Wave-Former: Through-Occlusion 3D Reconstruction via Wireless Shape Completion

Laura Dodds[1], Maisy Lam[1], Waleed Akbar[1], Yibo Cheng[1], Fadel Adib[1,2]
{ldodds, mllam, wakbar, yiboc, fadel}@mit.edu
[1] Massachusetts Institute of Technology, [2] Cartesian Systems

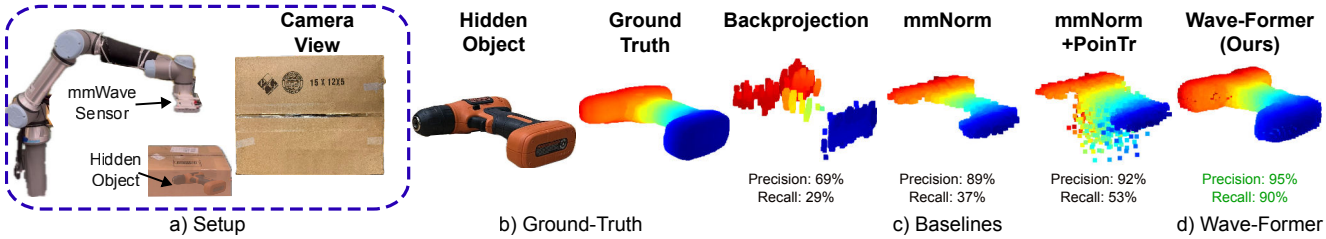

Figure 1. **Wave-Former:** a) Wave-Former can reconstruct objects hidden inside closed boxes using a millimeter-wave sensor. b) The ground-truth RGB and point cloud of a hidden object. c) Two state-of-the-art mmWave reconstruction methods (Backprojection and mmNorm) cannot reconstruct the entire object shape. Simply applying vision-based shape completion models (PoinTr) produces noisy reconstructions. d) By incorporating mmWave physical properties, Wave-Former can reconstruct the complete 3D shape of a hidden object from mmWave signals.

Abstract

We present Wave-Former, a novel method capable of high-accuracy 3D shape reconstruction for completely occluded, diverse, everyday objects. This capability can open new applications spanning robotics, augmented reality, and logistics. Our approach leverages millimeter-wave (mmWave) wireless signals, which can penetrate common occlusions and reflect off hidden objects. In contrast to past mmWave reconstruction methods, which suffer from limited coverage and high noise, Wave-Former introduces a physics-aware shape completion model capable of inferring full 3D geometry. At the heart of Wave-Former's design is a novel three-stage pipeline which bridges raw wireless signals with recent advancements in vision-based shape completion by incorporating physical properties of mmWave signals. The pipeline proposes candidate geometric surfaces, employs a transformer-based shape completion model designed specifically for mmWave signals, and finally performs entropy-guided surface selection. This enables Wave-Former to be trained using entirely synthetic point-clouds, while demonstrating impressive generalization to real-world data. In head-to-head comparisons with state-of-the-art baselines, Wave-Former raises recall from 54% to 72% while maintaining a high precision of 85%.

1. Introduction

Reconstructing the 3D geometry of a fully occluded object, such as one inside a closed box or beneath clutter, is an open challenge in computer vision. This capability could enable numerous applications spanning robotic manipulation, augmented reality, and shipping and logistics. However, optical sensors, such as cameras and LiDARs, are inherently limited in such scenarios. In contrast, radio-frequency signals, such as millimeter-wave (mmWave) signals, can traverse through occlusions and reflect off hidden objects. This ability has inspired recent interest in using these reflections for partial reconstruction of hidden objects [15, 17].

However, estimating the complete 3D shape of hidden objects with mmWave signals remains challenging. This limitation arises from the unique physical properties of mmWave reflections. Unlike visible light, which scatters diffusely off most surfaces, mmWave signals reflect primarily in a specular (mirror-like) manner [42] (Fig. 2). As a result, large portions of the surface will reflect signals away from the mmWave sensor, causing them to be effectively invisible and severely limiting the coverage. This issue is further compounded by the fact that mmWave signals are significantly noisier[1] and capture surface reflections at much lower resolutions, making complete 3D reconstruction even more difficult. For example, we show two state-of-the-art mmWave reconstruction methods in Fig. 1 c, Backprojection [15] and mmNorm [17], on a real-world, through-occlusion experiment. Both methods only capture a small portion of the object.

To overcome this, one natural approach is to apply existing vision-based shape completion models [66, 73, 77] to these partial mmWave reconstructions. Unfortunately, this strategy fails to produce reliable reconstructions, since these existing models are designed for visible-light sensors with high coverage and resolution, and do not account

[1]Compared to the Kinect depth camera, point clouds produced with mmWave signals have roughly 5× more noise [17, 36].

for the unique physical properties of mmWave reflections. For instance, we show an example in Fig. 1c where applying a state-of-the-art vision-based model, PoinTr [73], to a mmWave partial point cloud produces a highly inaccurate and noisy reconstruction.

In this paper, we present Wave-Former, the first complete mmWave 3D reconstruction method that can estimate the geometry of diverse objects through occlusions. Wave-Former bridges the gap between wireless sensing and modern shape completion techniques by embedding unique mmWave physics into the learning process. In Fig. 1d, we demonstrate that Wave-Former improves reconstruction quality on a real-world through-occlusion example, where the object is completely hidden from view.

Our core idea is a physics-aware training framework that embeds the physical characteristics of mmWave signals directly into the learning process. We train a transformer-based shape completion network entirely on synthetic 3D point clouds (e.g., ShapeNet [11]) while introducing a specularity-aware inductive bias that models the sparse, specular reflections of mmWave signals. Furthermore, our framework incorporates reflection-dependent visibility patterns, enabling the model to anticipate regions of the object that may be inherently unobserved. Finally, to enhance robustness against real-world noisy measurements, we adapt the model's loss function to jointly refine the noisy input and complete missing surfaces.

We then incorporate this design into a real-world inference process that transforms raw mmWave signals into several candidate surface hypotheses, applies our physics-aware shape completion model to each, and performs entropy-guided surface selection to output a single high-fidelity 3D reconstruction.

Our approach achieves state-of-the-art results on through-occlusion reconstruction. Across 61 diverse objects from the YCB dataset [10], Wave-Former achieves a 72% recall, an 18% improvement over the next-best baseline, while maintaining a high precision of 85%.

Contributions. We summarize our main contributions:

- **Physics-Aware mmWave Shape Completion.** We introduce the first through-occlusion mmWave 3D shape completion framework for diverse objects. It features a physics-aware training pipeline and real-world inference process, which together enable training entirely on synthetic data while showing 3D reconstruction on real-world data.
- **State-of-the-Art Performance.** On the real-world MITO dataset [16], we boost recall from 54% to 72% over existing mmWave reconstruction methods, (Backprojection [15], mmNorm [17], RMap [47]), while retaining 85% precision.
- **Ablation Study Against Vanilla Vision-Based Models.** We outperform vanilla vision-based completion models applied to mmWave partials, increasing recall by 12% and achieving the highest precision of 85%.

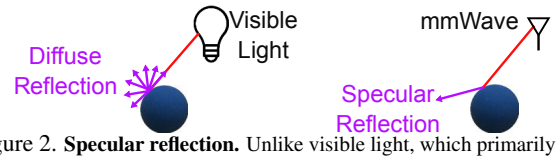

Figure 2. **Specular reflection.** Unlike visible light, which primarily scatters, mmWave signals undergo primarily specular (mirror-like) reflections.

2. Related Work

Shape Completion in Visible Spectrum. Shape completion in the visible spectrum has been a long-standing problem in computer vision, with methods evolving from early geometric approaches [5–7, 33] and encoder-decoder architectures [57, 74] to more recent transformer- and diffusion-based frameworks [8, 18, 32, 40, 44, 63, 66, 70, 73, 77, 78]. These models achieve remarkable performance on synthetic, camera, and LiDAR datasets, benefiting from point clouds with relatively high coverage and low noise compared to mmWave partial observations. However, these models are not directly applicable to the mmWave domain, as they fail to capture the physical characteristics of mmWave propagation, such as specularity, low spatial resolution, and low signal-to-noise ratios (SNRs). Consequently, vision-based completion methods struggle to recover reliable geometry from mmWave point clouds (as we demonstrate empirically in Sec. 5.3), underscoring the need for architectures that explicitly account for mmWave-specific physical properties.

mmWave Imaging. The vast majority of work in the mmWave space can image and reconstruct only the radar-facing surfaces of occluded objects or scenes, since they operate based on reflections that return from the surface. This includes volumetric [9, 15, 38, 39], interferometric [19–21, 23], surface-normal based [17] estimation approaches, as well as learning-based image enhancement [2, 24, 30, 31, 51] and scene reconstruction [22, 43, 47, 49, 62, 75]. This limitation is why airport mmWave scanners [46, 48] require a large infrastructure to rotate antennas around the entire human body, and require a large non-commercially available bandwidth to achieve high-accuracy depth estimation. In contrast, Wave-Former is able to perform complete 3D reconstruction with limited coverage of an object and only commercially available bandwidth through a physics-aware shape completion pipeline.

Additionally, some prior works have investigated mmWave full mesh reconstruction by focusing on a very limited set of objects (i.e., one to five categories) through strong data priors. For example, some works can only reconstruct humans due to their strong assumptions about human morphology [12, 67–69, 71, 72, 76]. Similarly, other work can only reconstruct up to five categories of objects (chairs, cars, robot arms, boxes, and desks) [53–55]. These works are fundamentally limited due to the fact that real-world and simulated mmWave data is extremely scarce. In contrast, Wave-Former eliminates the need for real-world

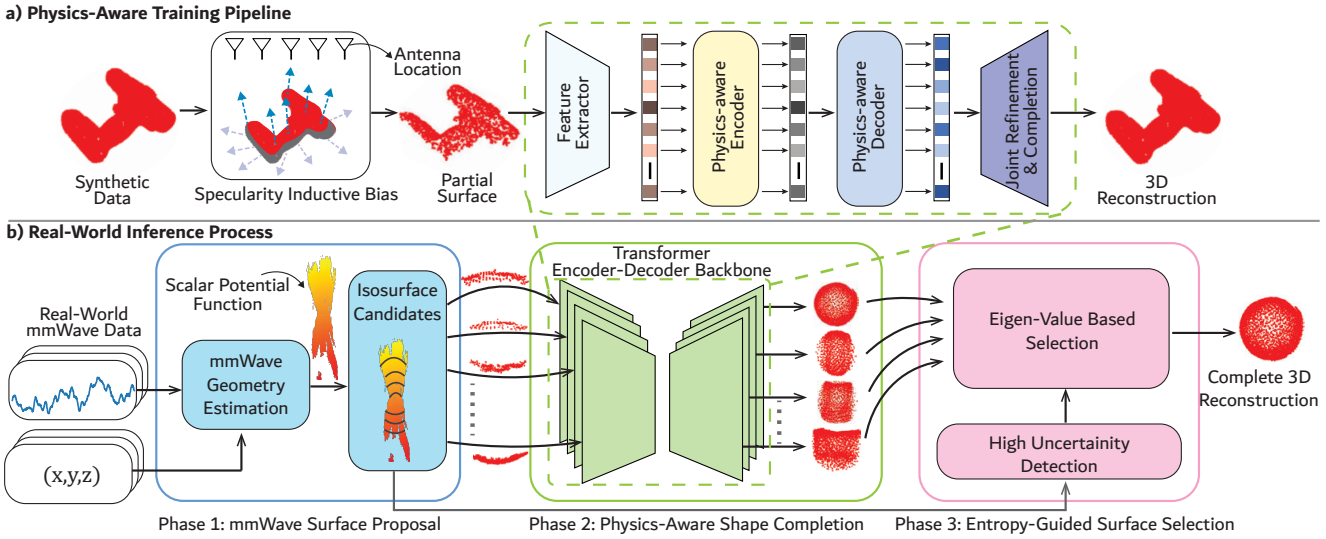

Figure 3. **mmWave Reconstruction Pipeline.** a) Wave-Former's physics-aware training pipeline incorporates physical properties through a specularity-aware inductive bias, reflection-dependent visibility, and joint refinement and completion framework to enable training on entirely synthetic data. b) Wave-Former's real-world inference process leverages a three-stage pipeline to reconstruct a complete 3D object from real mmWave signals.

mmWave training data by incorporating physical mmWave properties with readily-available synthetic 3D datasets, thus enabling generalization to diverse objects.

Alternative Through-Occlusion Perception Modalities. Recovering geometry through occlusions has been explored with various sensing modalities. For example, X-rays can penetrate opaque materials but use ionizing radiation, making them unsuitable for extended human exposure [60]. Additionally, acoustic or ultrasound is impractical for imaging through closed boxes due to the change in material properties (i.e., impedance mismatch) between air and cardboard [34].[2] Furthermore, thermal imaging [27, 29, 45, 50] can only image objects via heat, and thus is more suitable for imaging living things (due to their body temperature) rather than inanimate objects. Finally, around-the-corner laser [13, 25, 28, 35, 59] imaging can produce images in non-line-of-sight settings but cannot perceive through opaque obstacles, making them ill-suited for scenarios where objects are in a closed box or beneath clutter. In contrast, mmWave sensing combines material penetration and human safety, making it a practical modality for through-occlusion perception [15, 17, 24, 26, 26, 37, 41, 61, 69].

3. Preliminaries

Here, we describe the process of classical mmWave imaging at a high-level, and refer readers to [58] for more details.

Forming Images from mmWave Signals. Millimeter-wave (mmWave) radars transmit a frequency-modulated continuous wave (FMCW) signal which travels through occluding materials and reflects off hidden objects before being received back at the radar. These reflections carry information about how far the wave traveled, as a frequency shift, and the direction from which they arrived, as phase shifts across sensor locations. Together, this allows us to reconstruct a 3D image of the radar-facing surface of the object, where the image value $S(v)$ at each voxel v estimates the power reflected from that point in space [3]:

$$S(v) = \sum_{k=1}^{N} \sum_{t=1}^{T} h_k(t) e^{j2\pi(2||p_k - v||)/\lambda_t} \quad (1)$$

where p_k is the k$^{\text{th}}$ sensor location, λ_t is the wavelength of the t$^{\text{th}}$ sample (out of T), and $h_k(t)$ is the t$^{\text{th}}$ sample of the time-domain baseband received signal from the k$^{\text{th}}$ location.

Specularity. One primary limitation of these 3D mmWave images is that they only measure the radar-facing surface of an occluded object. Specifically, since mmWave signals experience primarily specular (mirror-like) reflections [42], many parts of an objects surface will reflect signals away from the sensor and thus are not present in the image. Wave-Former overcomes this limitation by bridging raw mmWave signals with physics-aware shape-completion.

4. Method

We design Wave-Former to perform high-fidelity, complete 3D reconstruction of hidden objects leveraging mmWave signals. Our core insight lies in a physics-aware training pipeline (Fig. 3a) to enable learning shape completion for mmWave properties using entirely synthetic data (Sec. 4.2). We additionally propose a real-world inference process (Fig. 3b) that (1) converts raw, mmWave signals into candidate partial surface representations (Sec. 4.3.1), (2) performs physics-aware shape completion (Sec. 4.3.2), and (3) identifies the optimal reconstruction (Sec. 4.3.3).

[2] Gel minimizes the mismatch when imaging through human tissue [4].

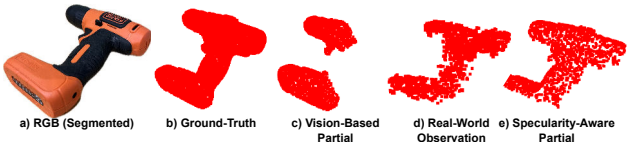

Figure 4. **Specularity-Aware Inductive Bias.** Specularity-aware inductive bias generates training partials (e) that resemble real mmWave visibility (d), unlike standard masking used in vision-based models (c).

4.1. Problem Definition

Our goal is to reconstruct a complete 3D point cloud $\hat{F}$ of a fully occluded object from a sequence of raw, complex-valued, time-domain mmWave measurements. Let $H \in \mathbb{C}^{N\times T}$ denote the collection of all measurements acquired from N known sensor positions ($P \in \mathbb{R}^{N\times 3}$), where each sensor position yields T temporal samples. Our pipeline learns a mapping $f_p : (H, P) \mapsto \hat{F}$.

4.2. Physics-Aware Training Pipeline

Training a shape-completion model for mmWave signals is fundamentally different from training one for visible-light sensors, such as cameras or LiDAR. Millimeter-wave returns are specular, anisotropic, and noisy, while existing completion models are trained on vision-like partials, which implicitly assume diffuse reflections and wide, uniform coverage, making them ill-suited for mmWave inputs.

To overcome this, our key insight is to embed mmWave physics directly into the training data and objective, creating a physics-consistent observation model that teaches the network physically plausible mmWave shape completion. By doing so, we enable training entirely on readily-available synthetic 3D datasets while achieving state-of-the-art generalization to real mmWave signals, as shown in Sec. 5.2.

4.2.1. Specularity-Aware Inductive Bias

Existing vision-based completion models inherently encode inductive bias consistent with visible light and incompatible with mmWave signals. This arises from their camera-like partial observations which assume diffuse reflections and wide coverage. To address this, we reformulate the inductive bias with physically consistent partials which emulate the specular reflections of mmWave signals. Specifically, for a given 3D model, we only include points in our partial which 1) produce a specular return and 2) lie on the radar facing surface (and are not further occluded by the object itself). Formally, we create partial observations O as:

$$O = \left\{ s_i \in F \,\middle|\, \theta_P(s_i) < \tau \cap V(s_i) = 1 \right\} \tag{2}$$

where s_i is a point in the full 3D cloud F. $\theta_P(s_i)$ denotes the smallest angular mismatch between s_i and all sensor positions in P. It is defined as $\theta_P(s_i) = \min_{k\in P} |\text{acos}\,(n_i \cdot u_{k,i})|$ where n_i is the normal at point s_i and $u_{k,i} = (p_k - s_i)/||p_k - s_i||$ is the unit vector pointing from s_i to p_k. $V(s_i)$ is 1 when a point s_i lies on a radar facing surface and 0 otherwise. [3]

[3] We compute this using Open3D's `hidden_point_removal`.

Applying this inductive bias during training teaches the model to expect mmWave-specific partial observations, guiding it to focus on physically plausible surfaces. This results in more accurate reconstructions and better generalization to real-world mmWave measurements (See Sec. 5.4).

4.2.2. Reflection-Dependent Visibility

Unlike optical sensors, mmWave visibility is strongly anisotropic: the measurable reflections depend on the incident angles and the reflection strength of an object. As a result, two objects with the same geometry may have markedly different visibility depending on their material properties. To model this behavior, we introduce a reflection-dependent visibility pattern that attenuates surface points according to physically-guided angular and material constraints. This replaces the common assumption of isotropic coverage and teaches the network that mmWave visibility is inherently non-uniform and angle dependent. Formally, for each specularity-aware partial O, we generate anisotropic partials O_A as:

$$O_A = \left\{ s_i \in O \middle| \theta_H(n_i) < \tau_H \cap \theta_V(n_i) < \tau_V \right\} \tag{3}$$

where $\theta_H = |\text{acos}\,(n_i \cdot [1,0,0])|$ and $\theta_V = |\text{acos}\,(n_i \cdot [0,1,0])|$ are the angles of the specular return from point s_i in the horizontal and vertical dimensions. τ_H and τ_V are parameters that can be changed to simulate different material properties. We train across a range of parameters to ensure robust mmWave completion.

Combined with our specularity-aware inductive bias, this visibility pattern further constrains the model to learn from physically plausible partial observations.

4.2.3. Joint Refinement and Completion

Existing vision-based shape completion models are designed for noise and resolution properties typical of a camera or LiDAR sensor, and therefore assume input partials can be directly concatenated with the reconstructed points. However, mmWave signals experience significantly higher noise levels (e.g., up to 5× higher than depth-camera point clouds [17, 36]) and a reduction in resolution. Thus, existing concatenation strategies would propagate significant distortions into the final reconstruction.

To address this, we introduce a joint refinement and completion method. Instead of preserving a noisy partial, we allow the model to simultaneously denoise and complete the object. To enable this, we incorporate noise during training to reflect the behavior of real-world mmWave signals. Then, we reformulate the loss function to allow the model to output the complete 3D shape (without concatenating the input), allowing it to reinterpret unreliable points rather than preserve them. Formally, we define the loss as the bi-directional chamfer distance between the complete, denoised output, $\hat{F}$ and the ground-truth shape, F:

$$\mathcal{L} = \frac{1}{|\hat{F}|} \sum_{s_i \in \hat{F}} \min_{g\in F} ||s_i - g|| + \frac{1}{|F|} \sum_{g\in F} \min_{s_i \in \hat{F}} ||g - s_i|| \tag{4}$$

This further improves Wave-Former's generalization to real-world measurements, as shown empirically in Sec. 5.4

Together, these three techniques establish a new physics-aware learning paradigm for mmWave shape completion, enabling state-of-the-art through-occlusion reconstruction capabilities.

Physics-Aware Encoder-Decoder. We integrate the above techniques into a transformer-based encoder-decoder for shape completion which maps a partial observation to an estimate of the complete 3D geometry: $\hat{F} = f_p(O)$. Built on the PoinTr backbone [73], our encoder–decoder is trained with a physics-consistent observation model (Eqs. 2 and 3) and a denoising–completion objective (Eq. 4), biasing the learned representation toward physically plausible mmWave surfaces. Our techniques enable us to train entirely with large, readily available synthetic 3D datasets [14, 52, 65], without relying on scarcely available real-world mmWave data. Critically, this allows Wave-Former to generalize to many different objects.

4.3. Real-World Inference Process

After training a physics-aware shape completion model, we next develop an inference process that converts raw, real-world mmWave signals into complete 3D reconstructions (Fig. 3b). Our pipeline follows three phases: mmWave surface proposal, physics aware shape completion, and entropy-guided surface selection.

4.3.1. Phase 1: mmWave Surface Proposal

We first transform raw mmWave measurements into a set of candidate partial surfaces which accurately capture the geometric information contained in the reflections.

Typically, partial mmWave point cloud estimation relies on thresholding a mmWave 3D power image (Eq. 1). However, this results in point clouds with significant erroneous points, as shown by the Backprojection baseline in Sec. 5.2. Instead, we leverage recent advancements in mmWave imaging [17] which can transform raw reflections into a space of geometrically consistent partial surfaces.

Formally, we compute a scalar potential function $f(v)$ consistent with mmWave reflections, where each isosurface [4] of this function is one possible partial surface [17]:

$$f(v) = \sum_{r \in R} N(v_r) \cdot d_r \tag{5}$$

where $N(v_r)$ is the estimated mmWave normal vector field, and $f(v_0) = 0$ at a reference voxel v_0, and R is the discrete path connecting v_0 and v. The value $f(v)$ represents the accumulated field integral along R, and d_r is the direction vector along R pointing from v_{r-1} to v_r

Unlike prior approaches which directly attempt to estimate the best candidate surface, we instead summarize the entire physically-plausible partial space by sampling candidate surfaces throughout this scalar function. This preserves available geometric information used during shape completion, and prevents critical errors caused by prematurely selecting the wrong surface. Formally:

$$C_{p,i} = \{v \mid |f(v) - I(i)| < \delta\} \tag{6}$$

where $C_{p,i}$ is the i$^{\text{th}}$ candidate partial surface, $I(i)$ is the value of the i$^{\text{th}}$ isosurface we sample, and δ is a small constant to account for numerical discrepancies.

4.3.2. Phase 2: Physics-Aware Shape Completion

Next, we convert the space of partial observations into a space of complete 3D reconstructions. We apply our physics-aware shape completion model, trained to incorporate physical properties as described in Sec. 4.2, to each candidate partial surface. This produces a set of physically-plausible candidate complete reconstructions, $C_{F,i}$:

$$C_{F,i} = f_p(C_{p,i}) \tag{7}$$

4.3.3. Phase 3: Entropy-Guided Surface Selection

In our final phase, we identify the optimal reconstruction from the space of candidate reconstructions.

Under high signal-to-noise conditions, we follow prior work [17] and select the optimal surface by comparing the simulated mmWave response of each candidate to the measured signals. However, when reflections are weak and the normal field contains significant noise, this often leads to erroneous surface selection. To handle these challenging cases, we introduce an entropy-guided selection strategy.

Detecting High-Uncertainty Reconstructions. We identify partial observations corrupted by noise, where weak reflections or noisy normals cause the scalar field $f(v)$ (Eq. 5) to produce irregular, vertically stacked voxels instead of smooth isosurfaces. We use the vertical spread of these voxels to serve as a proxy for reconstruction uncertainty. Formally, we classify reconstructions as high uncertainty when $\frac{|C^{xy}_{p,i}|}{|C_{p,i}|} < 0.6$, where $C^{xy}_{p,i}$ is constructed by projecting the observation $C_{p,i}$ to 2D and removing duplicate points.

Determining the Optimal Reconstruction. For high-uncertainty cases, we observe that physically consistent inputs produce continuous, locally planar reconstructions, whereas inaccurate or noisy inputs yield high-entropy point clouds dispersed over a larger spatial volume. We therefore select the candidate reconstruction with the lowest degree of local entropy.

We quantify entropy by sampling local neighborhoods [5] and computing the covariance eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$. Since planar regions have two dominant eigenvalues, whereas high-entropy regions have three of similar magnitude, we define an entropy score λ_3/λ_1 [64]. We then combine this score across all neighborhoods, and select the reconstruction with the smallest ratio:

$$i^\star = \arg\min_i \left(\texttt{pcntl} \left(\left\{ \frac{\lambda_{3,p}}{\lambda_{1,p}} \forall p \in C_{F,i} \right\}, 75 \right) \right) \tag{8}$$

[4] An isosurface is a surface of a constant value within a function.

[5] We use kNN with k=30.

Then, our final surface selection is $\hat{F} = C_{F,i^*}$. Together, these three phases together output a single, high-fidelity 3D reconstruction of a fully occluded object.

5. Experiments

5.1. Dataset and Implementation

Datasets: To train Wave-Former, we leverage three publicly available 3D object datasets: OmniObject3D [65], Toys4K-3D [52], and the Thingiverse subset of Objaverse [14] , totaling over 25K 3D object point clouds.

We evaluate Wave-Former on a real-world dataset of mmWave measurements, MITO [16]. The dataset contains measurements of 61 objects from the YCB dataset [10]. These items consist of a diversity of tasks (kitchen items, tools, food items, toys, etc), materials (wood, metal, cardboard, plastic, etc), and geometries (complex shapes, sharp edges, flat and curved surfaces, etc). This includes experiments in both line-of-sight and fully occluded settings for each object. More details can be found in [16].[6]

Training: We train our model using exclusively synthetic data, using the techniques in Sec 4.2. Each of the synthetic datasets are shuffled and split into 80% for training and 20% for testing. Additional details can be found in A.1.

Evaluation: For real-world evaluation, we directly input the raw mmWave radar data into Wave-Former without any additional masking or augmentations. We compare to the ground-truth 3D reconstructions provided in the MITO dataset [16]. Following prior work [66, 73], we align the reconstructed partial point clouds generated in Sec. 4.3.1 with the ground-truth point cloud for inference and evaluation. [7]

Evaluation Metrics: We adopt four standard metrics [56]: Chamfer Distance (CD), Precision, Recall, and F-Score, which jointly assess the accuracy and completeness of the reconstructions. We provide detailed definitions in A.2.

Baselines: We evaluate our approach against four state-of-the-art mmWave reconstruction baselines. We include detailed descriptions in A.3:

- *Backprojection [15]:* This is the classical and most widely used mmWave imaging approach. It is a first-principle method for volumetric mmWave reconstruction.
- *mmNorm [17]:* This is a recently-introduced state-of-the-art mmWave 3D reconstruction approach. It is also a first-principle method which estimates surface normal vectors and uses them to reconstruct the surface of the object
- *RMap [47]:* This is a state-of-the-art learning-based method for mmWave reconstruction, originally designed for scene level understanding. This model expects mmWave point cloud inputs, so we use mmNorm to create the input.
- *RMap (Finetuned) [47]:* We finetune RMap for object reconstruction on the same training data as Wave-Former.

[6] We note that the training datasets do not include the YCB meshes of the real-world objects which Wave-Former is evaluated on.

[7] Note that this alignment consist only of a vertical translation, which could be inferred using depth information from mmWave reflections [17].

5.2. Baseline Comparisons

5.2.1. Qualitative Results

To start, we compare qualitative results across Wave-Former and all four baselines using real-world measurements. Fig. 5 shows an isometric view of the ground-truth RGB (segmented) and point cloud, as well as the reconstruction for each method across several fully-occluded objects. Wave-Former is able to consistently reconstruct the complete shape of the object, even for complex geometries such as a power drill and clamp. In contrast, the baselines suffer from low accuracy, limited coverage, high noise, and, in some cases, even a complete a lack of discernible geometry. These results demonstrate the significant advancement of Wave-Former over prior state-of-the-art mmWave 3D reconstruction methods.

5.2.2. Quantitative Results

Table 1 reports the average chamfer distance, F-Score, precision, and recall across Wave-Former and all baselines. Notably, Wave-Former achieves a significant increase in recall, from 54% to 72% over the best-performing baseline, RMap (Finetuned), while maintaining a high precision of 85%. [8] Additionally, Wave-Former achieves the lowest chamfer distance of 0.069, compared to 0.18 for the best-performing baseline. This demonstrates the value of our techniques for enabling complete and high-accuracy 3D reconstruction of fully occluded objects.

5.3. Comparison to Vision-Based Shape Completion

We also evaluate whether state-of-the-art vanilla vision-based shape completion models can achieve high-accuracy mmWave 3D reconstruction. To produce radar-visible partial point clouds, we use the state-of-the-art mmWave reconstruction baseline with the highest precision per Table 1 (mmNorm), which is then fed into vision-based completion models. To ensure a fair comparison, all models are fine-tuned with the same synthetic datasets as used to train Wave-Former, while applying their own partial point cloud creation strategies. Table 2 reports the performance of Wave-Former compared to four state-of-the-art models. Wave-Former outperforms all models across all metrics, achieving an increase in recall from 60% to 72%, while also achieving the highest precision of 85%. This shows the importance of incorporating physical properties into the shape completion model. We also provide qualitative examples of Wave-Former compared to each vanilla model in A.4.

5.4. Ablation Study

Next, we analyze how different components of Wave-Former's design contribute to the overall performance. Table 3 shows the average and 75$^{\text{th}}$ percentile CD, as well as the marginal percent increase, of Wave-Former compared to three different partial implementations.

[8] mmNorm shows slightly higher precision since, as a first-principles method, it does not infer full geometry, leading to much lower recall.

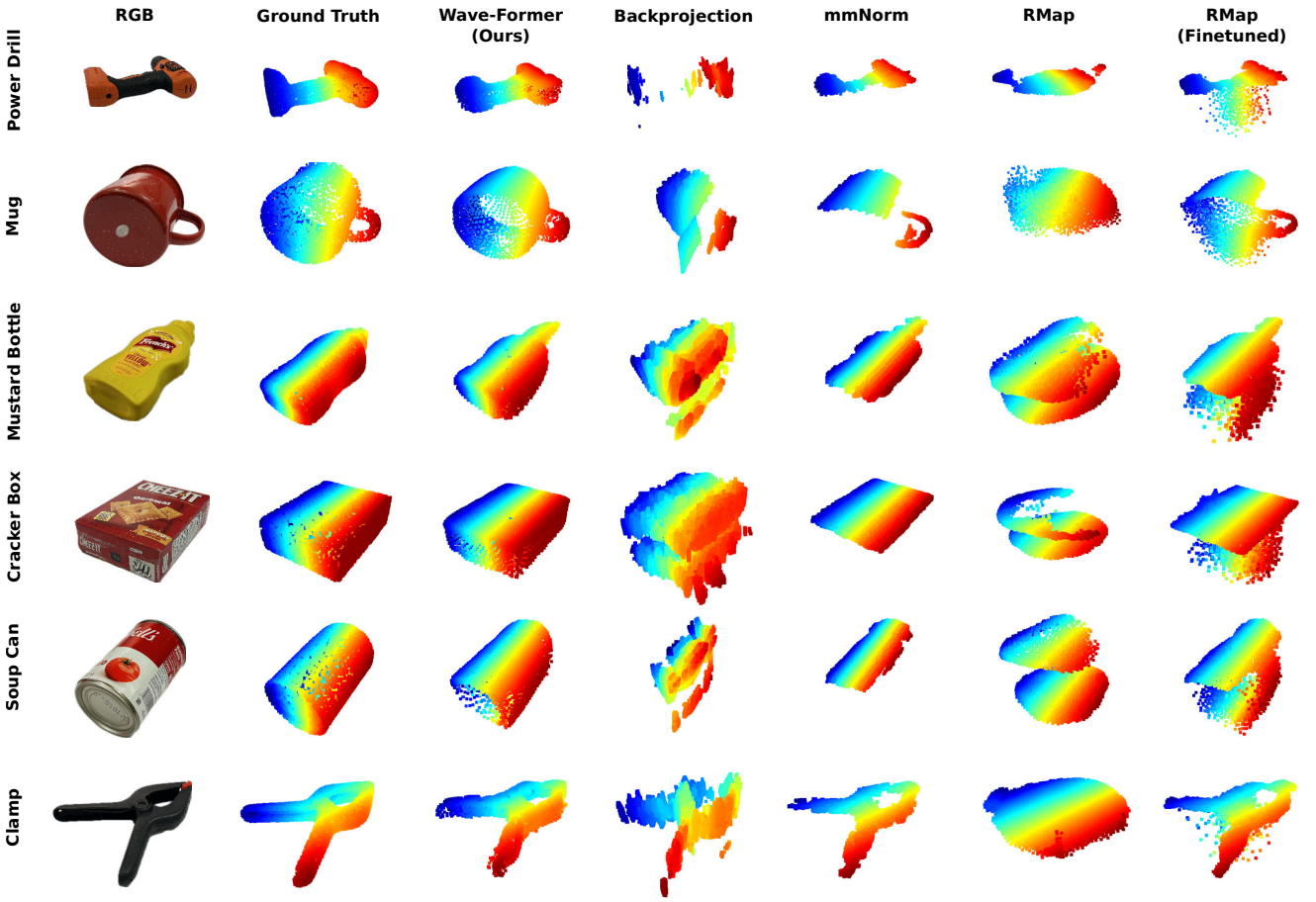

Figure 5. **Qualitative Results.** Visual comparison of mmWave 3D reconstruction on real-world, fully occluded objects. State-of-the-art baselines suffer from artifacts such as high noise and limited coverage, while Wave-Former consistently reconstructs shapes with high fidelity.

	CD ↓	FS ↑	Precision ↑	Recall ↑
Backprojection [15]	0.180	40%	43%	45%
mmNorm [17]	0.214	45%	**89%**	34%
R-Map [47]	0.273	23%	40%	17%
R-Map (Finetuned) [47]	0.330	62%	81%	54%
Wave-Former	**0.069**	**75%**	85%	**72%**

Table 1. Comparison between Wave-Former and state-of-the-art mmWave reconstruction baselines.

	CD ↓	FS ↑	Precision ↑	Recall ↑
mmNorm + PoinTr	0.104	62%	81%	53%
mmNorm + SnowFlakeNet	0.097	66%	80%	60%
mmNorm + SeedFormer	0.095	66%	83%	59%
mmNorm + PCN	0.138	58%	70%	56%
Wave-Former	**0.069**	**75%**	**85%**	**72%**

Table 2. Comparison with state-of-the-art vision-based vanilla shape completion models applied to the state-of-the-art mmWave reconstruction method.

Model	Physics	JRAC	Entropy	Average CD	Average % Inc.	75th Pcntl CD	75th Pcntl % Inc.
Wave-Former	✓	✓	✓	**0.069**	–	**0.072**	–
A		✓	✓	0.105	52%	0.120	67%
B			✓	0.115	10%	0.122	2%
C				0.116	1%	0.145	19%

Table 3. Ablation study of average and 75th percentile CD, as well as marginal % increase, for different components of Wave-Former.

First, when removing our specularity-aware inductive bias and reflection-dependent visibility (model A), we see a significant drop in performance, with the average chamfer distance increasing by 52% and the 75th percentile increasing by 67%. Next, when also removing our joint reconstruction and completion (model B), average chamfer distance increases by an additional 10%. Finally, when also removing our entropy-aware surface selection (model C), the 75th percentile continues to increase an additional 19%. We further demonstrate the benefit of our entropy-aware surface selection through qualitative examples in A.5, which show that for some objects, our selection process leads to drastically improved reconstructions. Together, these results demonstrate how each component of Wave-Former contributes to its overall performance.

5.5. Microbenchmarks

We performed microbenchmark experiments to understand the impact of various factors on Wave-Former. Additional microbenchmarks are available in the supplementary.

5.5.1. Impact of Occlusions

	Line-of-Sight		Fully Occluded	
	CD ↓	F-Score ↑	CD ↓	F-Score ↑
Backprojection [15]	0.177	40%	0.183	40%
mmNorm [17]	0.198	46%	0.231	44%
R-Map [47]	0.273	23%	0.272	24%
R-Map (Finetuned) [47]	0.091	64%	0.581	61%
Wave-Former	**0.058**	**78%**	**0.080**	**73%**

Table 4. Average performance in line-of-sight & fully occluded settings.

First, we compare the performance of Wave-Former and all baselines in line-of-sight and fully-occluded settings. Table 4 reports the average chamfer distance and F-Score. Notably, Wave-Former performs similarly in both settings, experiencing only a minor decrease in performance in fully-occluded settings. This demonstrates the ability of Wave-Former to not only reconstruct objects in visual line of sight, but also objects that are fully hidden from view.

5.5.2. Impact of Object Size

	Large	Medium	Small
Backprojection [15]	0.092	0.152	0.293
mmNorm [17]	0.078	0.201	0.361
R-Map [47]	0.136	0.239	0.452
R-Map (Finetuned) [47]	0.057	0.621	0.158
Wave-Former	**0.044**	**0.057**	**0.109**
Improvement	0.013	0.095	0.049

Table 5. Average CD for different object sizes.

Next, we evaluate the performance of Wave-Former and all baselines across objects of different sizes. We divide all objects into three categories based on their longest dimension: Large (>20 cm), Medium (10 cm-20 cm), and Small (<10 cm). Table 5 reports the average chamfer distance across different object sizes. As expected, Wave-Former performs best for large objects, with a chamfer distance of 0.04, compared to 0.06 for the best-performing baseline. As objects decrease in size, mmWave perception becomes increasingly challenging (See A.6). Despite this, Wave-Former outperforms all baselines, achieving the lowest CD across all sizes. This shows the value of our techniques for reconstructing various object sizes.

5.5.3. Impact of mmWave Coverage

Finally, we assess how varying levels of coverage affect the performance of Wave-Former. To start, we compute the percent of ground-truth points "covered" by the mmWave partial observation, where a point is considered covered if its nearest neighbor in the input lies within 0.08 m (for objects scaled to a unit sphere). We then define three categories of objects: Moderate (>36%), Challenging (18%-36%), and Extreme (<18%). [9] It is important to note that these percentages are significantly more challenging than most standard vision-based shape completion models, which only consider point clouds with 25% to 75% coverage [73, 77]. This is due to the specular nature of mmWave signals, which result in significantly lower coverage than visible light.

[9] Thresholds were selected for balanced category distribution.

	Moderate	Challenging	Extreme
Backprojection [15]	0.076	0.140	0.325
mmNorm [17]	0.064	0.155	0.434
R-Map [47]	0.126	0.221	0.481
R-Map (Finetuned) [47]	0.050	0.076	0.904
Wave-Former	**0.033**	**0.048**	**0.126**
Improvement	0.017	0.028	0.199

Table 6. Average CD for different initial object coverage.

Table 6 reports the average chamfer distance of Wave-Former and all baselines. In all categories, Wave-Former outperforms all baselines. As the objects become more challenging, Wave-Former outperforms the best baseline by an increasing amount, with a 0.2 improvement for the Extreme category. This highlights the benefit of our techniques for enabling through-occlusion 3D mmWave reconstruction even on extremely challenging examples. We also show qualitative examples from each category in A.7.

6. Conclusion & Future Opportunities

We presented Wave-Former, the first mmWave 3D reconstruction method that can operate on fully-occluded, diverse, everyday objects. Wave-Former leverages a novel physics-aware training pipeline to integrate core physical properties of mmWave signals into the completion process. Then, Wave-Former uses a real-world inference process which proposes candidate surfaces, leverages our physics-aware shape completion model, and selects a final reconstruction. These techniques allow Wave-Former to be trained on entirely synthetic 3D data while demonstrating impressive generalization to real-world signals.

Building on these contributions, we believe Wave-Former opens up many exciting directions for future research. For example, it would be interesting to use Wave-Former to develop novel downstream capabilities, such as fully-occluded robotic grasping, augmented reality with through-occlusion perception, or automatic package verification in shipping and logistics. Additionally, as the first paper to tackle the challenge of diverse mmWave 3D reconstruction, we hope that Wave-Former inspires future research towards reconstruction with even higher fidelity or using increasingly smaller radar scans. Finally, our techniques for training entirely on synthetic data may bring benefit to other mmWave tasks, such as fully-occluded classification, segmentation, and pose estimation.

More generally, we hope this work marks an important step towards generalizable through-occlusion perception.

Acknowledgments. We thank the anonymous reviewers and the Signal Kinetics group for their help and feedback. This research is sponsored by NSF (Award #2313234), Amazon MIT Science Hub, and the MIT Media Lab. Laura Dodds is supported by the Amazon Robotics Thriving Stars Fellowship.

References

[1] Ti iwr1443boost. https://www.ti.com/product/IWR1443#tech-docs, 2023. 12

[2] Aakriti Adhikari, Hem Regmi, Sanjib Sur, and Srihari Nelakuditi. Mishape: Accurate human silhouettes and body joints from commodity millimeter-wave devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(3):1–31, 2022. 2

[3] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. Capturing the human figure through a wall. *ACM Transactions on Graphics (TOG)*, 34(6):219, 2015. 3

[4] John E Aldrich. Basic physics of ultrasound imaging. *Critical care medicine*, 35(5):S131–S137, 2007. 3

[5] Marc Alexa, Johannes Behr, Daniel Cohen-Or, Shachar Fleishman, David Levin, and Claudio T Silva. Point set surfaces. In *Proceedings Visualization, 2001. VIS'01.*, pages 21–29. IEEE, 2001. 2

[6] Nina Amenta, Marshall Bern, and Manolis Kamvysselis. A new voronoi-based surface reconstruction algorithm. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 415–421, 1998.

[7] Nina Amenta, Sunghee Choi, and Ravi Krishna Kolluri. The power crust. In *Proceedings of the sixth ACM symposium on Solid modeling and applications*, pages 249–266, 2001. 2

[8] Burak Bekci, Nassir Navab, Federico Tombari, and Mahdi Saleh. Escape: Equivariant shape completion via anchor point encoding. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 6480–6489, 2025. 2

[9] Pingping Cai and Sanjib Sur. Millipcd: Beyond traditional vision indoor point cloud generation via handheld millimeter-wave devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6 (4):1–24, 2023. 2

[10] Berk Calli, Arjun Singh, James Bruce, Aaron Walsman, Kurt Konolige, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M Dollar. Yale-cmu-berkeley dataset for robotic manipulation research. *The International Journal of Robotics Research*, 36(3):261–268, 2017. 2, 6

[11] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. ShapeNet: An Information-Rich 3D Model Repository. Technical Report arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago, 2015. 2

[12] Anjun Chen, Xiangyu Wang, Kun Shi, Shaohao Zhu, Bin Fang, Yingfeng Chen, Jiming Chen, Yuchi Huo, and Qi Ye. Immfusion: Robust mmwave-rgb fusion for 3d human body reconstruction in all weather conditions. *arXiv preprint arXiv:2210.01346*, 2022. 2

[13] Wenzheng Chen, Simon Daneau, Fahim Mannan, and Felix Heide. Steady-state non-line-of-sight imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6790–6799, 2019. 3

[14] Matt Deitke, Ruoshi Liu, Matthew Wallingford, Huong Ngo, Oscar Michel, Aditya Kusupati, Alan Fan, Christian Laforte, Vikram Voleti, Samir Yitzhak Gadre, Eli VanderBilt, Aniruddha Kembhavi, Carl Vondrick, Georgia Gkioxari, Kiana Ehsani, Ludwig Schmidt, and Ali Farhadi. Objaverse-xl: A universe of 10m+ 3d objects. *arXiv preprint arXiv:2307.05663*, 2023. 5, 6

[15] Laura Dodds, Hailan Shanbhag, Junfeng Guan, Saurabh Gupta, and Haitham Hassanieh. Around the corner mmwave imaging in practical environments. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 1–15, 2024. 1, 2, 3, 6, 7, 8, 14

[16] Laura Dodds, Tara Boroushaki, Cusuh Ham, and Fadel Adib. Mito: A millimeter-wave dataset and simulator for non-line-of-sight perception, 2025. 2, 6, 12, 14, 16

[17] Laura Dodds, Tara Boroushaki, Kaichen Zhou, and Fadel Adib. *Non-Line-of-Sight 3D Object Reconstruction via mmWave Surface Normal Estimation*, page 445–458. Association for Computing Machinery, New York, NY, USA, 2025. 1, 2, 3, 4, 5, 6, 7, 8, 14

[18] Yi Du, Zhipeng Zhao, Shaoshu Su, Sharath Golluri, Haoze Zheng, Runmao Yao, and Chen Wang. Superpc: a single diffusion model for point cloud completion, upsampling, denoising, and colorization. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 16953–16964, 2025. 2

[19] Jingkun Gao, Yuliang Qin, Bin Deng, Hongqiang Wang, and Xiang Li. A novel method for 3-d millimeter-wave holographic reconstruction based on frequency interferometry techniques. *IEEE Transactions on Microwave Theory and Techniques*, 66(3):1579–1596, 2017. 2

[20] Jingkun Gao, Yuliang Qin, Bin Deng, Hongqiang Wang, and Xiang Li. Novel efficient 3d short-range imaging algorithms for a scanning 1d-mimo array. *IEEE Transactions on Image Processing*, 27(7):3631–3643, 2018.

[21] Jingkun Gao, Bin Deng, Yuliang Qin, Xiang Li, and Hongqiang Wang. Point cloud and 3-d surface reconstruction using cylindrical millimeter-wave holography. *IEEE Transactions on Instrumentation and Measurement*, 68(12):4765–4778, 2019. 2

[22] Ruixu Geng, Yadong Li, Dongheng Zhang, Jincheng Wu, Yating Gao, Yang Hu, and Yan Chen. Dream-pcd: Deep reconstruction and enhancement of mmwave radar pointcloud. *IEEE Transactions on Image Processing*, 2024. 2

[23] Borja Gonzalez-Valdes, Yuri Alvarez-Lopez, Jose Angel Martinez-Lorenzo, Fernando Las Heras Andres, and Carey Rappaport. On the use of improved imaging techniques for the development of a multistatic three-dimensional millimeter-wave portal for personnel screening. *Progress In Electromagnetics Research*, 138:83–98, 2013. 2

[24] Junfeng Guan, Sohrab Madani, Suraj Jog, Saurabh Gupta,

and Haitham Hassanieh. Through fog high-resolution imaging using millimeter wave radar. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11464–11473, 2020. 2, 3

[25] Otkrist Gupta, Thomas Willwacher, Andreas Velten, Ashok Veeraraghavan, and Ramesh Raskar. Reconstruction of hidden 3d shapes using diffuse reflections. *Optics express*, 20 (17):19096–19108, 2012. 3

[26] Chenming He, Chengzhen Meng, Chunwang He, Xiaoran Fan, Beibei Wang, Yubo Yan, and Yanyong Zhang. See through vehicles: Fully occluded vehicle detection with millimeter wave radar. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 740–754, 2024. 3

[27] Yingdong He, Hui Zhang, Edward Arens, Alexander Merritt, Charlie Huizenga, Ronnen Levinson, Andy Wang, Ali Ghahramani, and Ana Alvarez-Suarez. Smart detection of indoor occupant thermal state via infrared thermography, computer vision, and machine learning. *Building and Environment*, 228:109811, 2023. 3

[28] Felix Heide, Lei Xiao, Wolfgang Heidrich, and Matthias B Hullin. Diffuse mirrors: 3d reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3222–3229, 2014. 3

[29] Marina Ivašić-Kos, Mate Krišto, and Miran Pobar. Human detection in thermal imaging using yolo. In *Proceedings of the 2019 5th International Conference on Computer and Technology Applications*, pages 20–24, 2019. 3

[30] Sakila S Jayaweera, Sai Deepika Regani, Yuqian Hu, Beibei Wang, and KJ Ray Liu. mmid: High-resolution mmwave imaging for human identification. In *2023 IEEE 9th World Forum on Internet of Things (WF-IoT)*, pages 1–6. IEEE, 2023. 2

[31] Sakila S Jayaweera, Sai Deepika Regani, Yuqian Hu, Beibei Wang, and KJ Ray Liu. Hrnet: High-resolution neural network for human imaging using mmwave radar. *IEEE Internet of Things Journal*, 2024. 2

[32] Yoni Kasten, Ohad Rahamim, and Gal Chechik. Point cloud completion with pretrained text-to-image diffusion models. *Advances in Neural Information Processing Systems*, 36: 12171–12191, 2023. 2

[33] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, 2006. 2

[34] Yang-Hann Kim. *Sound propagation: an impedance based approach*. John Wiley & Sons, 2010. 3

[35] Ahmed Kirmani, Tyler Hutchison, James Davis, and Ramesh Raskar. Looking around the corner using ultrafast transient imaging. *International journal of computer vision*, 95:13–28, 2011. 3

[36] Gregorij Kurillo, Evan Hemingway, Mu-Lin Cheng, and Louis Cheng. Evaluating the accuracy of the azure kinect and kinect v2. *Sensors (Basel)*, 22(7):2469, 2022. 1, 4

[37] Haowen Lai, Gaoxiang Luo, Yifei Liu, and Mingmin Zhao. Enabling visual recognition at radio frequency. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 388–403, 2024. 3

[38] Jaime Laviada, Ana Arboleya-Arboleya, Yuri Álvarez, Borja González-Valdés, and Fernando Las-Heras. Multiview three-dimensional reconstruction by millimetre-wave portable camera. *Scientific reports*, 7(1):6479, 2017. 2

[39] Jaime Laviada, Miguel Lopez-Portugues, Ana Arboleya-Arboleya, and Fernando Las-Heras. Multiview mm-wave imaging with augmented depth camera information. *IEEE Access*, 6:16869–16877, 2018. 2

[40] Shanshan Li, Pan Gao, Xiaoyang Tan, and Mingqiang Wei. Proxyformer: Proxy alignment assisted point cloud completion with missing part sensitive transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9466–9475, 2023. 2

[41] Chris Xiaoxuan Lu, Stefano Rosa, Peijun Zhao, Bing Wang, Changhao Chen, John A Stankovic, Niki Trigoni, and Andrew Markham. See through smoke: robust indoor mapping with low-cost mmwave radar. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, pages 14–27, 2020. 3

[42] Jonathan S Lu, Patrick Cabrol, Daniel Steinbach, and Ravikumar V Pragada. Measurement and characterization of various outdoor 60 ghz diffracted and scattered paths. In *MILCOM 2013-2013 IEEE Military Communications Conference*, pages 1238–1243. IEEE, 2013. 1, 3

[43] Kai Luan, Chenghao Shi, Neng Wang, Yuwei Cheng, Huimin Lu, and Xieyuanli Chen. Diffusion-based point cloud super-resolution for mmwave radar data. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11171–11177. IEEE, 2024. 2

[44] Zhaoyang Lyu, Zhifeng Kong, Xudong Xu, Liang Pan, and Dahua Lin. A conditional point diffusion-refinement paradigm for 3d point cloud completion. *arXiv preprint arXiv:2112.03530*, 2021. 2

[45] Tomohiro Maeda, Yiqin Wang, Ramesh Raskar, and Achuta Kadambi. Thermal non-line-of-sight imaging. In *2019 IEEE International Conference on Computational Photography (ICCP)*, pages 1–11. IEEE, 2019. 3

[46] Douglas McMakin, David Sheen, Thomas Hall, Jonathan Tedeschi, and A Mark Jones. New improvements to millimeter-wave body scanners. *Proceedings of 3DBODY. TECH*, 2017. 2

[47] Ajay Narasimha Mopidevi, Kyle Harlow, and Christoffer Heckman. Rmap: Millimeter-wave radar mapping through volumetric upsampling. *arXiv preprint arXiv:2310.13188*, 2023. 2, 6, 7, 8, 14

[48] National Academies of Sciences, Medicine, Division on Engineering, Physical Sciences, National Materials, Manufacturing Board, Committee on Airport Passenger Screening, and Millimeter Wave Machines. *Airport passenger screening using millimeter wave machines: compliance with guidelines*. National Academies Press, 2018. 2

[49] Akarsh Prabhakara, Tao Jin, Arnav Das, Gantavya Bhatt, Lilly Kumari, Elahe Soltanaghaei, Jeff Bilmes, Swarun Kumar, and Anthony Rowe. High resolution point clouds from mmwave radar. *arXiv preprint arXiv:2206.09273*, 2022. 2

[50] EFJ Ring and Kurt Ammer. Infrared thermal imaging in medicine. *Physiological measurement*, 33(3):R33, 2012. 3

[51] Peyman Rostami, Hojatollah Zamani, Mohammad Fakharzadeh, Arash Amini, and Farokh Marvasti. A deep learning approach for reconstruction in millimeter-wave imaging systems. *IEEE Transactions on Antennas and Propagation*, 71(1):1180–1184, 2022. 2

[52] Stefan Stojanov, Anh Thai, and James M. Rehg. Using shape to categorize: Low-shot learning with an explicit shape bias. 2021. 5, 6

[53] Yue Sun, Zhuoming Huang, Honggang Zhang, Zhi Cao, and Deqiang Xu. 3drimr: 3d reconstruction and imaging via mmwave radar based on deep learning. In *2021 IEEE International Performance, Computing, and Communications Conference (IPCCC)*, pages 1–8. IEEE, 2021. 2

[54] Yue Sun, Zhuoming Huang, Honggang Zhang, and Xiaohui Liang. 3d reconstruction of multiple objects by mmwave radar on uav. In *2022 IEEE 19th International Conference on Mobile Ad Hoc and Smart Systems (MASS)*, pages 491–495. IEEE, 2022.

[55] Yue Sun, Honggang Zhang, Zhuoming Huang, and Benyuan Liu. R2p: A deep learning model from mmwave radar to point cloud. In *International Conference on Artificial Neural Networks*, pages 329–341. Springer, 2022. 2

[56] Maxim Tatarchenko*, Stephan R. Richter*, René Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas Brox. What do single-view 3d reconstruction networks learn? 2019. 6, 12, 13

[57] Lyne P. Tchapmi, Vineet Kosaraju, Hamid Rezatofighi, Ian Reid, and Silvio Savarese. Topnet: Structural point cloud decoder. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 383–392, 2019. 2

[58] Texas Instruments. The fundamentals of millimeter wave radar sensors. https://www.ti.com/lit/wp/spyy005a/spyy005a.pdf?ts=1762694207085, 2020. 3

[59] Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Moungi Bawendi, Diego Gutierrez, and Ramesh Raskar. Femto-photography: capturing and visualizing the propagation of light. *ACM Transactions on Graphics (ToG)*, 32(4):1–8, 2013. 3

[60] BF Wall, GM Kendall, AA Edwards, Simon Bouffler, CR Muirhead, and JR Meara. What are the risks from medical x-rays and other low dose radiation? *The British journal of radiology*, 79(940):285–294, 2006. 3

[61] Chao Wang, Feng Lin, Zhongjie Ba, Fan Zhang, Wenyao Xu, and Kui Ren. Wavesdropper: Through-wall word detection of human speech via commercial mmwave devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(2):1–26, 2022. 3

[62] Wenbo Wang, Wei Wang, Xixin Yu, and Weibin Zhang. C4rfnet: Camera and 4d-radar fusion network for point cloud enhancement. *IEEE Sensors Journal*, 2025. 2

[63] Guangshun Wei, Yuan Feng, Long Ma, Chen Wang, Yuanfeng Zhou, and Changjian Li. Pcdreamer: Point cloud completion through multi-view diffusion priors. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 27243–27253, 2025. 2

[64] Martin Weinmann, Boris Jutzi, Stefan Hinz, and Clément Mallet. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105: 286–304, 2015. 5

[65] Tong Wu, Jiarui Zhang, Xiao Fu, Yuxin Wang, Liang Pan Jiawei Ren, Wayne Wu, Lei Yang, Jiaqi Wang, Chen Qian, Dahua Lin, and Ziwei Liu. Omniobject3d: Large-vocabulary 3d object dataset for realistic perception, reconstruction and generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 5, 6

[66] Peng Xiang, Xin Wen, Yu-Shen Liu, Yan-Pei Cao, Pengfei Wan, Wen Zheng, and Zhizhong Han. Snowflake point deconvolution for point cloud completion and generation with skip-transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):6320–6338, 2023. 1, 2, 6

[67] Qian Xie, Qianyi Deng, Ta Ying Cheng, Peijun Zhao, Amir Patel, Niki Trigoni, and Andrew Markham. mmpoint: Dense human point cloud generation from mmwave. In *BMVC*, pages 194–196, 2023. 2

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

[69] Hongfei Xue, Qiming Cao, Yan Ju, Haochen Hu, Haoyu Wang, Aidong Zhang, and Lu Su. M4esh: mmwave-based 3d human mesh construction for multiple subjects. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 391–406, 2022. 2, 3

[70] Xingguang Yan, Liqiang Lin, Niloy J. Mitra, Dani Lischinski, Danny Cohen-Or, and Hui Huang. Shapeformer: Transformer-based shape completion via sparse representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 2

[71] Jiarui Yang, Songpengcheng Xia, Yifan Song, Qi Wu, and Ling Pei. mmbat: A multi-task framework for mmwave-based human body reconstruction and translation prediction. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8446–8450. IEEE, 2024. 2

[72] Jiarui Yang, Songpengcheng Xia, Zengyuan Lai, Lan Sun, Qi Wu, Wenxian Yu, and Ling Pei. mmdear: mmwave point cloud density enhancement for accurate human body reconstruction. *arXiv preprint arXiv:2503.02375*, 2025. 2

[73] Xumin Yu, Yongming Rao, Ziyi Wang, Zuyan Liu, Jiwen Lu, and Jie Zhou. Pointr: Diverse point cloud completion with geometry-aware transformers. In *ICCV*, 2021. 1, 2, 5, 6, 8, 12

[74] Wentao Yuan, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. PCN: Point completion network. In *2018 International Conference on 3D Vision (3DV)*, pages 728–737. IEEE, 2018. 2

[75] Ruibin Zhang, Donglai Xue, Yuhan Wang, Ruixu Geng, and Fei Gao. Towards dense and accurate radar perception via efficient cross-modal diffusion model. *IEEE Robotics and Automation Letters*, 2024. 2

[76] Mingmin Zhao, Yingcheng Liu, Aniruddh Raghu, Tianhong Li, Hang Zhao, Antonio Torralba, and Dina Katabi. Through-wall human mesh recovery using radio signals. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10113–10122, 2019. 2

[77] Haoran Zhou, Yun Cao, Wenqing Chu, Junwei Zhu, Tong Lu, Ying Tai, and Chengjie Wang. Seedformer: Patch seeds based point cloud completion with upsample transformer. *arXiv preprint arXiv:2207.10315*, 2022. 1, 2, 8

[78] Linqi Zhou, Yilun Du, and Jiajun Wu. 3d shape generation and completion through point-voxel diffusion. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5826–5835, 2021. 2

A.1. Additional Training Details

We include additional details on training Wave-Former.

Specularity-Aware Inductive Bias Calculation: Prior to training, we uniformly sample points from each synthetic object mesh (or directly use points from a point cloud if available), and center and scale the object to a unit-sphere, following prior work [73]. We also apply our specularity-aware inductive bias for each object at six different angles. First, we compute the points on a radar's top facing surface ($V(s_i)$ in Eq. 2) for each of the six angles, by applying Open3D's `hidden_point_removal` function. We apply this function from four viewpoints near the center of the simulated radar aperture, [10] and find the OR of these outputs to produce our final $V(s_i)$.

Then, we use the normals of the object to compute which points would produce a specular reflection towards our aperture. We define an aperture as a 2 m x 2 m square located 1 m above the object. Then, we find whether each normal points within this aperture. This can be combined with $V(s_i)$ to compute our partial input, O (Eq. 2).

During training, for each training sample, we randomly sample one of the six angles and apply the corresponding specularity-aware inductive bias to compute O. Then, following prior work, all point clouds are either randomly downsampled or resampled to 8,192 points to ensure a consistent input size.

Outlier Removal: To account for radar's low spatial resolution, we additionally perform outlier removal on the points in O. We cluster all points in O, such that all points in a cluster have a nearest neighbor in the cluster less than 0.03 m away (after scaling objects to a unit sphere). We discard any clusters with fewer than 100 points.

Reflection-Dependent Visibility: We also randomly sample τ_H and τ_V to apply our reflection-dependent visibility (described in Sec. 4.2.2). To ensure physically plausible values, we sample these values within the 3 dB and 6 dB beamwidth of the mmWave sensor's antenna [1].

[10] To ensure this function does not see "through" points to the opposite side of the object, we sample point clouds of 40,960 points, and downsample to 8192 after computing $V(s_i)$.

Object Noise: For each training sample, we simulate real-world sensing noise. We apply Gaussian noise, $\mathcal{N}(0, 0.025)$, to each of the x, y, and z coordinates of each point in the cloud. In addition, we add a uniform height bias, sampled from $\mathcal{U}(0, 0.01)$, to the z coordinate of all points to replicate the consistent vertical offsets observed in real radar measurements.

Training: We train on entirely synthetic data. To properly account for both low and high uncertainty objects, we train two models. For low-uncertainty objects, we follow prior work [73] and apply a chamfer distance loss function on both sparse and dense point clouds. For high uncertainty objects, we train a second model with reflection-dependent visibility (described in Sec. 4.2.2) and use a density aware chamfer distance loss function [8] on the dense point cloud, defined as:

$$d_{\mathrm{DCD}}(F,\hat{F}) = \frac{1}{2}\left(\frac{1}{|\hat{F}|}\sum_{s_i\in\hat{F}}\left(1-\frac{1}{n_{\hat{s_i}}}e^{-\alpha\|s_i-\hat{g}\|_2}\right) + \frac{1}{|F|}\sum_{g\in F}\left(1-\frac{1}{n_{\hat{g}}}e^{-\alpha\|g-\hat{s_i}\|_2}\right)\right) \quad (9)$$

where $\hat{g} = \min_{g\in F}||s_i - g||_2$, $\hat{s_i} = \min_{s_i\in\hat{F}}||g - s_i||_2$, and α is a scale factor. We define subsets $\hat{F}^{\hat{g}} \subseteq \hat{F}$ and $F^{\hat{s_i}} \subseteq F$, where points in the subsets query $\hat{g}$ and $\hat{s_i}$, respectively, and $n_{\hat{g}} = |\hat{F}^{\hat{g}}|$ and $n_{\hat{s_i}} = |F^{\hat{s_i}}|$.

Following prior work [73], we adopt the AdamW optimizer with an initial learning rate of $5{\times}10^{-4}$ and a weight decay of $5{\times}10^{-4}$. The learning rate follows a LambdaLR schedule, decaying by 24% every 21 epochs until reaching 2% of its initial value. Simultaneously, the Batch Normalization momentum is halved at the same interval to stabilize convergence.

The model is trained for 2000 epochs with a batch size of 8. All experiments are conducted on a machine running Ubuntu 22.04 equipped with Intel Xeon CPUs and an NVIDIA GeForce GTX 1080 Ti GPU.

Evaluation: After training the model, we evaluate on 123 real-world experiments from the MITO dataset [16] across 61 diverse, everyday objects.

A.2. Evaluation Metric Definitions

We define each metric used to evaluate Wave-Former.

Chamfer Distance (CD) measures the average closest-point error between a reconstructed point cloud $\hat{F}$ and the ground-truth point cloud F using the L2-norm [56]:

$$CD = \frac{1}{2}\left(\frac{1}{|\hat{F}|}\sum_{s_i\in\hat{F}}\min_{g\in F}\|s_i-g\| + \frac{1}{|F|}\sum_{g\in F}\min_{s_i\in\hat{F}}\|g-s_i\|\right) \quad (10)$$

where s_i and g are points in the reconstructed point cloud

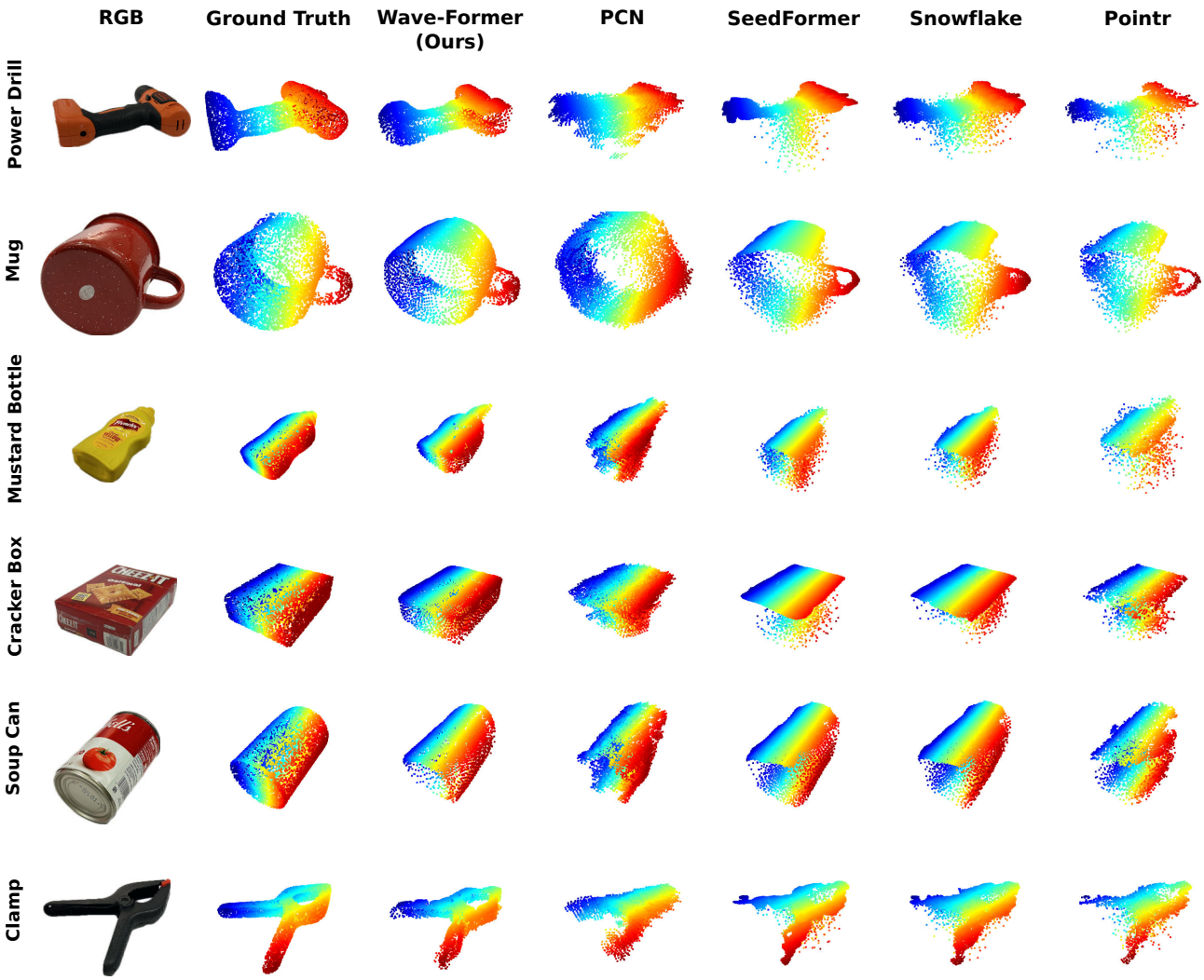

Figure A.6. **Qualitative Results.** Visual comparison of shape completion on real world data for select objects. Compared to existing state-of-the-art vision-based shape completion models, Wave-Former produces more accurate and complete object reconstructions from mmWave signals.

and ground-truth point cloud, respectively.

Lower chamfer distance indicates more accurate reconstructions.

Precision (P) measures how precise the reconstructed point cloud is compared to the ground truth. A point in $s_i \in \hat{F}$ is considered precise if its nearest neighbor $g \in F$ lies within a distance threshold of τ_F [56]:

$$P = \frac{1}{|\hat{F}|} \sum_{s_i \in \hat{F}} \left[\min_{g \in F} |s_i - g| < \tau_F \right], \tag{11}$$

τ_F is defined as 0.08 m in our implementation.

Larger precision indicates more accurate reconstructions.

Recall (R) measures the proportion of ground-truth points that are recovered by the reconstruction. A point in $g \in F$ is considered recalled if its nearest neighbor $s_i \in \hat{F}$ lies within τ_F [56]:

$$R = \frac{1}{|F|} \sum_{g \in F} \left[\min_{s_i \in \hat{F}} |g - s_i| < \tau_F \right] \tag{12}$$

Larger recall indicates reconstructions which have a higher coverage of the ground-truth.

F-Score (FS) captures the overall geometric agreement between the reconstructed and ground-truth point clouds by balancing precision and recall [56]:

$$\text{FS} = \frac{2 \times P \times R}{P + R} \tag{13}$$

A higher F-score indicates better alignment and completeness of the reconstructed shape. We compute the F-Score for each object independently, then average across all objects to compute the results in Sec. 5.

A.3. Baseline Description

We provide a more detailed description of our four baselines:

Backprojection [15]: This baseline is the most common approach for mmWave 3D reconstruction. It uses a first-principles method to reconstruct the 3D structure of objects from raw radar measurements. The raw mmWave measurements are first converted to a 3D image (Eq. 1), where the intensity of each voxel represents the power reflected from that position in space. This image is then converted to a point cloud by adding one point for each voxel with intensity values above a selected threshold. [11]

mmNorm [17]: This recent first-principles method allows for higher accuracy 3D reconstruction, but only recovers the top radar-facing surface of the object. It first estimates surface normals from the radar measurements and integrates them to form isosurfaces. Then, an isosurface optimization is performed to select a single surface reconstruction, by simulating reflections from each isosurface and comparing it with actual received signals.

RMap [47]: RMap is a learning-based method originally designed for scene understanding. It takes the radar point clouds and sensor trajectory as inputs, and employs a generative transformer network, UpPoinTr, which performs upsampling, denoising, and infilling on to produce higher-accuracy 3D point clouds.

RMap (Finetuned) [47]: In addition to employing RMap directly, we also finetune it on object level reconstructions using the same datasets used to train Wave-Former.

A.4. Qualitative Comparison to Vision-Based Shape Completion Models

In addition to the quantitative analysis in Sec. 5.3, we provide qualitative examples which compare Wave-Former to the combination of mmNorm and vanilla, state-of-the-art (SOTA) vision-based shape completion models. Fig. A.6 shows RGB photos and ground-truth point clouds for several objects, as well as Wave-Former's and each state-of-the-art model's output. Wave-Former consistently matches the ground truth, whereas state-of-the-art models produce noisy and inaccurate reconstructions. It is important to note that some state-of-the-art models accurately reconstruct the radar-facing surface of the object, such as the top of the cracker box in SeedFormer and SnowFlakeNet. This is because these models simply concatenate their input with their output reconstruction, and thus this accurate surface portion comes from mmNorm, as opposed to the vanilla model itself. This qualitative result reinforces our quantitative analysis to further show the benefit of integrating physics properties into the shape completion process.

[11] Following prior work [4], we select the threshold which maximizes the performance of this baseline across all objects.

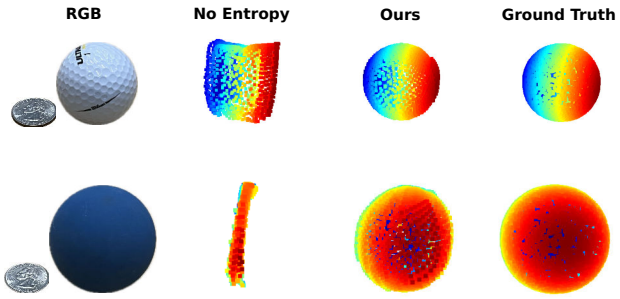

Figure A.7. **Analysis of Entropy-Aware Surface Selection.** Comparison of our method with and without entropy-guided surface selection. (quarter for scale)

A.5. Qualitative Comparison of Benefit of Entropy-Guided Surface Selection

In addition to the quantitative analysis in Sec. 5.4, we provide qualitative examples of the benefit of our Entropy-Guided surface selection (described in Sec. 4.3.3). We remove our entropy guided surface selection, and follow prior work [17] to select the surface which best matches the received signals, regardless of their estimated uncertainty.

Fig. A.7 shows an RGB and ground-truth point cloud of two objects. A quarter is included to show the scale of these objects. When we remove our entropy-guided surface selection, both objects produce reconstructions which drastically differ from the object's shape, due to selecting the wrong reconstruction. In contrast, with our entropy-guided surface selection, Wave-Former is able to select a more accurate reconstruction, recovering an accurate shape for both objects. This shows the benefit of our technique for ensuring accurate reconstructions.

A.6. Why are smaller objects harder to reconstruct?

The quality of a mmWave image and the resulting normal vector based surface reconstruction is determined in part by the (horizontal) resolution of the mmWave system. This resolution is itself determined by the aperture, or distance between first and last radar measurements, and is defined as [16]:

$$\delta_x = \frac{\lambda z_0}{2D_x}, \quad \lambda_y = \frac{\lambda z_0}{2D_y} \tag{14}$$

where δ_x and δ_y are the resolutions in the x and y dimensions, respectively. D_x and D_y are the apertures in the x and y dimensions, z_0 is the distance from the radar to the object, and λ is the starting wavelength of the FMCW sweep.

This resolution creates a spread of points in the horizontal (x and y) dimensions. For objects that are smaller in size, this resolution becomes a larger percentage of the object's size. This results in input observations with decreased accuracy, leading to larger distortions in the models output.

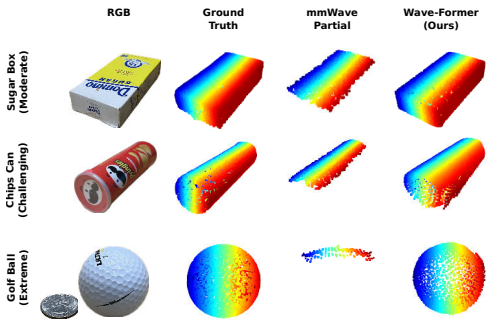

Figure A.8. **Varying Coverage Levels.** Three qualitative examples from three different coverage categories (Moderate, Challenging, and Extreme). We show the ground truth RGB (Quarter for scale) and point cloud, mmWave partial observation, and Wave-Former output.

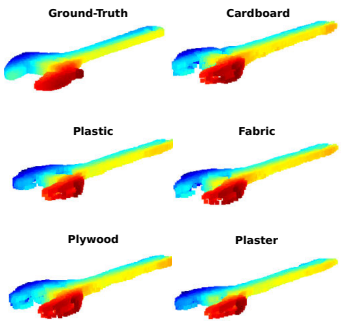

Figure A.9. **Varying Occlusion Materials.** Qualitative results across different occlusion materials including cardboard, plastic, fabric, 0.25" plywood, and 0.5" plaster.

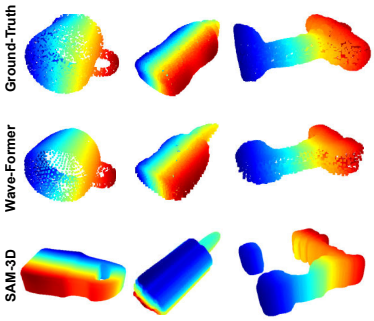

Figure A.10. **Comparison to SAM-3D.** Qualitative results comparing Wave-Former to the SAM-3D model operating directly on mmWave images.

Despite this, we are able to successfully reconstruct objects that are only ∼4.3 cm long, as shown by the two examples in Fig. A.7 (quarter for scale).

A.7. Qualitative Results for Different Coverage Levels

To complement the quantitative results shown in Sec. 5.5.3, we provide qualitative results for Wave-Former in three different mmWave coverage scenarios: Moderate, Challenging, and Extreme. Fig. A.8 shows the RGB and ground truth point clouds of objects in the three categories, along with the mmWave partial input and Wave-Former's output. For the moderate category, the shape of the object allows the entire top surface to be reconstructed in the partial input, resulting in a very high accuracy shape completion. For objects with more curvature, such as a cylinder can or spherical golf ball (for the Challenging and Extreme cases, respectively), a smaller portion of the object can be reconstructed in the partial input. In fact, only 6.3% of the object is visible in the final case of the golf ball. Despite this, Wave-Former is still able to output accurate reconstructions. This shows the benefit of designing a mmWave shape completion model for full object perception even in the presence of limited coverage.

A.8. Qualitative Results for Different Occlusion Materials

We provide an additional qualitative result which compares the performance of Wave-Former across different occlusion materials. For this experiment, the same item was placed in a fixed location and different occlusion materials were placed between the object and the mmWave radar. The same Wave-Former model is applied to each experiment without any retraining or fine-tuning. Fig. A.9 shows the ground-truth point cloud and Wave-Former's reconstruction across different occlusion materials including cardboard, plastic, fabric, 0.25" plywood, and 0.5" plaster. Wave-Former is able to produce accurate 3D reconstructions across all occlusion materials, with only a marginal performance degredation under thicker occlusions. This shows that Wave-Former can generalize to different occluding materials thanks to its physics-aware design.

A.9. Qualitative Comparison to SAM-3D

We further compare Wave-Former to SAM-3D [7], a state-of-the-art model trained to convert 2D RGB images to 3D reconstructions. We use the mmWave images (projected to 2D by averaging along the height dimension) as input to SAM-3D's online demo and manually select points in the image until the 2D segmentation covers the entire object. Fig. A.10 presents the ground-truth point clouds as well as Wave-Former's and SAM-3D's reconstructions. We note that SAM-3D fails to capture accurate geometric information. For example, the mug is reconstructed as a flat surface, the mustard bottle is cylindrical, and the power drill is missing accurate curvature and a large portion of the base. This is expected since SAM-3D is not trained for unique mmWave physics. In contrast, Wave-Former combines 3D mmWave information with a physics-aware model to produce accurate 3D shapes.

A.10. Discussion & Limitations

We believe that Wave-Former marks an important step towards complete mmWave 3D object reconstruction. Below, we discuss aspects of Wave-Former's design, current limitations, and avenues to overcome them.

Design Rationale. Wave-Former is intentionally designed to operate on mmNorm partial reconstructions rather than raw mmWave signals or normal fields. While an end-to-end, fully data-driven approach is appealing in principle, operating on raw signals would require ingesting orders of magnitude more data and a substantially larger model, necessitating far more training samples. At present, no large-scale object-level mmWave dataset exists, and generating

synthetic mmWave signals at scale is prohibitively expensive.[12] Instead, our design leverages existing large-scale 3D datasets by introducing a physically grounded intermediate representation aligned with mmWave sensing.

Common Limitations of mmWave. Wave-Former's primary limitations are common to all mmWave-based perception systems. For example, mmWave signals are unable to sense objects beneath metal or very dense occlusions (e.g., concrete, brick). Also, scenarios with very low SNR may suffer from poor initial reconstruction. Similar to any vision-based shape completion model, severely degraded inputs may lead to decreased reconstruction accuracy.

Performance on Small Objects. Similar to state-of-the-art mmWave reconstruction methods, Wave-Former performs worse for smaller objects (See Sec. 5.5.2). While Wave-Former's techniques allow it to outperform state-of-the-art methods across all object sizes, it would be interesting future work to further improve reconstructions of small objects. To do so, future work might investigate whether combining information across all isosurfaces might provide additional context to further improve reconstruction of small objects.

Accuracy of Isosurface Selection. Similar to small objects, Wave-Former's performance also decreases when only a smaller portion of the object is covered by initial surface reconstruction. This is expected, since the model has reduced information on the overall shape of the object. To overcome this, future work might investigate whether alternative robot scanning trajectories (e.g., spherical trajectories which might recover a portion of the sides of the object) can increase initial coverage of the object[13] and thus increase Wave-Former's complete reconstruction quality.

[12] Simulating our training set (25k objects at 6 views each) would require ∼8.5 years using existing state-of-the-art GPU simulators [16].

[13] Note that even with a more extensive scanning trajectories, it is infeasible to directly measure the entire object, since there is presumably a portion of the object occluded by the surface it rests on and perhaps occluded by other parts of the object or other objects.