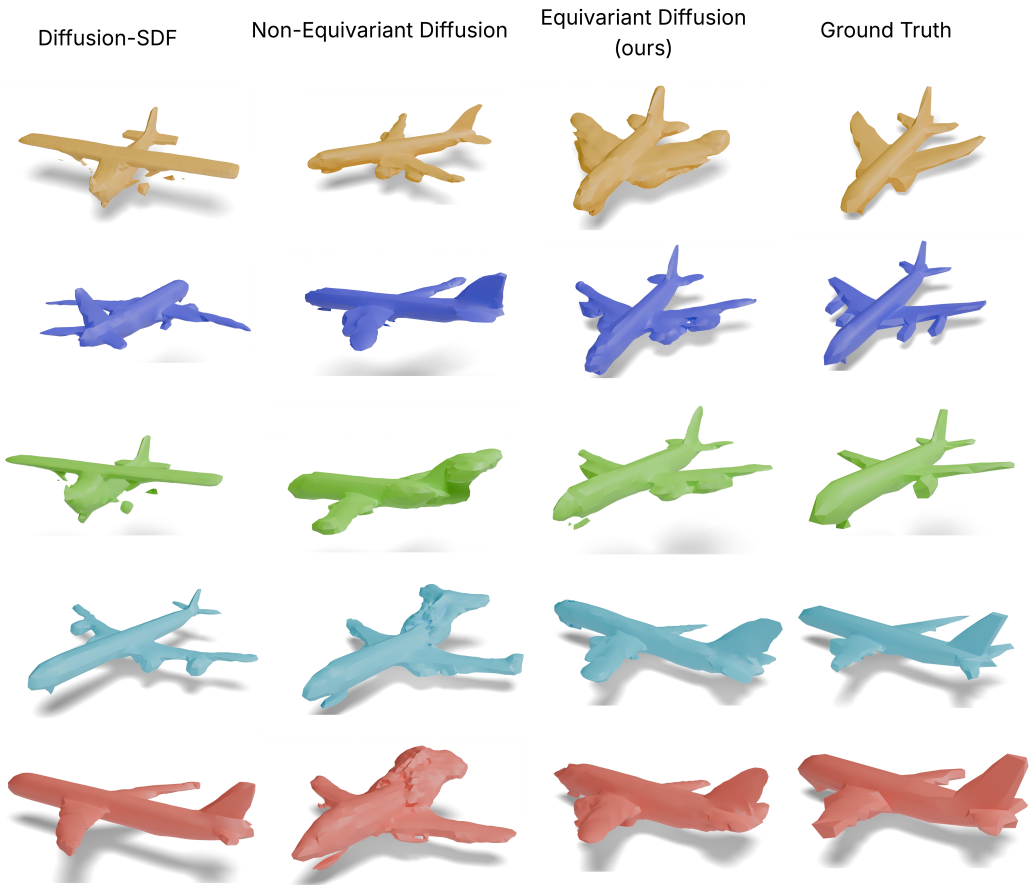

Figure 6: **Generated Manifold40 Airplane Samples.** We generate meshe samples using Fully observed radar signals from the Manifold40 Airplane dataset. We compare the ground truth mesh with meshes generated from the Diffusion-SDF, the non-equivariant diffusion model, and our equivariant diffusion model.

While all models drop in performance, our method achieves the highest TMD despite lower IoU and F-Score (1%). This indicates that while the baselines generate meshes similar to the ground truth mesh, our method generates a more diverse range of meshes rather than overfitting to a single one, as expected when reconstructing complete shapes from partially observed radar signals.

6 Conclusion

In this work, we present an equivariant diffusion model for the inverse problem of reconstructing the 3D geometry given its radar signature. We also introduced an equivariant FiLM layer for conditioning diffusion models on radar signals and developed a new architecture for conditional latent diffusion models that are equivariant to SO(3) group transformations in this work. Our method outperforms strong baselines in F-score (1%) , TMD and IOU metrics for Manifold40 Airplane dataset under fully observed radar signal setting and has competitive results for frusta dataset. We also show that when provided with a partially observed radar signal, our model gives more diverse point clouds compared to the baselines. We also show that our equivariant model requires fewer training epochs compared to non-equivariant methods.

AI use statement

In this work, we used generative AI tools only to polish the writing, such as improving clarity, grammar, and conciseness of the text. We have not used generative AI tools for generating research

ideas, writing code, running experiments, analyzing results, or producing figures. All AI-suggested edits were reviewed by the authors to ensure they preserved the intended meaning and technical accuracy. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

Reproducibility statement

To ensure reproducibility, we provide full details of our model architecture and all hyperparameters used for training and evaluation in the paper. The datasets, radar simulation setup, and evaluation metrics are described in the experiments section, and the proof of equivariance for the EqFiLM layer is given in Proposition 4.1. We plan to publicly release our code upon acceptance.

Author Contributions

If you'd like to, you may include a section for author contributions as is done in many journals. This is optional and at the discretion of the authors.

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments, including those to funding agencies, go at the end of the paper.

References

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*, pp. 9321–9330, 2023.

Constantine A Balanis. *Advanced engineering electromagnetics*. John Wiley & Sons, 2012.

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*, pp. 2262–2272, 2023.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016.

Taco Cohen, Mario Geiger, Jonas Kohler, and Max Welling. Spherical cnns. In *International Conference on Learning Representations*, 2018.

Congyue Deng, Or Litany, Yueqi Duan, Adrien Poulenard, Andrea Tagliasacchi, and Leonidas Guibas. Vector neurons: A general framework for so(3)-equivariant networks, 2021. URL https://arxiv.org/abs/2104.12229.

Carlos Esteves, Christine Allen-Blanchette, Ameesh Makadia, and Kostas Daniilidis. Learning so (3) equivariant representations with spherical cnns. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 52–68, 2018.

Haoqiang Fan, Hao Su, and Leonidas Guibas. A point set generation network for 3d object reconstruction from a single image, 2016. URL https://arxiv.org/abs/1612.00603.

Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in neural information processing systems*, 33: 1970–1981, 2020.

Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks, 2022. URL https://arxiv.org/abs/2207.09453.

Mario Geiger, Tess Smidt, Alby M., Benjamin Kurt Miller, Wouter Boomsma, Bradley Dice, Kostiantyn Lapchevskyi, Maurice Weiler, Michał Tyszkiewicz, Simon Batzner, Dylan Madisetti, Martin Uhrin, Jes Frellsen, Nuri Jung, Sophia Sanborn, Mingjian Wen, Josh Rackers, Marcel Rød, and Michael Bailey. Euclidean neural networks: e3nn, April 2022. URL https://doi.org/10.5281/zenodo.6459381.

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)*, pp. 770–778, 2016.

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

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.

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

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.

Colin Kohler, Purvik Patel, Nathan Vaska, Justin Goodwin, Matthew C. Jones, Robert Platt, Rajmonda S. Caceres, and Robin Walters. Bridging equivariant GNNs and spherical CNNs for structured physical domains. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=yh4DPshiWZ.

Manoj Kumar, Neil Houlsby, and Emiel Hoogeboom. Conditional diffusion on web-scale image pairs leads to diverse image variations, 2024. URL https://arxiv.org/abs/2405.14857.

Jiahui Lei, Congyue Deng, Karl Schmeckpeper, Leonidas Guibas, and Kostas Daniilidis. Efem: Equivariant neural field expectation maximization for 3d object segmentation without scene supervision, 2023. URL https://arxiv.org/abs/2303.15440.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In *The Eleventh International Conference on Learning Representations*, 2022.

Yi-Lun Liao, Brandon Wood, Abhishek Das, and Tess Smidt. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. *arXiv preprint arXiv:2306.12059*, 2023.

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*, pp. 347–353. 1998.

J. Lundén and V. Koivunen. Deep learning for HRRP-based target recognition in multistatic radar systems. In *IEEE Radar Conf.*, pp. 1–6, 2016. doi: 10.1109/RADAR.2016.7485271.

Ramya Muthukrishnan, Justin Goodwin, Adam Kern, Nathan Vaska, and Rajmonda S Caceres. Invrt: Solving radar inverse problems with transformers. In *AAAI*, 2023.

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. doi: 10.1109/TGRS.2021.3055061.

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation, 2019a. URL http://arxiv.org/abs/1901.05103.

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation, 2019b. URL https://arxiv.org/abs/1901.05103.

Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents, 2022. URL https://arxiv.org/abs/2204.06125.

Neel Sortur, Justin Goodwin, Rajmonda S. Caceres, and Robin Walters. Do you see the shape? diffusion models for noisy radar scattering problems. In *Frontiers in Probabilistic Inference: Learning meets Sampling*, 2025a. URL https://openreview.net/forum?id=fO7HfkOpKq.

Neel Sortur, Justin Goodwin, Purvik Patel, Luis Enrique Martinez Jr, Tzofi Klinghoffer, Rajmonda S. Caceres, and Robin Walters. Radar2shape: 3d shape reconstruction from high-frequency radar using multiresolution signed distance functions, 2025b. URL https://arxiv.org/abs/2511.17484.

Jiaxiang Tang, Zhaoshuo Li, Zekun Hao, Xian Liu, Gang Zeng, Ming-Yu Liu, and Qinsheng Zhang. Edgerunner: Auto-regressive auto-encoder for artistic mesh generation. *arXiv preprint arXiv:2409.18114*, 2024.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

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. doi: 10.1109/ACCESS.2020.2969450.

Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images, 2018. URL https://arxiv.org/abs/1804.01654.

Rundi Wu, Xuelin Chen, Yixin Zhuang, and Baoquan Chen. Multimodal shape completion via conditional generative adversarial networks, 2020. URL https://arxiv.org/abs/2003.07717.

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. ISSN 0165-1684. doi: https://doi.org/10.1016/j.sigpro.2018.09.041. URL https://www.sciencedirect.com/science/article/pii/S0165168418303220.

Jingyun Yang, Congyue Deng, Jimmy Wu, Rika Antonova, Leonidas Guibas, and Jeannette Bohg. Equivact: Sim(3)-equivariant visuomotor policies beyond rigid object manipulation, 2024. URL https://arxiv.org/abs/2310.16050.

Xiaohui Zeng, Arash Vahdat, Francis Williams, Zan Gojcic, Or Litany, Sanja Fidler, and Karsten Kreis. Lion: Latent point diffusion models for 3d shape generation, 2022. URL https://arxiv.org/abs/2210.06978.