Supplementary Material
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

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 [9]. 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, [1] 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].

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 [9] 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:

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

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 [9], we adopt the AdamW optimizer with an initial learning rate of 5×10^{-4} and a weight decay of 5×10^{-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

[1] 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 $V(s_i)$.

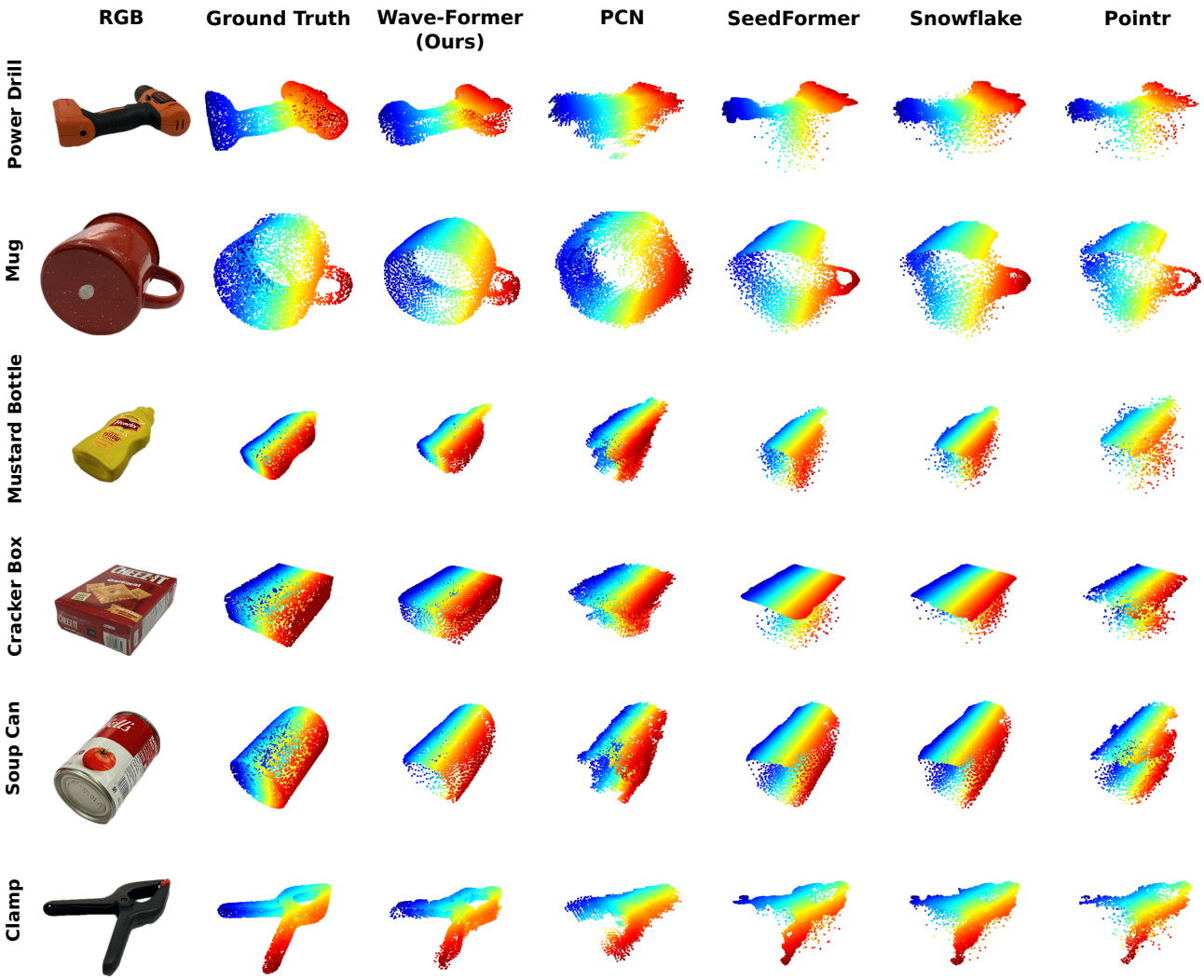

Figure A.1. **Qualitative Results.** Visual comparison of shape completion on real world data for select objects. Compared to existing state-of-the-art vision-based shape completion models, Wave-Former produces more accurate and complete object reconstructions from mmWave signals.

NVIDIA GeForce GTX 1080 Ti GPU.

Evaluation: After training the model, we evaluate on 123 real-world experiments from the MITO dataset [3] 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 [6]:

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

where s_i and g are points in the reconstructed point cloud 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 [6]:

$$P = \frac{1}{|\hat{F}|}\sum_{s_i \in \hat{F}} \left[\min_{g \in F} |s_i - g| < \tau_F\right], \tag{3}$$

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 [6]:

$$R = \frac{1}{|F|} \sum_{g \in F} \left[\min_{s_i \in \hat{F}} |g - s_i| < \tau_F \right] \tag{4}$$

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 [6]:

$$\text{FS} = \frac{2 \times P \times R}{P + R} \tag{5}$$

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 [2]: 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. [2]

mmNorm [16]: 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 [5]: 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) [5]: 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.1 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.

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

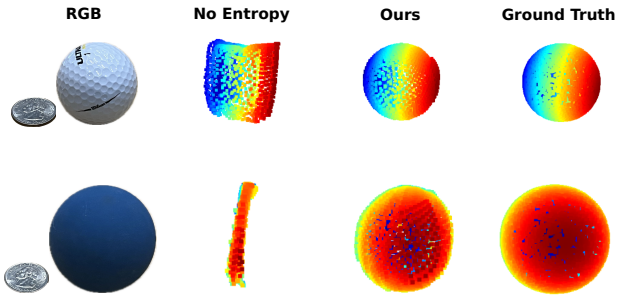

Figure A.2. **Analysis of Entropy-Aware Surface Selection.** Comparison of our method with and without entropy-guided surface selection. (quarter for scale)

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 [16] to select the surface which best matches the received signals, regardless of their estimated uncertainty.

Fig. A.2 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

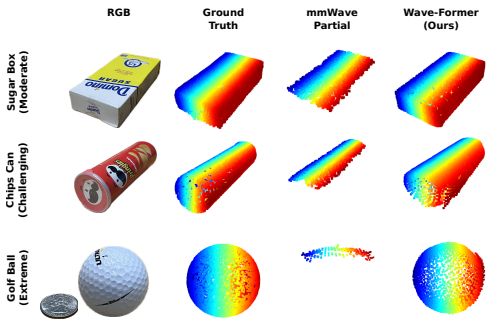

Figure A.3. **Varying Coverage Levels.** Three qualitative examples from three different coverage categories (Moderate, Challenging, and Extreme). We show the ground truth RGB (Quarter for scale) and point cloud, mmWave partial observation, and Wave-Former output.

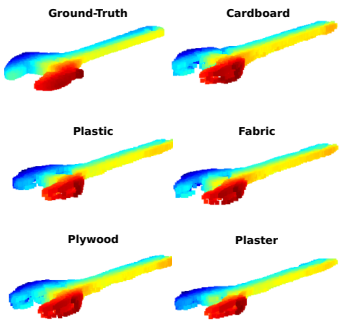

Figure A.4. **Varying Occlusion Materials.** Qualitative results across different occlusion materials including cardboard, plastic, fabric, 0.25" plywood, and 0.5" plaster.

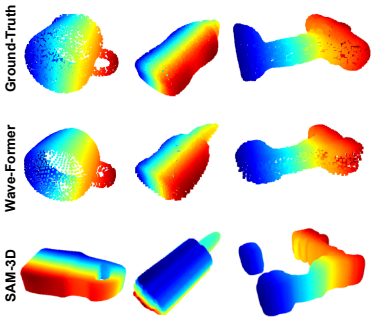

Figure A.5. **Comparison to SAM-3D.** Qualitative results comparing Wave-Former to the SAM-3D model operating directly on mmWave images.

between first and last radar measurements, and is defined as [3]:

$$\delta_x = \frac{\lambda z_0}{2D_x}, \quad \lambda_y = \frac{\lambda z_0}{2D_y} \tag{6}$$

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.

Despite this, we are able to successfully reconstruct objects that are only $\sim$4.3 cm long, as shown by the two examples in Fig. A.2 (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.3 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.4 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.5 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.[3] 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[4] and thus increase Wave-Former's complete reconstruction quality.

[3] Simulating our training set (25k objects at 6 views each) would require ∼8.5 years using existing state-of-the-art GPU simulators [3].

[4] 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.

References

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

[2] 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. 3

[3] Laura Dodds, Tara Boroushaki, Cusuh Ham, and Fadel Adib. Mito: A millimeter-wave dataset and simulator for non-line-of-sight perception, 2025. 2, 4, 5

[4] 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. 3

[5] Ajay Narasimha Mopidevi, Kyle Harlow, and Christoffer Heckman. Rmap: Millimeter-wave radar mapping through volumetric upsampling. *arXiv preprint arXiv:2310.13188*, 2023. 3

[6] Maxim Tatarchenko*, Stephan R. Richter*, René Ranftl, Zhuwen Li, Vladlen Koltun, and Thomas Brox. What do single-view 3d reconstruction networks learn? 2019. 2, 3

[7] SAM 3D Team, Xingyu Chen, Fu-Jen Chu, Pierre Gleize, Kevin J Liang, Alexander Sax, Hao Tang, Weiyao Wang, Michelle Guo, Thibaut Hardin, Xiang Li, Aohan Lin, Jiawei Liu, Ziqi Ma, Anushka Sagar, Bowen Song, Xiaodong Wang, Jianing Yang, Bowen Zhang, Piotr Dollár, Georgia Gkioxari, Matt Feiszli, and Jitendra Malik. Sam 3d: 3dfy anything in images. 2025. 4

[8] Junzhe Zhang Tai WANG Ziwei Liu Dahua Lin Tong Wu, Liang Pan. Density-aware chamfer distance as a comprehensive metric for point cloud completion. In *In Advances in Neural Information Processing Systems (NeurIPS), 2021*, 2021. 1

[9] 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