DO YOU SEE THE SHAPE? DIFFUSION MODELS FOR NOISY RADAR SCATTERING PROBLEMS *

Neel Sortur[1] **Justin Goodwin**[2] **Rajmonda S. Caceres**[2] **Robin Walters**[1]

[1]Northeastern University
Boston, MA, USA

[2]MIT Lincoln Laboratory
Lexington, MA, USA

sortur.n@northeastern.edu

ABSTRACT

Determining the shape of 3D objects from high-frequency radar signals is analytically complex but critical for applications like aerospace and autonomous driving. Previous methods using deep learning have successfully been applied to this task, but the radar response in practical applications contains noise, which is hard to model deterministically. In addition, we often only observe the 3D object from partial viewing angles, leading to a complex one-to-many mapping task. In this work, we demonstrate that diffusion models are a suitable learning paradigm for radar inverse modeling due to their probabilistic learning and denoising properties. We present the radar2Shape model, which approximates the distribution of shape parameters conditioned on radar responses that are representative of practical applications. In addition to being more accurate than a deterministic competitive baseline across levels of noise, we show that the probabilistic nature of radar2Shape is important to capture the uncertainty associated with object reconstruction with partial data.

1 INTRODUCTION

Radar is a reliable sensing mechanism in adverse light and weather conditions, offering wide-ranging applications such as robotics Barnes & Posner (2020), autonomous driving Bilik et al. (2019), and remote sensing Bergen et al. (2002). Radar-based object characterization and reconstruction still presents a challenging inverse learning task. At long ranges, radar signals are often noisy and provide poor resolution Kissinger (2012). Furthermore, radar sensors often do not fully observe an object at all aspect angles, which introduces uncertainties for the reconstruction process.

Many existing deep learning algorithms developed for radar object characterization primarily infer object class rather than full shape. They typically encode the spatial information utilizing 1D convolutional neural networks (CNNs) Lundén & Koivunen (2016); Wan et al. (2020) or recurrent neural networks (RNNs) Xu et al. (2019). Some of these approaches also apply attention to spatial encodings Pan et al. (2022); Wan et al. (2020); Xu et al. (2019), increasing model performance. In Muthukrishnan et al. (2023), a custom transformer model was designed to encode both the spatial and temporal structure of the radar signature to reconstruct the full shape. However, the radar-based reconstruction process remains a challenging task with high sensitivity across geometries and noisy, partial observability settings. Generative models, such as Generative Adversarial Networks (GANs) Truong & Yanushkevich (2019) offer an alternative approach, but suffer from training instability problems and limited generalization Becker et al. (2022).

*DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited. This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. 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 Defense for Research and Engineering. © 2025 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.

Diffusion models have emerged as the model of choice to address some of these challenges (Ho et al., 2020). Their applications ranging from domains like bioinformatics (Guo et al., 2024) to climate science (Li et al., 2024; Bassetti et al., 2024) demonstrate their versatility, even when subject to noisy inputs (Webber & Reader, 2024). They have been shown to outperform GANs and generate diverse high-fidelity images for super-resolution and inpainting (Rombach et al., 2022; Dhariwal & Nichol, 2021).

In this work, we present the radar2Shape model, an extension of the Denoising Diffusion Probabilistic Model (DDPM). Radar2Shape denoises in the space of parametric roll-symmetric shapes and is conditioned on radar signatures of the shape. The robustness and versatility across domains motivates the use of diffusion models to model radar object characterization in noisy, partially observable settings, where there is high uncertainty about the surfaces of the shape that are not observed by the radar. We find that conditioning the denoising process on radar responses is robust under noise and partial observability, which leads to improvements over a competitive transformer-based (Vaswani, 2017) baseline. We also empirically quantify how well the predicted distribution matches the ground truth distribution of shapes, which is typically difficult for inverse problems such as these. Additionally, we explore the versatility of the model to instead condition on shapes and produce radar responses, which it is capable of doing to a high degree of accuracy.

2 Background

2.1 Radar Modeling and Simulation

Techniques for radar 3D object modeling depend on the relative size of the object l and the radar wavelength λ. Most commercial and defense-related applications use high-frequency radar waveforms. This leads to the optical regime, $l >> \lambda$, where radar sensors provide a signal complementary to optical imaging techniques while providing a capability that, for automotive applications, penetrates fog and smoke with scattering particle sizes smaller than the wavelength. Additionally, scattering can often be reduced to a summation of discrete scattering centers taking advantage of the Geometric Theory of Diffraction (GTD) Keller (1962). This allows the use of parametric, component-based, scattering models that reduce the radar modeling to simulating each component. Examples of components are discrete points, spheres, rings, and triangles. This paper will focus on the use of triangles.

2.2 Denoising Diffusion Probabilistic Models

Denoising Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020) 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 a noisy sample 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 matching the true data distribution, over all timesteps 1 to T. For a more comprehensive background on training DDPMs, see Appendix B.

2.3 Denoising Diffusion Implicit Models

Denoising Diffusion Implicit Models (DDIMs) (Song et al., 2021) introduce a non-Markovian reformulation of the reverse process in DDPMs, enabling faster and deterministic sampling while maintaining high sample quality. Unlike DDPMs, which rely on a stochastic Markov chain to generate samples, DDIMs parameterize the reverse process using a deterministic mapping, effectively skipping intermediate steps without additional training. This makes DDIMs particularly useful for applications where fast sample generation is important. For further details on DDIMs see Appendix B.6.

3 METHOD

3.1 RADAR MODELING

We utilize Physical Optics (PO) Balanis (2012) to model electromagnetic scattering from perfectly electrically conducting metallic meshes composed of triangular facets for high-frequency applications. To generate a range profile, a common view for radars, we calculate the scattering response of a mesh over a linear set of frequencies, such as those used in the common Linear Frequency Modulated (LFM) waveform Skolnik (1962). Details of our PO model are provided in Appendix C.2.

The range profile is a collection of real and imaginary scattering responses across a discrete set of N_R range bins and different viewing directions $\boldsymbol{u}$ denoted

$$R(\boldsymbol{u}) = \{R_i(\boldsymbol{u})\}, \text{ for } i = 1, \ldots, N_R, \tag{1}$$

where $\boldsymbol{u}$ is the observational unit vector given in spherical coordinates as $\boldsymbol{u} = (\sin\alpha\cos\phi, \sin\alpha\sin\phi, \cos\alpha)$ for $\alpha \in [0, \pi]$ and $\phi \in [0, 2\pi]$.

We consider 3D, roll-symmetric geometries, parameterized by vectors $(\mathbf{r}, \mathbf{z})$, corresponding to radial and axial parametrization parameters (see Figure 1 for an illustration). Since the geometry is roll symmetric, the radar response R is identical across all ϕ and it is sufficient to index R by range bin and aspect angle α. We discritize α into N_α bins such that $R \in \mathbb{R}^{N_\alpha \times N_R}$.

3.2 INVERSE RADAR MODELING VIA CONDITIONED DIFFUSION

We employ a Denoising Diffusion Probabilistic Model (DDPM) as described in Section 2.2 to generate 3D shapes conditioned a radar response in order to solve the radar-to-geometry inverse learning task. Our model is a latent space diffusion model consisting of three networks f_{enc}, f_{dec}, and ϵ_θ. Figure 2 contains an overview. The shape encoder $f_{enc} : (\mathbf{r}, \mathbf{z}) \mapsto \mathbf{h}$ maps the geometry parametrization to a latent vector $\mathbf{h} \in \mathbb{R}^{N_\alpha}$ and the shape decoder $f_{dec} : \mathbf{h} \mapsto (\mathbf{r}, \mathbf{z})$ performs the reverse mapping. The encoder consists of a single linear layer followed by a ReLU nonlinearity, and the decoder is linear.

Our latent shape diffusion model uses a U-Net (Ronneberger et al., 2015) as its noise prediction network ϵ_θ. See Appendix G for specific architecture details. In order to condition the diffusion model on the radar response R, at each step in the denoising process, we concatenate the latent shape vector $\mathbf{h_t} \in \mathbb{R}^{N_\alpha}$ with R along the range dimension giving $x_t = [\mathbf{h}_t \; R] \in \mathbb{R}^{N_\alpha \times (N_R+1)}$. The U-Net is a 1-dimensional convolutional network which convolves over the N_α-dimensional aspect angle axis and treats the range axis as $N_R + 1$ channels.

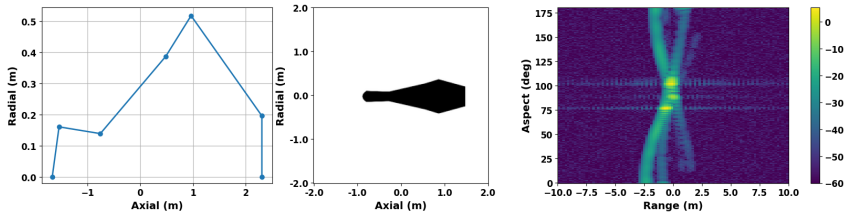

Figure 1: Cross section of a roll-symmetric frusta object parameterized by the radial and axial coordinates (left), the true shape cross section (middle) and the radar response including noise (right). The bright values in the right plot correspond to specular returns when the observation vector is perpendicular to the frustum section on the left.

By design, the U-Net $\epsilon_\theta \colon x_t \mapsto \epsilon$ outputs a noise vector $\epsilon \in \mathbb{R}^{N_\alpha \times (N_R+1)}$ of the same shape as the input x_t. However, we only wish to denoise the latent shape vector $\mathbf{h}_t$, and so we disregard the component of the noise vector corresponding to R. Denote the components of ϵ corresponding to $\mathbf{h}$ and R as $\epsilon = [\epsilon_h \; \epsilon_R]$. The loss function then becomes

$$L = \frac{1}{N_\alpha} \sum_{i=0}^{N_\alpha} (\epsilon_h - [\epsilon_\theta(x_t, t)]_h)^2. \tag{2}$$

During inference, since only $\mathbf{h}$ changes and R remains constant, the denoising process is defined

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\sqrt{1-\alpha_t}}{\sqrt{1-\bar{\alpha}_t}} \left[\epsilon_\theta(x_t, t)_h \;\; \mathbf{0} \right] \right). \tag{3}$$

The tensor $[\epsilon_\theta(x_t, t)_h \;\; \mathbf{0}]$ denotes $\epsilon_\theta(x_t, t)$ with ϵ_R set to 0. The denoised sample is defined $\mathbf{h} = [x_0]_h$. See Appendix B for more details on the diffusion process including the scales α_t and $\bar{\alpha}_t$.

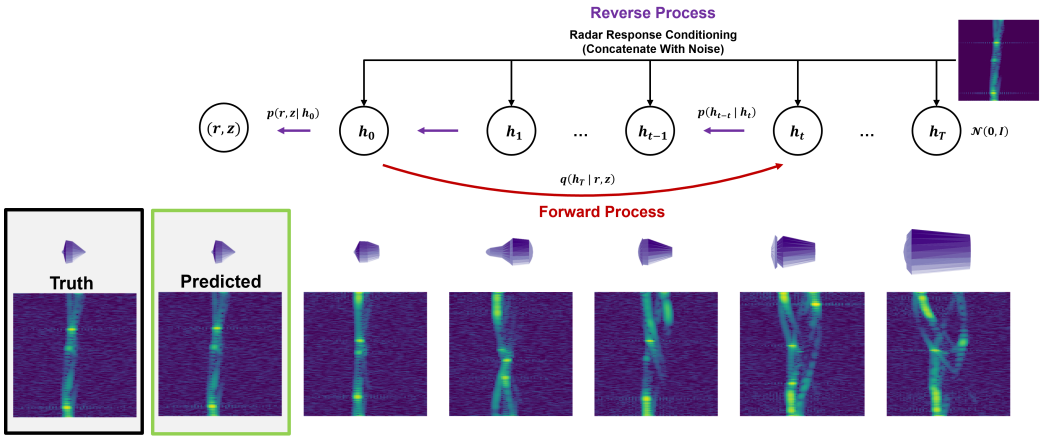

Figure 2: Overview of the radar2Shape model. To generate shapes that have a given radar response, parameters $(\mathbf{r}, \mathbf{z})$ are generated by denoising latent shape vectors $\mathbf{h}_t$ from random noise $\mathbf{h}_T$. This is done by iteratively sampling from the distribution $p(\mathbf{h}_{t-1}|\mathbf{h}_t)$ until $t = 0$. To train the model, a small amount of noise using q is added to each latent vector. The U-Net learns to denoise this, conditioned on the radar response. In successful cases, after denoising is completed at $t = 0$, the predicted shape's radar response is almost identical to the ground truth response that the model was conditioned on.

3.3 Dataset Generation

In this work consider the *Frusta* dataset introduced in Muthukrishnan et al. (2023); Kohler et al. (2023). We generate 3D geometries, where each corresponding mesh is a combination of roll-symmetric frusta stacked together, parameterized by a sequence of radial and body axis coordinates $(\mathbf{r}, \mathbf{z})$. For futher details on dataset generation, see Appendix F.

3.4 Observability Settings

We consider two observability settings controlling for the number of aspect angles or viewing orientations on the 3D object. In the fully observed regime, the diffusion model is conditioned on the full radar signal ($0 \leq \alpha \leq \pi$). For the partially observed setting, a mask is chosen to expose less than 70% of the object's surface such that the unmasked region is contiguous in aspect. We incorporate masking information into our dataset by setting masked aspects to a constant value of 0, and generate three randomly masked radar responses for a given shape to promote generalization.

3.5 Metrics of Evaluation

We develop multiple methods to evaluate the accuracy of shape samples generated by radar2Shape. Given the complex phenomenology of radar we aim to develop metrics that capture the model's ability to estimate the geometric shape, the shapes impact on radar phenomenology, and matching of the shape topology. Additionally, we aim to examine the ability of diffusion models to predict an accurate distribution of shapes. The following is a brief description of metrics used to evaluate these characteristics.

3.5.1 IOU-S: SHAPE OVERLAP

This metric evaluates the quality of shape predictions using the binary mask intersection-over-union (IoU). Ground truth and predicted shape parameters, $(\mathbf{r}, \mathbf{z})_{\texttt{gt}}$ and $(\mathbf{r}, \mathbf{z})_{\texttt{pred}}$, are used to generate corresponding mesh representations, which are then converted into 2D binary images of the shapes as shown in the middle plot of Figure 1. The IoU is computed to quantify the overlap between the predicted and ground truth image corresponding to the object cross-section, a value defined in the range of 0 and 1 where 0 indicates no overlap in shape and 1 is perfect overlap.

3.5.2 IOU-IM: RADAR SIGNAL OVERLAP

This metric evaluates the ability of the predicted shapes to generate the ground truth RF phenomenology, specifically speculars. Speculars occur when segments of the frustum are perpendicular to the observation vector $\boldsymbol{u}$. To detect the location of speculars across the aspect angle α, we develop a method that generates a binary mask used for intersection-over-union (IoU) calculations.

Ground truth and predicted shape parameters, $(\mathbf{r}, \mathbf{z})_{\texttt{gt}}$ and $(\mathbf{r}, \mathbf{z})_{\texttt{pred}}$, are used to generate corresponding mesh representations. For a given aspect angle, the sum of the projected area of all triangles in a mesh perpendicular to the observation is computed as $P_S(\boldsymbol{u})$ (see Appendix C.2 for more details). A binary mask is then crated for both predictions and ground truth by setting values to 1 when $P_S(\boldsymbol{u}) > 0.05$ and 0 otherwise. The IoU is calculated to measure how accurately the prediction captures the bright specular reflections. Similar to IOU-S, IOU-IM is in the $[0, 1]$ range, with 1 indicating the shape will produce speculars at the correct aspect angles.

3.5.3 MATCH: SHAPE MATCHING COST

This metric evaluates the matching accuracy between ground truth and predicted shape segments, excluding the hemisphere elements, corresponding to shapes parametrized by $(\mathbf{r}, \mathbf{z})_{\texttt{gt}}$ and $(\mathbf{r}, \mathbf{z})_{\texttt{pred}}$. Let $S_{\texttt{gt}}$ denote the number of ground truth segments and $S_{\texttt{pred}}$ the number of predicted segments. The linear sum assignment problem Crouse (2016) is solved to minimize the matching error cost, $\min \sum_i^{S_{\texttt{gt}}} \sum_j^{S_{\texttt{pred}}} C_{i,j} X_{i,j}$, where X is a binary matrix such that $X_{i,j} = 1$ when ground truth segment i is assigned to predicted segment j, and $C_{i,j}$ is the cost of matching i to j. The MATCH metric is defined as the total cost of the assignments represented by the above sum divided by the number of assigned segments, $\min\{S_{\texttt{gt}}, S_{\texttt{pred}}\}$.

3.5.4 DIST-MATCH: DISTRIBUTION ACCURACY METRIC

This metric evaluates how well the predicted distribution of our diffusion model matches the ground truth distribution, important in partially observable settings. Define the deterministic radar simulation used to generate R as $f : X \to Y$, where X represents shape parameters and Y represents the radar response. There exists a set $f^{-1}(y) = \{x \in X | f(x) = y\}$, in other words, all the possible shapes that correspond to an observed radar response. Our diffusion models h can be framed as trying to estimate this set (as samples from a conditional distribution), $h(y) = \widehat{f^{-1}(y)}$, which we must compare to the ground truth $f^{-1}(y)$.

However, in practice, it is difficult to obtain the ground truth because it is an ill-posed problem, and there is no model that can reliably extrapolate unknown shape information into a radar response that matches our masked observation. This is analogous to inverse problems in computer vision, like image inpainting and deblurring, that diffusion variants like DDRM (Kawar et al., 2022) attempt to solve. However, image quality and perceptual metrics like Frenchet Inception Distance and Peak Signal-to-Noise Ratio are not correlated well with evaluations on latent shape vectors that we use in this work, as they mostly focus on 2D RGB images. Even these metrics do not necessarily capture $\widehat{f^{-1}(y)}$ coverage of the ground truth distribution when large amounts of inpainting are required, because similarity to training set may not necessarily correlate with accuracy of generated pixels (Chen et al., 2024). For the radar domain, we must estimate both the accuracy of our produced shapes for the aspect angles in which they are observable, as well as the coverage of our diffusion model's shape estimations over the unknown shape distribution. Therefore, this metric is composed of two parts: ACC and COV.

Define a data sample (x_j, y_j) for shape j in the test set, with unmasked aspect angles (visible regions of the radar response) denoted as y_j^0 and masked angles as y_j^1. y_j without a superscript denotes both masked and unmasked aspect angles are included. We generate N samples from our diffusion model, conditioned on y_j, which takes the form $h(y_j) = \hat{x}_{ij}$ such that $x_i \in \widehat{f^{-1}(y_j)}$. We can obtain a radar response $\hat{y}_{ij}$ for this shape by running the radar simulation in Section 3.3. Equipped with a metric to measure distance in the space of y, the accuracy takes the form,

$$\text{ACC} = \frac{1}{M}\sum_{j=0}^{M}\frac{1}{N}\sum_{i=0}^{N}||\hat{y}_{ij}^0 - y_j^0||^2. \tag{4}$$

For coverage, we want to evaluate how accurately our generated shape distribution fits the true distribution of possible shapes under low observability. More accurate generated shape samples will be closer to the shape sample we have from the ground truth distribution $x_j \in f^{-1}(y_j)$ where in practice, x_j is the $(\mathbf{r}, \mathbf{z})$ parametrization. Across M shape samples with varying levels of partial observability, we can estimate coverage as the following,

$$\text{COV} = \frac{1}{M}\sum_{j=0}^{M}\min_i(||\hat{x}_{ij} - x_j||^2). \tag{5}$$

We similarly must choose a metric to measure distance in the space of x. The min operator takes the distance between the generated sample that is closest to x_j among our N generated samples. Intuitively, this metric ignores high variance in samples, and rewards such behavior when at least one of the samples is close to x_j - therefore, it should be observed in tandem with ACC. We do not have the distribution that x_j belongs to, so when we average over M samples, we average over the distribution of possible shapes (if M is sufficiently large) therefore becoming a proxy for coverage.

4 EXPERIMENTS

We measure how accurately radar2Shape can predict *Frusta* shapes, conditioned on a noisy and partially observable radar responses. We compare across a variety of sampling methods, noise levels and observability levels. Our models are trained on a variety of noise levels and observability settings indicated in Table 1 as data augmentations.

The training datasets consist of 25000 shape samples, each with three randomly applied masks, for a total of 75000 data samples. We set the number of timesteps for the diffusion process to 1000. Training is done for 120000 steps, using the Adam optimizer with an initial learning rate of $6\text{e}{-5}$, exponential decay every 400 steps with $\gamma = 0.985$. Once the diffusion models are trained, we evaluate our metrics with standard DDPM sampling, as well as DDIM sampling with 100 timesteps and $\eta = 0$. These methods are benchmarked against the InvRT transformer (Muthukrishnan et al., 2023) trained with our mask and observability augmented dataset described in Section 3.3.

Throughout these experiments, we denoise in the latent shape representation defined by the autoencoder f_{enc} and f_{dec} defined in Section 3.2. It is pretrained for a single epoch (to convergence) with 500 samples, using a mean squared error reconstruction loss, Adam (Kingma & Ba, 2015) optimizer, learning rate of 0.01, and batch size of 32.

4.1 FULLY OBSERVED

In the fully observed setting, our model has the radar response from every aspect (or viewing) angle of the shape - therefore, we can expect all metrics to be better across the board. We show in Table 1 our model's ability to reconstruct the radial and axial profile (and thus, the shape) across noise levels. In the high-noise setting, DDPM outperforms InvRT by a large margin on the IOU of the shape and it's associated radar response, as well as the matching error between the predicted and ground truth shape. In the low and medium noise settings, DDPM still shows improvement over InvRT, but the performance gap is slightly smaller. Since performance is observed to increase as noise increases, it

Table 1: Performance of our diffusion models (DDPM and DDIM) compared to the baseline transformer InvRT. DDPM, DDIM, and InvRT are trained and tested against different noise levels (low $=-80$dB, medium $=-60$dB, and high $=-40$dB), with randomly sampled masks up to 70%. In the full observability setting, metrics are computed over data with no applied mask. In the partial observability setting, we uniformly apply masks in all aspect of sizes between 0% and 70% with a starting aspect uniformly sampled. Metrics are averaged across 20 random (frusta shape, radar response) pairs.

		Full Observability			**Partial Observability**		
Noise	Model	*IOU-IM* (↑)	*IOU-S* (↑)	*MATCH* (↓)	*IOU-IM* (↑)	*IOU-S* (↑)	*MATCH* (↓)
low	DDPM	0.67 ± 0.22	$\mathbf{0.73 \pm 0.24}$	$\mathbf{0.11 \pm 0.09}$	$\mathbf{0.62 \pm 0.24}$	$\mathbf{0.67 \pm 0.25}$	$\mathbf{0.12 \pm 0.12}$
	DDIM	0.61 ± 0.24	0.68 ± 0.27	0.12 ± 0.12	0.58 ± 0.24	0.63 ± 0.26	0.13 ± 0.12
	InvRT	$\mathbf{0.70 \pm 0.13}$	0.66 ± 0.20	0.16 ± 0.11	0.61 ± 0.25	0.66 ± 0.20	0.18 ± 0.21
medium	DDPM	$\mathbf{0.71 \pm 0.18}$	$\mathbf{0.76 \pm 0.20}$	$\mathbf{0.11 \pm 0.11}$	$\mathbf{0.66 \pm 0.21}$	$\mathbf{0.71 \pm 0.24}$	$\mathbf{0.12 \pm 0.11}$
	DDIM	0.64 ± 0.24	0.69 ± 0.24	0.12 ± 0.14	0.63 ± 0.22	0.68 ± 0.24	0.14 ± 0.15
	InvRT	0.70 ± 0.24	0.64 ± 0.18	0.18 ± 0.09	0.63 ± 0.23	0.66 ± 0.15	0.19 ± 0.12
high	DDPM	$\mathbf{0.77 \pm 0.16}$	$\mathbf{0.79 \pm 0.17}$	$\mathbf{0.10 \pm 0.10}$	$\mathbf{0.70 \pm 0.19}$	$\mathbf{0.74 \pm 0.21}$	0.14 ± 0.12
	DDIM	0.68 ± 0.20	0.74 ± 0.21	0.14 ± 0.16	0.66 ± 0.21	0.71 ± 0.22	$\mathbf{0.13 \pm 0.15}$
	InvRT	0.70 ± 0.20	0.72 ± 0.13	0.26 ± 0.22	0.63 ± 0.23	0.67 ± 0.17	0.27 ± 0.20

is possible that DDPM overfits to the structure of the radar response in low noise settings, while in high noise settings, the model is forced to learn the most robust features.

DDIM exhibits slightly lower accuracy across metrics compared to DDPM, of about 0.05 IOU, but we observe an inference time improvement from 116.6 to 11.8 seconds on a single V100 GPU (with 100 DDIM steps), which can be more useful in certain remote sensing applications. We note that this drop off is somewhat smaller than the shape conditional radar response generation in Appendix A. We also show the limits of the models' performance under noise in Figure 5 for DDPM and Figure 6 for DDIM. They exhibit extremely consistent performance under the IOU metrics until around -20 dB, where a drop off of ~ 0.1 IOU occurs across samples.

4.2 Partially Observed

Compared to deterministic models, diffusion models theoretically perform better when conditioned on partially observable radar responses, as they naturally model the uncertainty of unobserved surfaces. We evaluate radar2Shape by measuring prediction accuracy against test data and assessing its distributional properties.

In Table 1, we can see that our DDPM model outperforms InvRT on all metrics in the high noise setting - this highlights how robust diffusion modeling is to noise, and suggests that the U-Net style encoding of the radar response is also beneficial. In the low and medium noise setting, DDPM still shows slightly better performance across metrics, and DDIM seems to match the performance of InvRT. *IOU-IM* and *IOU-S* typically exhibit decreases and *MATCH* increases from full observability to partial observability for a given model and noise level, especially in medium and high noise levels. This indicates slightly worse performance as expected for partial data. Both diffusion models and invrt show about equal performance drops across observability. Instead of averaging over mask levels, we also consider observing performance for a given object as masking increases - these results can be seen in Figure 7 for DDPM and Figure 8 for DDIM. We fix an arbitrary noise level to the medium setting of -60dB. There is little observed difference in the performance drop between DDPM and DDIM sampling. Overall, the model's strength is that it remains consistent in performance to about a mask fraction of 0.5, when a slight drop off of 0.1 to 0.2 IOU occurs for most objects in the mask fraction range of 0.5 to 0.7. Objects that are predicted with high accuracy in the fully observed setting also tend to have a sharper performance drop off as the mask fraction increases, compared to objects that have a lower accuracy in the fully observed setting - this is especially apparent in the left of Figure 8.

In addition to heldout test data, it is also important to understand that these ground truth data points are samples from a larger ground truth distribution. Therefore, we must also compare our predictions to the underlying data distribution, which gets wider as mask fraction increases. We expand upon

Table 2: Results for *DIST-MATCH* accuracy (ACC) and coverage (COV) metrics, with Peak-Bin as described in Appendix A as the distance between two radar responses y_{ij}^0 and y_j^0, and IOU-S as the distance between two shapes x_j and $\hat{x}_{ij}$. N and M are set to 20. We compare across different noise levels (low =-80dB, medium =-60dB, and high =-40dB and use random masks uniformly sampled up to 70% to induce uncertainty. On average, the diffusion model can generate samples that theoretically cover more of the ground truth distribution than a deterministic model such as InvRT, while maintaining accurate shape predictions corresponding to visible regions of the radar response.

Noise	Model	ACC : Peak-Bin (↓)	COV : IOU-S (↑)
low	DDPM	1.23 ± 1.40	0.91 ± 0.03
	InvRT	1.25 ± 0.68	0.72 ± 0.16
medium	DDPM	1.32 ± 1.49	0.85 ± 0.03
	InvRT	3.04 ± 2.51	0.74 ± 0.13
high	DDPM	1.66 ± 1.71	0.86 ± 0.01
	InvRT	9.47 ± 5.22	0.72 ± 0.11

this concept in terms of our masked radar response dataset by introducing the *DIST-MATCH* metric in Section 3.5.4, and discuss the results here. In Table 2, ACC : Peak-Bin shows that the diffusion model more accurately models the visible portion of the object than InvRT. COV: IOU-S also shows that it generates samples that theoretically cover more of the ground truth data distribution, while maintaining accuracy under visible regions of the radar response. We note that InvRT is deterministic so will produce the same results across N in these experiments, which inherently highlights the advantage of probabilistic modeling to properly model distribution variance.

Additionally, we measure radar2Shape's correlation between variance of generated shapes and the mask level (uncertainty) of the conditional radar response. Using the notation from Section 3.5.4, and defining y^m as a radar response with mask percentage m taken from the test dataset, we compute the mean of $x_i - y^m$ where $x_i \in \widehat{f^{-1}(y^m)}$. We obtain samples x_i by sampling repeatedly from y^m, and compare this mean and its variance across noise levels (low, medium, high) in Figure 9, Figure 10, Figure 11 respectively. We do indeed see a positive correlation between masking and variance of generated samples across noise levels. This means our model accurately represents the level of uncertainty in it's diffusion process, as domain knowledge suggests that the difficulty of the inverse problem increases drastically after $m = 0.5$ due to lack of specular information as defined in Section 3.5.2.

5 Conclusion

We present a novel method of using denoising diffusion models for radar object characterization, which in many settings, proves to be more accurate than deterministic modeling. Our object estimations also show robustness under perturbations of the conditional variables, which are high-frequency radar responses that are partially observable and contain varying amounts of noise. Additionally, although quantifying a ground-truth distribution to benchmark models against remains a challenge in probabilistic modeling for inverse problems, we develop a domain-specific metric to show that the probabilistic nature of diffusion modeling accurately represents the one-to-many mapping between partially observable radar responses and corresponding objects. Speeding up inference for these models with DDIM sampling maintains relatively strong performance in these settings. Future work can further evaluate our diffusion model in radar settings outside the optical regime, improve sampling time by using one-step distillation approaches, include more complex shapes with more general representations, and incorporate the geometric characteristics of far-field radar to improve sample-efficiency and performance.

Acknowledgements

This work is supported in part by NSF 2107256 and NSF 2134178. This material is also based on Neel Sortur's work supported by the NSF GRFP under Grant No. DGE-2439018.

REFERENCES

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

Dan Barnes and Ingmar Posner. Under the radar: Learning to predict robust keypoints for odometry estimation and metric localisation in radar. pp. 9484–9490, 05 2020. doi: 10.1109/ICRA40945.2020.9196835.

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.

Evan Becker, Parthe Pandit, Sundeep Rangan, and Alyson K. Fletcher. Instability and local minima in gan training with kernel discriminators. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

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*, ACM International Conference Proceeding Series. Digital Government Research Center, 2002. URL http://dblp.uni-trier.de/db/conf/dgo/dgo2002.html#BergenBGD02a.

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. doi: 10.1109/MSP.2019.2926573.

RB Blackman and JW Tukey. The measurement of power spectra dover publications. *Inc, New York*, 1958.

Tianyi Chen, Jianfu Zhang, Yan Hong, Yiyi Zhang, and Liqing Zhang. Assessing image inpainting via re-inpainting self-consistency evaluation. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pp. 291–300, 2024.

David F Crouse. On implementing 2d rectangular assignment algorithms. *IEEE Transactions on Aerospace and Electronic Systems*, 52(4):1679–1696, 2016.

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

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.

Roger F. Harrington. *Field Computation by Moment Methods*. Wiley-IEEE Press, 1993. ISBN 0780310144.

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

Wolfgang JR Hoefer. The transmission-line matrix method-theory and applications. *IEEE Transactions on Microwave Theory and Techniques*, 33(10):882–893, 1985.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022.

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

Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, San Diega, CA, USA, 2015.

D. Kissinger. *Radar Fundamentals*, pp. 9–19. Springer US, Boston, MA, 2012. ISBN 978-1-4614-2290-7. doi: 10.1007/978-1-4614-2290-7_2. URL https://doi.org/10.1007/978-1-4614-2290-7_2.

Eugene F Knott, John F Schaeffer, and Michael T Tulley. *Radar cross section*. SciTech Publishing, 2004.

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.

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.

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.

Per-Gunnar Martinsson. *Fast Multipole Methods*, pp. 498–508. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015. ISBN 978-3-540-70529-1. doi: 10.1007/978-3-540-70529-1_448. URL https://doi.org/10.1007/978-3-540-70529-1_448.

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

Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pp. 8162–8171. PMLR, 2021.

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.

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*, pp. 10684–10695, 2022.

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. doi: 10.1007/978-3-319-24810-3_154.

Merrill I Skolnik. Introduction to radar. *Radar handbook*, 2:21, 1962.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.

Thomas Truong and Svetlana Yanushkevich. Generative adversarial network for radar signal synthesis. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–7, 2019. doi: 10.1109/IJCNN.2019.8851887.

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

Apoorv Vyas, Angelos Katharopoulos, and François Fleuret. Fast transformers with clustered attention. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 21665–21674. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/f6a8dd1c954c8506aadc764cc32b895e-Paper.pdf.

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.

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

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.

Kane Yee. Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. *IEEE Transactions on antennas and propagation*, 14(3):302–307, 1966.

Table 3: Test performance of our diffusion model used for shape conditional static pattern generation. Results are averaged over 20 random (frusta shape, radar response) pairs. DDPM sampling proves to generate much higher quality samples than DDIM (with $\eta = 0$), also evidenced by Figure 3 and Figure 4.

Sampling	Inference Steps (↓)	Inference Time (s) (↓)	RMSE (dB) (↓)	Peak-Bin (↓)
DDPM	1000	224.0	22.99 ± 3.03	0.52 ± 0.36
DDIM	100	22.1	23.66 ± 2.30	5.68 ± 1.65

A Appendix: Shape-condition diffusion for static pattern generation

The nature of the denoising process makes the DDPM framework very flexible, allowing it to model arbitrary signals. The U-Net noise prediction backbone used in this work denoises both the shape latent vector and the shape's radar response, but as shown in Section A.1, we adjust our weights only according to the noise prediction for the shape component. For the following experiments, we use a different objective to only train the network to predict noise in the radar response space.

A.1 Objective Function

We condition our diffusion model on the latent shape vector $\mathbf{h}$ defined in Section 3.2, with goal of denoising the shape's radar response $R \in \mathbb{R}^{N_\alpha \times N_R}$. The UNet $\epsilon_\theta : x_t \rightarrow \epsilon$ predicts noise $\epsilon \in \mathbb{R}^{N_\alpha \times (N_R+1)}$ from the signal $x_t \in \mathbb{R}^{N_\alpha \times (N_R+1)}$, which is the concatenation of $\mathbf{h} \in \mathbb{R}^{N_\alpha \times 1}$ and $R \in \mathbb{R}^{N_\alpha \times N_R}$. However, instead of predicting noise for $\mathbf{h}$ as in inverse modeling, we only want our objective to train the model to predict noise for R. Denote the components of ϵ corresponding to $\mathbf{h}$ and R as $\epsilon = [\epsilon_h \ \epsilon_R]$. The loss function then becomes

$$L = \frac{1}{N_\alpha} \sum_{i=0}^{N_\alpha} (\epsilon_R - [\epsilon_\theta(x_t, t)]_R)^2. \tag{6}$$

During inference, since only R changes and $\mathbf{h}$ remains constant, the denoising process is defined

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\sqrt{1-\alpha_t}}{\sqrt{1-\bar{\alpha}_t}} \left[\mathbf{0} \ \ \epsilon_\theta(x_t, t)_R \right] \right). \tag{7}$$

The tensor $[\mathbf{0} \ \ \epsilon_\theta(x_t, t)_R]$ denotes $\epsilon_\theta(x_t, t)$ with ϵ_h set to 0. The denoised sample is defined $R = [x_0]_R$. We note that this objective does not take full advantage of the probabilistic nature of diffusion models, since the mapping between shapes and full observations is one-to-one. However, they still generally outperform GANs and other generative techniques in sample quality. We can see evidence of the high sample quality in Figure 3 and Table 3, which show the high accuracy of our model with standard sampling. The RMSE metric between two radar responses is a naive metric that overly emphasizes the more trivial sidelobe effects, so we also use the *Peak-Bin* metric used in Kohler et al. (2023), which calculates the locations of the k local maxima with the highest overall magnitude, then uses this to the measure the average difference of the range bins between the local maxima in the prediction and ground truth. Lower Peak-Bin is better, as with RMSE.

For DDIM sampling, we use $\eta = 0$ due to the deterministic nature of the radar response generation problem. We observe a sharp drop in sample quality compared to DDPM as shown in Table 3 and Figure 4. This is likely due to the high dimensionality of the radar response denoising space compared to the shape latent space as previously explored. Since DDIM's denoising trajectory is deterministic, it may miss critical modes which compound if the denoising space is high dimensional. This also suggests that, for the radar-to-shape problem, DDIM may have issues with denoising higher dimensional shape representations than what was presented in this work.

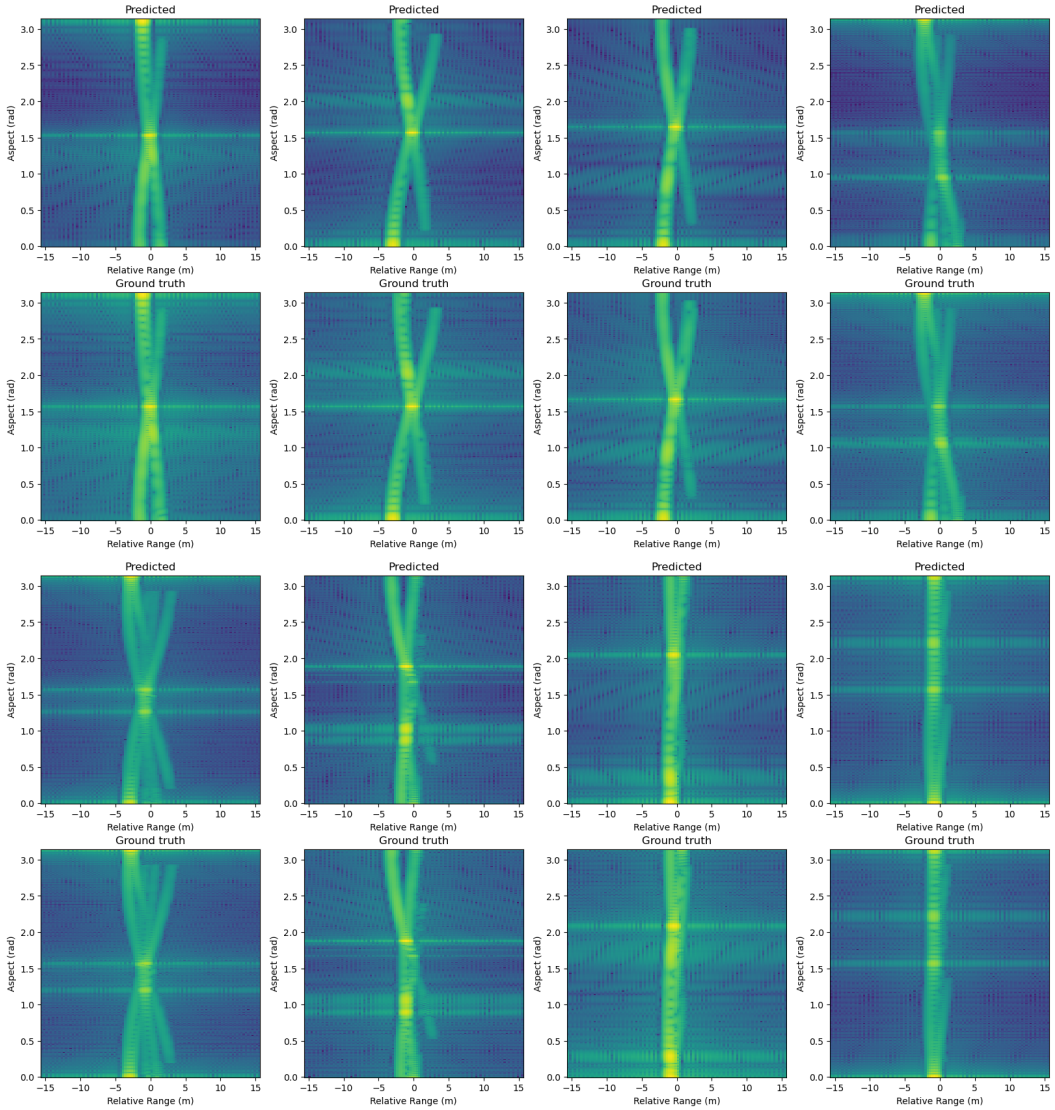

Figure 3: Test performance of the shape conditioned static pattern model in zero-noise setting with standard DDPM sampling. Two sets of four semantic comparison between generated static patterns (top) and ground truth static patterns (bottom) are shown, conditioned on heldout shapes. Approximations are semantically extremely accurate, with slight variations in the width and locations of speculars.

B Appendix: Diffusion Modeling Background

B.1 Reverse Process

We define β_t is the noise variance schedule for timestep t, and $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$ as the cumulative product of noise scales. The forward diffusion process is defined as the marginal distribution of x_t given the clean data x_0,

$$q(x_t \mid x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbf{I}). \tag{8}$$

The reverse process aims to denoise x_t back to x_0 and is parameterized as,

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)), \tag{9}$$

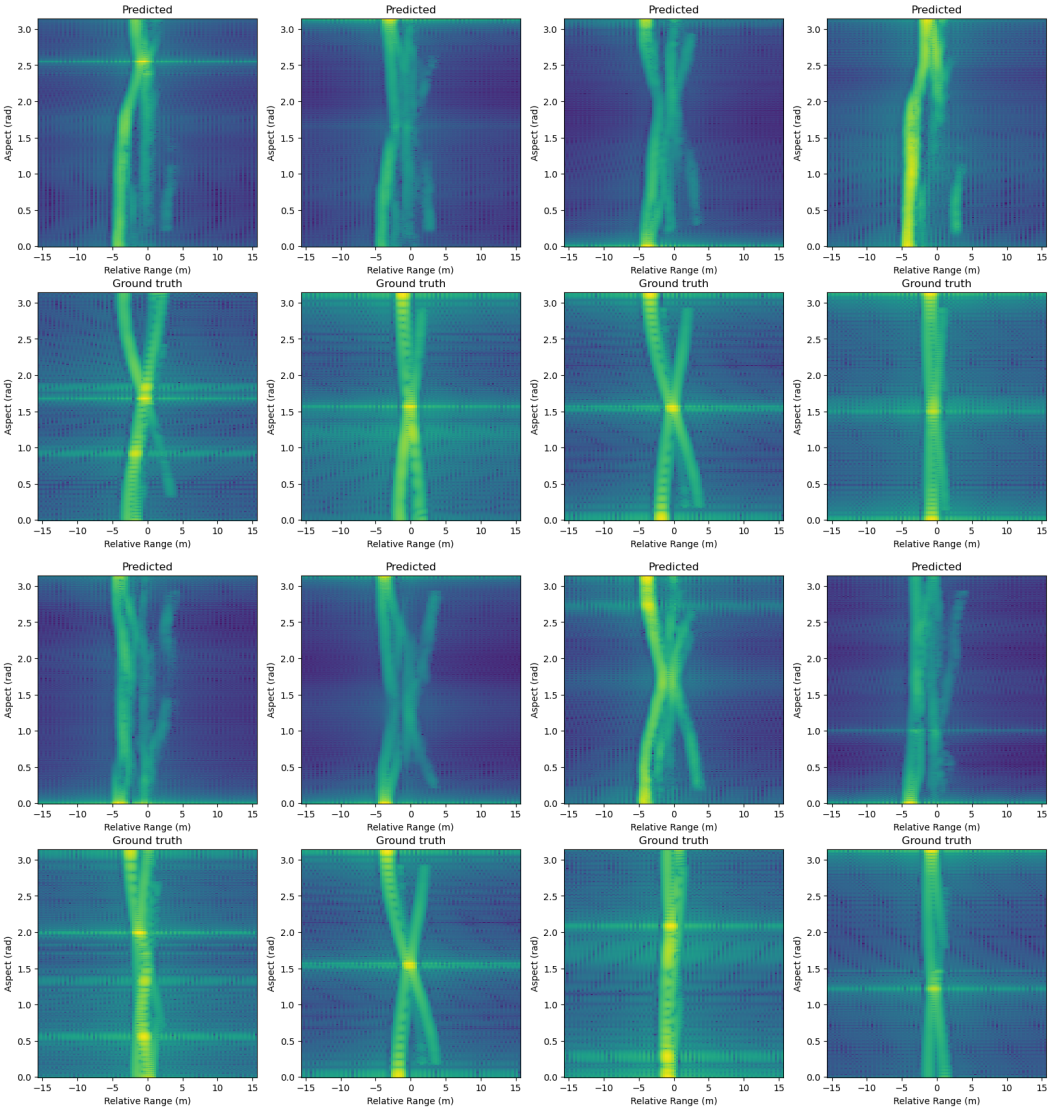

Figure 4: Test performance of the shape conditioned static pattern model in zero-noise setting with faster DDIM sampling (with $\eta = 0$). Two sets of four semantic comparison between generated static patterns (top) and ground truth static patterns (bottom) are shown, conditioned on heldout shapes. Approximations of the response have incorrect scatterer locations and incorrect specular locations throughout samples, showing DDIM sampling is not as effective in this higher dimensional denoising space, compared to the static pattern conditioned shape model in Table 1.

where $\mu_\theta(x_t, t)$ and $\Sigma_\theta(x_t, t)$ are learned parameters of the model. To reduce the model's complexity, the covariance matrix is typically set to a fixed diagonal matrix $\Sigma_\theta(x_t, t) = \beta_t I$, where β_t is a fixed noise variance.

B.2 Variational Lower Bound Loss

The training objective of DDPMs is to minimize the variational lower bound (VLB) of the negative log-likelihood. The VLB can be expressed as a sum of KL divergence terms over all timesteps,

$$L_{\text{VLB}} = \mathbb{E}_q \left[\sum_{t=1}^{T} L_t + L_0 + L_T \right],$$

where $L_t = \mathrm{KL}\big(q(x_{t-1} \mid x_t, x_0) \mid\mid p_\theta(x_{t-1} \mid x_t)\big)$, which measures how well the model approximates the posterior at each timestep t. $L_0 = -\log p_\theta(x_0 \mid x_1)$, which accounts for the final reconstruction of the clean image. $L_T = \mathrm{KL}(q(x_T \mid x_0) \mid\mid p(x_T))$, which ensures consistency with the prior distribution of noise at T.

In practice, L_0 is typically ignored since a separate decoder handles the reconstruction (Nichol & Dhariwal, 2021). L_T is ignored because the prior is a standard Gaussian, and its contribution is negligible during optimization. This simplifies the training objective to primarily minimizing L_t, which governs the intermediate denoising steps,

$$L_{\text{simple}} = KL(q(x_{t-1}|x_t, x_0)||p_\theta(x_{t-1}|x_t)). \tag{10}$$

B.3 Forward Process Posterior Distribution

To minimize L_{simple}, we can analytically compute the posterior distribution of $q(x_{t-1} \mid x_t, x_0)$ during the forward process as,

$$q(x_{t-1} \mid x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}(x_t, x_0), \tilde{\beta}_t \mathbf{I}),$$

where:

$$\tilde{\mu}(x_t, x_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t,$$

and $\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$. In practice, we usually fix the variance β_t to a schedule.

B.4 Reverse Process Simplified Mean Parameterization

Recall Equation 9, which we also need to analytically compute to minimize L_{simple}. The reverse process should similarly predict the mean $\mu_\theta(x_t, t)$, defined as,

$$\mu_\theta(x_t, x_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} x_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t,$$

where x_0 is the clean image, and $\bar{\alpha}_t$, β_t, and α_t are terms derived from the noise schedule. Since x_0 is unavailable during the reverse process, it is replaced using the reparameterization trick,

$$x_0 = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(x_t, t)\right), \tag{11}$$

where $\epsilon_\theta(x_t, t)$ is the model's prediction of the added noise at timestep t. Substituting this expression for x_0 into $\mu_\theta(x_t, x_0)$ allows the reverse process to operate entirely in terms of x_t and $\epsilon_\theta(x_t, t)$, without explicitly requiring x_0. Finally, we get,

$$\mu_\theta(x_t, x_0) = \mu_\theta(x_t, t) = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)).$$

B.5 Objective Function

The training objective is the variational bound on the negative log-likelihood, which simplifies to a weighted sum of KL divergences. For this simplified case, when we compare just the noise, the loss can be expressed as,

$$L_{\text{simple}} := \mathbb{E}_{t,x_0,\epsilon}\left[\left\|\epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t\right)\right\|^2\right],$$

which reduces to:

$$L_{\text{simple}} := \mathbb{E}_{t,x_0,\epsilon}\left[\|\epsilon - \epsilon_\theta(x_t, t)\|^2\right], \tag{12}$$

where $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ represents the added noise, and $\epsilon_\theta(x_t, t)$ is the model's prediction of the noise.

B.6 DENOISING DIFFUSION IMPLICIT MODELS

To define the DDIM process, we introduce its alternative formulation of the reverse distribution. Instead of sampling from the learned Gaussian transition in DDPMs, DDIM defines the reverse step deterministically as,

$$x_{t-1} = \sqrt{\bar{\alpha}t-1}x_0 + \sqrt{1-\bar{\alpha}t-1}\epsilon_\theta(x_t, t), \tag{13}$$

where x_0 is reconstructed using the same reparameterization trick as in DDPMs (Equation 11). The key insight of DDIM is that by choosing an implicit deterministic trajectory, we can generate high-quality samples in fewer steps.

The DDIM sampling process introduces a hyperparameter η that controls the balance between determinism and stochasticity. When $\eta > 0$, controlled noise is injected to retain stochasticity,

$$x_{t-1} = \sqrt{\bar{\alpha}t-1}x_0 + \sqrt{1-\bar{\alpha}t-1-\sigma_t^2}\epsilon_\theta(x_t, t) + \sigma_t z_t, \tag{14}$$

where $z_t \sim \mathcal{N}(0, \mathbf{I})$ and $\sigma_t = \eta\sqrt{(1-\bar{\alpha}_{t-1})/(1-\bar{\alpha}_t)}$. By adjusting η, DDIM can interpolate between standard DDPM sampling and a fully deterministic process.

The primary advantage of DDIM is that it allows for faster sampling by skipping intermediate steps while still producing high-quality samples.

C APPENDIX: RADAR MODELING AND SIMULATION

Radar 3d object modeling techniques depend on the relative size of the object l and the radar wavelength λ. Modeling in the Rayleigh region, when $l << \lambda$, is simplified as the entire object contributes to the RCS as a single-point scatterer. For modeling in the resonance regime, where $l \approx \lambda$, non-parametric numerical methods are often used. In this case, the object's geometry and materials all contribute to its RCS as a whole. Popular numeric methods include the Method of Moments Harrington (1993), the Fast Multipole Method Martinsson (2015), the Finite-Difference-Time-Domain Method Yee (1966), and the Transmission-Line-Matrix Method Hoefer (1985). These techniques provide exact solutions to Maxwell's equations, creating very accurate models of object scattering.

In most commercial and defense-related applications, objects are often illuminated using high-frequency radar waveforms. This leads to the optical regime, where $l >> \lambda$, where radar sensors provide a signal complementary to optical imaging techniques while providing a capability that, for automotive applications, penetrates fog and smoke with scattering particle sizes smaller than the wavelength. Additionally, scattering can often be reduced to a summation of discrete scattering centers taking advantage of Geometric Theory of Diffraction (GTD) Keller (1962). This allows the use of parametric, component-based, scattering models that reduce the radar modeling to simulating each component. Example of components are discrete points, spheres, rings, and triangles. The latter will be the focus of this paper.

We will utilize Physical Optics (PO) Balanis (2012) to model electromagnetic scattering from perfectly electrically conducting metallic meshes composed of triangular facets for high-frequency applications. Physical Optics simplifies the electromagnetic scattering problems by assuming that currents induced on a surface are the same as those in geometrical optics, neglecting edge effects. When applied to a meshed surface composed of triangular facets, PO calculates the scattered fields by integrating surface currents over each triangle, making it particularly effective for large, smooth objects Knott et al. (2004).

For a mesh composed of a collection of triangle facets, $\{T_t\}$, the simple operator that describes PO across the illuminated facet for a monostatic radar is,

$$F_t(\boldsymbol{u}, f) = \frac{ik}{2\pi}\int_{R^3} \langle n, \boldsymbol{u}\rangle e^{-i2k\langle \boldsymbol{u}, s\rangle} ds, \tag{15}$$

where n is the outward normal unit vector on the surface at the point s, $k = 2\pi f/c$ is the wave number, f is the wavefront frequency, c is the speed of light, and the observation unit vector is,

$$\boldsymbol{u} = (\sin\alpha\cos\phi, \sin\alpha\sin\phi, \cos\alpha), \tag{16}$$

for $\alpha \in [0, \pi]$ and $\phi \in [0, 2\pi]$. Given a set of illuminated facets dependent on the observation vector, $I_{\boldsymbol{u}}$, the mesh scattering response is the summation of individual triangle scattering responses,

$$F(\boldsymbol{u}, f) = \sum_{t \in I_{\boldsymbol{u}}} F_t(\boldsymbol{u}, f). \tag{17}$$

A common view for radar is the range profile typically calculated using a linear set of frequencies, such as those used in the common Linear Frequency Modulated (LFM) waveform Skolnik (1962). To emulate this waveform, define the set of frequencies, $\{f_i\}$, such that,

$$f_i \in [f_{\text{min}}, f_{\text{max}}], \text{ for } i = 1, \ldots, N.$$

We define the bandwidth of the signal as $B = f_{\text{max}} - f_{\text{min}}$. The range profile is a collection of range bins,

$$R(\boldsymbol{u}) = \{R_i(\boldsymbol{u})\}, \text{ for } i = 1, \ldots, N, \tag{18}$$

calculated by taking the Discrete Fourier Transform (DFT) of the scattering response across frequencies,

$$R_i(\boldsymbol{u}) = \sum_{m=0}^{N-1} F(\boldsymbol{u}, f_m) e^{-i2\pi im/N}. \tag{19}$$

Without zero padding or applying a taper function, the nominal bin size is the range resolution for the given bandwidth, $\Delta_r = c/(2B)$.

C.1 Illuminated Triangles

The set of illuminated triangles is used in Equation 17. Typically, this is done with ray tracing in order to properly handle shadowing and multi-bounce effects. Yet, ray tracing will often requires a significant amount of compute resources. For the purposes of this work we utilize a simple calculation to avoid this computation, namely,

$$t \in I_{\boldsymbol{u}} \;\; \texttt{iff} \;\; \langle -n_t, \boldsymbol{u} \rangle > 0.2.$$

This approach is simple and fast but does limit our ability to handle shadowing and multi-bounce interactions correctly.

C.2 Projected Area

Projected area for a mesh given an observation direction is used to calculate the IOU-IM metric. To calculate the projected area for a given triangle t, compute the area A_t and the normal n_t. Given an observation unit direction, $\boldsymbol{u}$, the projected area of a given triangle is

$$P_A^t(\boldsymbol{u}) = \langle n_t, \boldsymbol{u} \rangle A_t.$$

The total projected area is the sum of illuminated triangle areas projected along the observation direction,

$$P_A(\boldsymbol{u}) = \sum_{t \in I_{\boldsymbol{u}}} P_A^t(\boldsymbol{u}).$$

For the IOU-IM metric the set of illuminated triangles is defined by finding the triangles whose normal vector is parallel to the observation vector within some threshold,

$$I_S = \{t \mid \langle n_t, \boldsymbol{u} \rangle > 0.98\}.$$

Then the projected area of triangles contributing to specular responses is

$$P_S(\boldsymbol{u}) = \sum_{t \in I_S} P_A^t(\boldsymbol{u}).$$

D APPENDIX: DATASET GENERATION

Object samples are uniformly distributed with a minimum of one and maximum of $k = 5$ stacked frusta. Each sample could also have a hemisphere stacked onto either side, along the axis of symmetry. Since these shapes are roll-symmetric, we can parameterize them as a sequence of radial and body axis coordinates. There are a maximum of $k + 1$ coordinates that parameterize the frustum section of the object, and two more pairs that represent the radius of the optional side hemispheres. For diffusion modeling, we flatten and concatenate this parametrization into a vector with a fixed ordering of $(\mathbf{r}, \mathbf{z}) = [r_1, \ldots, r_{k+1}, z_1, \ldots, z_{k+1}, s_1, s_2] \in \mathbb{R}^{2k+4}$, where each $\mathbf{r}$ is the radial component of the frusta, $\mathbf{z}$ the axial (along the z axis) component, $\mathbf{s}$ the hemisphere radius, and k the maximum number of stacked frusta. When the shape parameter vector is shorter than the maximum length, or there is no hemipheres, we set the appropriate parameters to a padding value.

To calculate the range profile, $R(\boldsymbol{u})$, for each shape, we first generate a triangular mesh with the radial and axial parameters defining the surface, then calculate the scattering response using Equation 15. The illumination set for Equation 17 is set by choosing triangles such that $\langle -n, \boldsymbol{u} \rangle > 0.2$. We define a set of wavefront frequencies, $\{f_i\}$, to generate the range profile, where the center frequency is $f_c = 3e9$ Hz (S-Band) and the bandwidth is $B = 400e6$Hz. Given the set of frequencies we can apply the DFT as shown in Equation 19. To perform the DFT we utilize the Fast Fourier Transform (FFT) with enough zero padding to ensure a range bin size of 0.2m, while also applying a Hamming window function Blackman & Tukey (1958) to filter out sidelobes.

The dataset stores the real and imaginary components of the range profile across different observation vectors $\boldsymbol{u}$ defined in Equation 16. Since the objects are roll-symmetric, we do not need to create observations across roll angles ϕ, and therefore, we focus on sampling only across aspect angles α. The range profiles are stored as matrices such that $R(\boldsymbol{u}) \in \mathbb{R}^{N_\alpha \times N_R}$, where N_α and N_R represent the aspect and range dimensions, respectively. We use $N_\alpha = 360$ and $N_R = 100$. For training and testing, we perform min-max normalization on the magnitudes of the range profiles. We find the minimum and maximum parameters to be -152.606 and 29.327, respectively.

E APPENDIX: ABLATIONS FOR DDPM AND DDIM ACROSS NOISE AND OBSERVABILITY

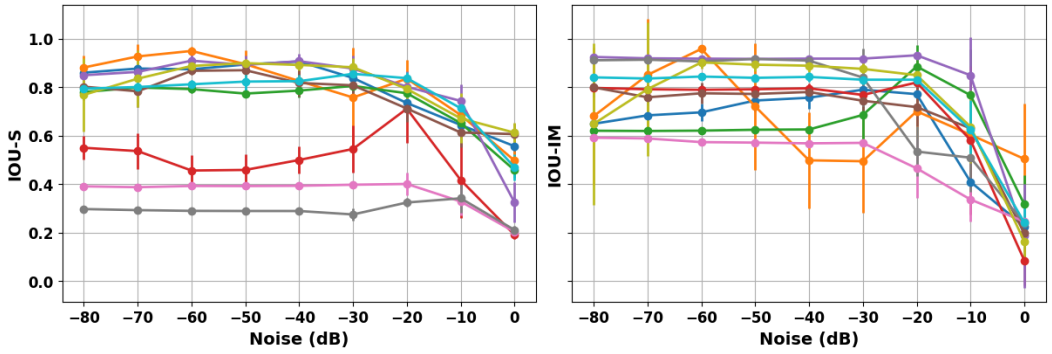

Figure 5: Performance metrics with standard DDPM sampling conditioned on fully observable radar responses for 10 different objects (indicated by color). (left) is the IOU metric for shape, (right) is the IOU metric for specular. Performance should degrade as noise masks signal.

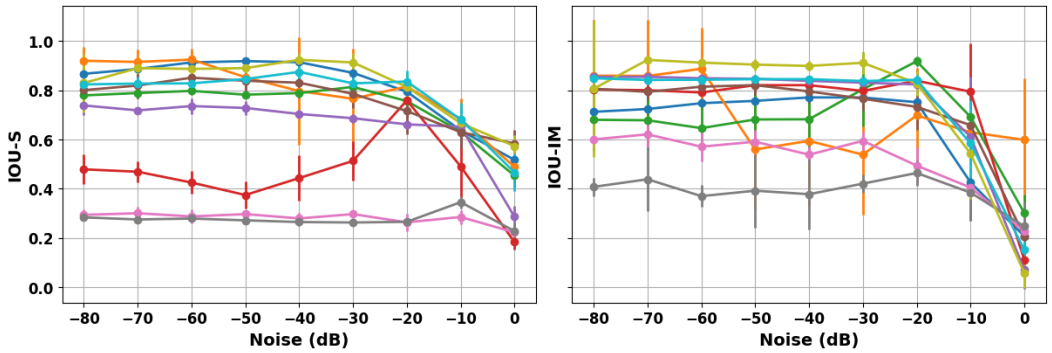

Figure 6: Performance metrics with DDIM sampling conditioned on fully observable radar responses for 10 different objects (indicated by color). (left) is the IOU metric for shape, (right) is the IOU metric for specular. Performance should degrade as noise masks signal. Performance is slightly lower than DDPM sampling, but sampling time is faster as discussed in Section 4.1.

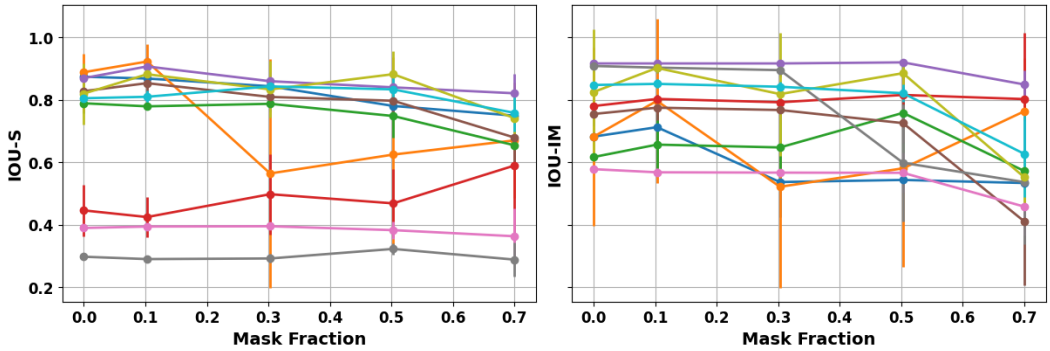

Figure 7: Performance metrics versus masking percentage with standard DDPM sampling for 10 different objects. The noise floor is set to -60 dB. (left) is the IOU metric for shape, (right) is the IOU metric for specular. Performance should degrade as noise masks signal.

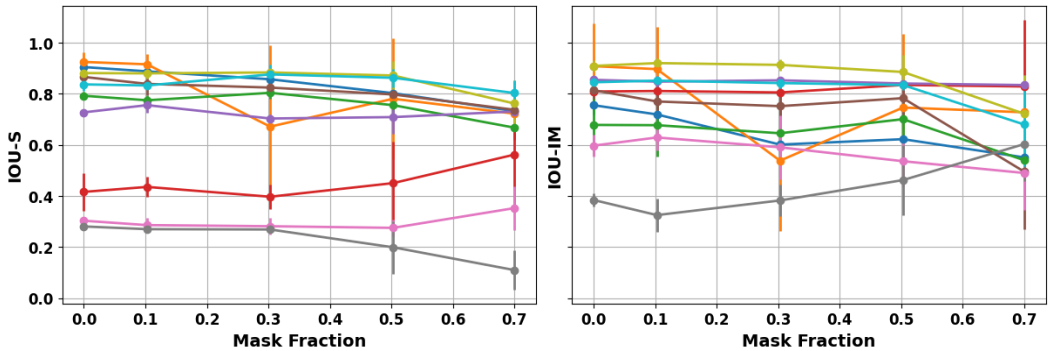

Figure 8: Performance metrics versus masking percentage with DDIM sampling for 10 different objects. The noise floor is set to -60 dB. (left) is the IOU metric for shape, (right) is the IOU metric for specular. Performance should degrade as noise masks signal. Performance is slightly lower than DDPM sampling, but sampling time is faster as discussed in Section 4.1.

F Appendix: Variance of generated samples across noise and observability

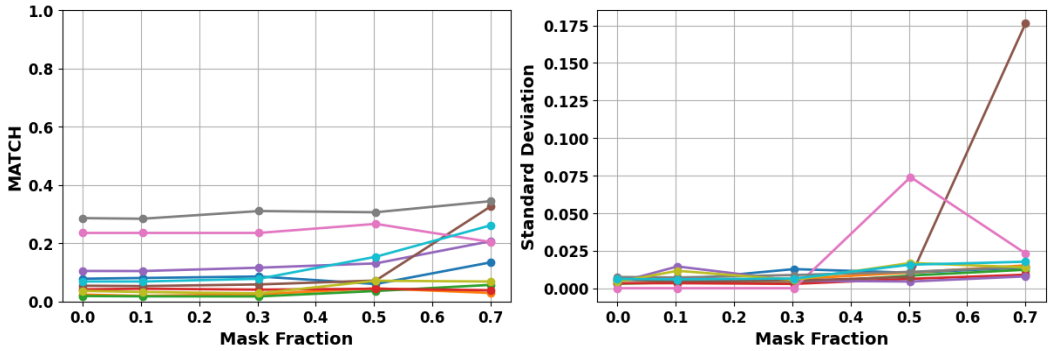

Figure 9: Performance metrics versus masking percentage with DDPM sampling for 10 different objects. The noise floor is set to -80 dB. (left) is the mean error of matching segments of the shape, (right) is the standard deviation of this metric across 20 samples for the diffusion model.

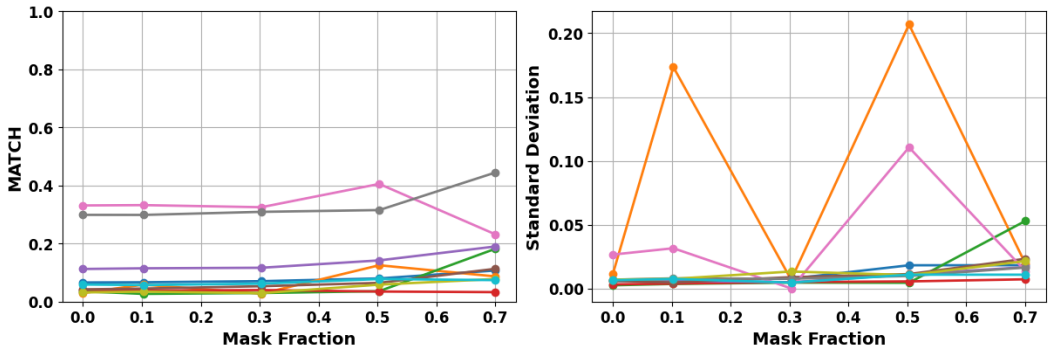

Figure 10: Performance metrics versus masking percentage with DDPM sampling for 10 different objects. The noise floor is set to -60 dB. (left) is the mean error of matching segments of the shape, (right) is the standard deviation of this metric across 20 samples for the diffusion model.

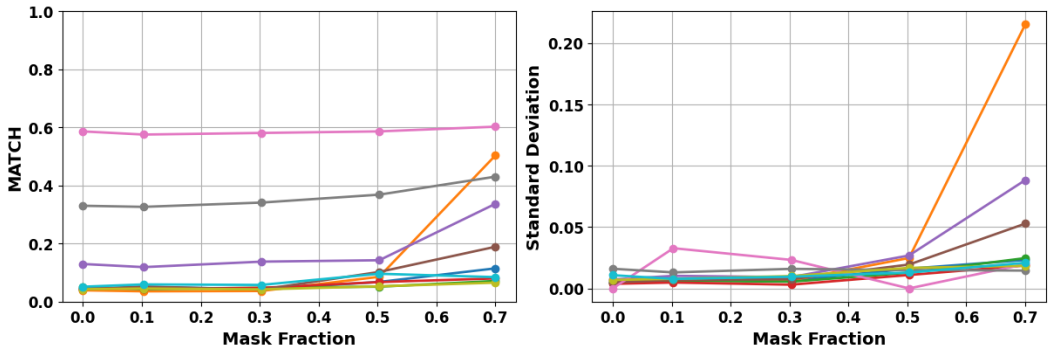

Figure 11: Performance metrics versus masking percentage with DDPM sampling for 10 different objects. The noise floor is set to -40 dB. (left) is the mean error of matching segments of the shape, (right) is the standard deviation of this metric across 20 samples for the diffusion model.

G Appendix: UNet Architecture Details

The network consists of an initial convolutional layer, 4 downsampling blocks, intermediate convolution and self attention, upsampling blocks, 1 resnet block, and a final convolution layer. Each downsampling block has 2 resnet blocks, groupnorm, linear self attention, residual connection, and a convolutional downsampling layer. Each upsampling block has 2 resnet blocks, groupnorm, linear self attention (Vyas et al., 2020), residual connection, and a convolutional upsampling layer. All convolutions are 1 dimensional, so the architecture shares weights across the aspect angle dimension. Intuitively, this means we can share information across aspects, since the shape is constant and only viewed from different angles. In practice, this convolutional downsampling acts as a radar response encoder. We treat the range dimension as channels of the signal. The total size of the model is 892 million trainable parameters.