3D Shape Reconstruction from Autonomous Driving Radars

Samah Hussein[1]
samah.husseinyoussef@epfl.ch
Junfeng Guan[1,3]
jguan1019@gmail.com
Swathi Shree Narashiman[1]
swathi.narashiman@epfl.ch
Saurabh Gupta[2]
saurabhg@illinois.edu
Haitham Hassanieh[1]
haitham.alhassanieh@epfl.ch

[1] École Polytechnique Fédérale de Lausanne
Switzerland
[2] University of Illinois Urbana-Champaign
USA
[3] Bosch Research
USA

Abstract

This paper presents RFconstruct, a framework that enables 3D shape reconstruction using commercial off-the-shelf (COTS) mmWave radars for self-driving scenarios. RFconstruct overcomes radar limitations of low angular resolution, specularity, and sparsity in radar point clouds through a holistic system design that addresses hardware, data processing, and machine learning challenges. The first step is fusing data captured by two radar devices that image orthogonal planes, then performing odometry-aware temporal fusion to generate denser 3D point clouds. RFconstruct then reconstructs 3D shapes of objects using a customized encoder-decoder model that does not require prior knowledge of the object's bound box. The shape reconstruction performance of RFconstruct is compared against 3D models extracted from a depth camera equipped with a LiDAR. We show that RFconstruct can accurately generate 3D shapes of cars, bikes, and pedestrians. Check our demo video and more on our project page.

1 Introduction

Self-driving cars require precise and high-resolution 3D perception of the environment. The ability to recover accurate depth information, 3D dimensions, and the shapes of objects in the scene can be essential for improving decision-making for safer and more efficient autonomous driving. Today, self-driving cars rely mainly on cameras or LiDAR to image the environment. However, both sensors fail in adverse weather conditions such as fog, smog, snowstorms, and sandstorms [29, 30, 56] which is a foundational challenge against realizing the true vision of autonomous driving. Millimeter-wave (mmWave) radars have recently received significant interest in the autonomous driving industry due to their unique ability to operate in adverse weather conditions [27, 28, 37, 49] since mmWave signals can penetrate

(a) Radar Point Cloud (b) Baseline PCN (c) **RFconstruct (Ours)** (d) Ground Truth

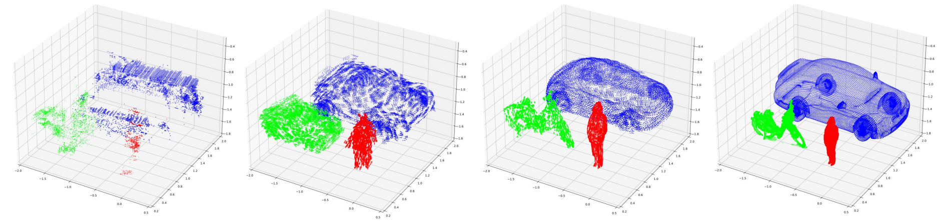

Figure 1: RFconstruct's ability to reconstruct 3D shapes from sparse specular point clouds.

through fog, smoke, sand, rain, etc. As a result, many radar manufacturers have extended the capabilities of their autonomous driving radars from 2D ranging to generating 3D point clouds [3, 40, 41, 47, 57]. Despite this significant progress, there is still a massive gap in the resolution compared to LiDAR and cameras, which prevents mmWave radars from providing detailed and interpretable shapes of objects.

Enhancing mmWave radar point clouds has been studied in past work. [34] uses LiDAR supervision to generate denser point clouds from radar heatmaps. [5, 18, 23] train neural networks to predict correct radar point clouds that match depth camera mappings of indoor environments. Other works focus on human pose estimation in controlled indoor scenarios[19, 59, 60]. However, none of the prior work on mmWave radars can accurately reconstruct the 3D shape of common autonomous driving objects (cars, bikes, pedestrians) from partial radar measurements. On the other hand, 3D shape completion has been studied in the context of LiDAR and depth cameras [6, 35, 48, 50, 54, 55]. However, these works cannot be directly used for mmWave radars due to the unique nature of RF signals and the resolution limitations of the radars.

Unlike LiDAR, Radar point clouds are incredibly sparse as can be seen in Fig. 1(a). The angular resolution of radars can be $100\times$ lower than cameras and LiDARs on the azimuth plane and $2000\times$ lower along the elevation plane [10, 14]. Moreover, in contrast to light, mmWave signals do not scatter as much and mainly reflect off surfaces [25]. Hence, cars are highly specular and act as a mirror reflector of mmWave signals and most reflections never trace back to the radar. This specularity makes certain portions of the car impossible to image, where a large portion of the car's surface is missing, as seen in Fig. 1(a). Finally, unlike vision, mmWave radar data is very scarce and highly dependent on the radar system that captures the data, making it very challenging to train deep neural networks.

In this paper, we introduce *RFconstruct*, the first system for 3D shape reconstruction from partial autonomous driving radar observations. It deploys COTS mmWave radars to generate interpretable 3D reconstructions of objects observed in autonomous driving scenarios, as shown Fig. 1(c). Enabling RFconstruct requires tackling the aforementioned domain-specific challenges like lack of high-resolution elevation information, sparsity of point clouds, signal specularity, and lack of training data.

RFconstruct overcomes these challenges through a holistic system design. First, to obtain elevation information, we build a radar system that uses two synchronized radars where one of the radars is flipped by 90°. This allows one radar to capture azimuth information and one radar to capture elevation information, which we then fuse together to obtain 3D radar point clouds. Second, we leverage odometry-aware temporal fusion to accumulate points over time to overcome specularity and generate denser point clouds. Third, we deploy a customized point-cloud completion neural network to generate complete, meaningful 3D representations from the sparse mmWave points. Finally, RFconstruct overcomes the scarcity of training

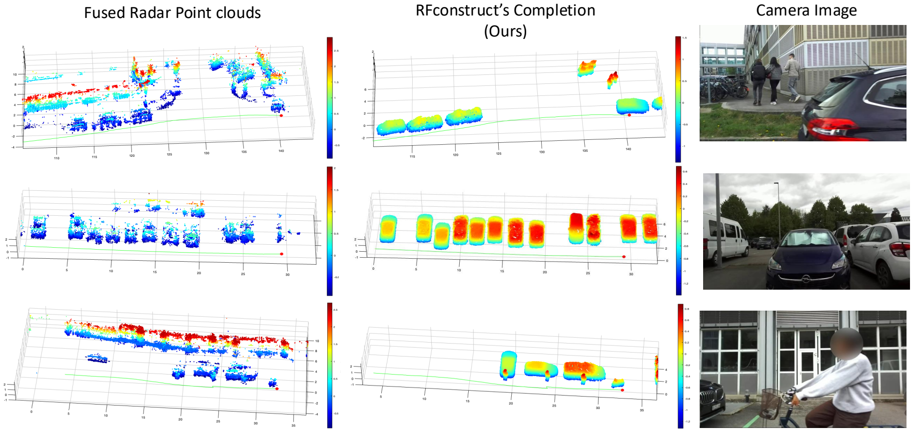

Figure 2: RFconstruct's reconstruction results. Note that we filter out background buildings and objects in our reconstruction and the color map corresponds to the height.

data[1] by designing a carefully augmented dataset that resembles partial radar inputs which we use for training before refining the results with real data collected by our radar system.

We build RFconstruct using two 77-GHz TI MMWCAS mmWave radars [46] which we mount on the side of a car as shown in Fig. 4 to collect a real-world dataset. Fig. 2 shows qualitative examples of RFconstruct's reconstruction from a video demo that can be found in the supplementary material. To evaluate RFconstruct quantitatively, we further collect a dataset of cars, humans, bikes, and motorbikes by mounting the radars on a TurtleBot3 [31] and measuring ground-truth from a full 3D scan of the objects using a depth camera and LiDAR-based reconstructions. We compare RFconstruct against the ground truth using standard metrics like Chamfer Distance (CD) and Earth Mover Distance (EMD) and show that it outperforms four baselines: PCN [55], AdaPointTr [54], ODGNet [6], and a class medoid baseline. Moreover, we show that RFconstruct's temporal fusion significantly outperforms Synthetic Aperture Radar techniques. We also present extensive microbenchmarks, ablation studies, and qualitative results.

This paper makes the following contributions:

- RFconstruct, to the best of our knowledge, is the *first* mmWave radar system capable of reconstructing 3D shapes and details of commonly found street-side objects (cars, pedestrians, bikes and motorbikes,) from partial observations of driving past the object.
- We present a novel system design that combines: (1) Radar fusion, (2) Odometry-aware temporal fusion, and (3) a radar shape completion deep neural network to reconstruct complete 3D shapes.
- We provide a new radar dataset composed of: (1) an augmented training dataset generated from ShapeNet-55 [1] to emulate mmWave radar point cloud imperfections such as specularity and noise, (2) a simulated radar dataset that captures radar characteristics like sinc leakage, (3) a real 3D mmWave radar dataset with equally good resolution in azimuth and elevation collected paired with depth camera and odometry data.
- We build, implement, and extensively evaluate a prototype of RFconstruct in real settings using COTS MIMO cascaded radars.

[1]Note that given our unique radar system, none of the online radar data sets can be used for training since none of them contain elevation information or odometry-aware temporally fused radar data.

2 Related Work

A. Millimeter-wave Radar Perception: Recent years have witnessed an increasing interest in mmWave radar perception for various applications, ranging from human posture tracking [17, 19, 59] to gesture recognition [20, 21]. On autonomous vehicles and autonomous mobile robots, mmWave radars have also been exploited for odometry [2, 24], mapping [8, 23, 34], object detection [4, 9, 11, 13, 14, 22, 26, 27, 28, 32, 33, 58], and scene flow estimation[10].These works only output high-level abstract features for specific end goals, such as semantic segmentation of radar point clouds [32, 38, 39], coordinates of objects [14], 2D BEV bounding boxes [9, 11, 26, 27, 58], and 3D bounding boxes [4, 22, 28, 33]. In contrast, RFconstruct aims to reconstruct 3D shapes of objects, which contain high-frequency details and contextual and perceptual information than the abstracted features.

B. 3D Reconstruction with mmWave Radars: There are prior works that try to reconstruct 3D shapes, but they focus on reconstructing 3D meshes of human bodies [7, 36, 51, 61]. They are specifically designed and trained for human targets and cannot generalize to other types of objects. [15, 18] create 2D depth maps of cars and indoor buildings from radar heatmaps. However, when converting depth maps to 3D point clouds, they suffer from common inaccuracies and artifacts and are incomplete because of occlusion. Sun et al. [43, 44, 45] take a step further, combining multiple GAN-generated depth maps from multiple views to generate a complete point cloud and feeding it to another generative model to reconstruct 3D shapes. These methods, however, have only been shown to work on cars and struggle with overfitting and insufficient input information due to requiring two stages of deep learning (Radar to depth map and depth map to 3D shape).

C. Point-Cloud Completion: Learning-based methods have become the current research trend in point cloud completion, and they can be further categorized into voxel-based and point-based methods. The voxel-based method entails the voxelization of the disordered point cloud, followed by the use of the voxelized 3-D model to accomplish shape completion [48, 50]. Point feature-based approaches leverage deep neural network architectures designed to work directly with point clouds like PointNet [35]. RFconstruct builds on top of PCN [55], which has an encoder-decoder architecture and uses the strategy of coarse-to-fine point generation. The encoder extracts global features from the partial input, and the decoder uses these global features to generate a dense, complete point cloud. AdaPtr [54] further advances this line by introducing an adaptive point transformer that dynamically warps a learned template to each input, improving completion quality especially under large missing regions. ODGNet [6] takes a complementary approach, employing an object-driven graph network that incorporates symmetry and geometry priors to refine local details and better preserve fine structures.

3 Proposed Method

RFconstruct is a 3D reconstruction framework designed for mmWave radars, which can recover complete 3D shapes of objects commonly seen by self-driving cars such as vehicles, motorcycles, bicycles and pedestrians. RFconstruct operates in two stages shown in Fig. 3; The first stage focuses on the enhancement of mmWave radar point clouds. The second stage leverages the enhanced partial point cloud to feed through a shape-completion network that generates a complete 3D shape from partial inputs.

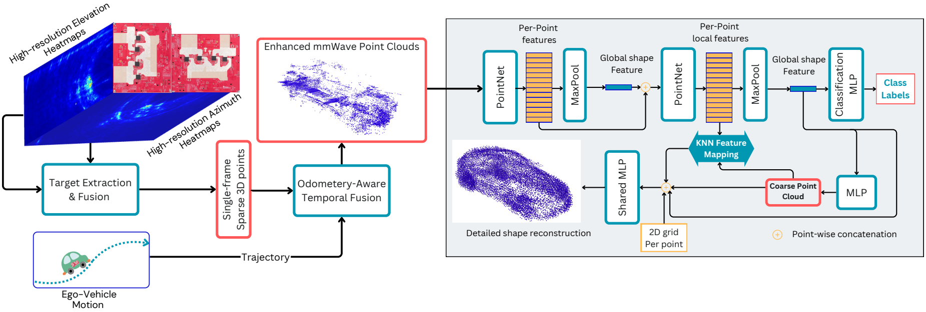

Figure 3: RFconstruct Overview. In Stage 1 (left), Enhanced, dense point clouds are generated by multi-radar fusion and odometry-aware fusion. In stage 2 (right), RFconstruct's completion network generates complete 3D shapes from the enhanced point clouds.

3.1 Enhanced 3D mmWave Point Clouds

Orthogonal Radars Fusion: In order to enhance the resolution and provide elevation information necessary for reconstructing the shapes of objects, we built a radar system consisting of two 1D MIMO radars placed orthogonal to each other, i.e., we flip one radar by 90° so it has good resolution along elevation. The intuition is that the horizontal radar can emulate a 1D horizontal antenna array with N elements, and achieve an equally high azimuth resolution as a prohibitively expensive 2D $N \times N$ elements radar. Similarly, the vertical rada has an equally high elevation resolution. *For a detailed background on mmWave imaging and MIMO radar, please refer to the supplementary material.*

Imaging a 3D scene using a 1D antenna array, however, results in ambiguities on the other dimension. For a target detected at the azimuth angle ϕ_t with the horizontal radar, if we could estimate its elevation angle using vertical radar, we would get the best of both worlds and enable high-resolution 3D imaging. Finding corresponding targets in multiple radar heatmaps is challenging, however, because radar heatmaps lack rich features as in RGB camera images that can be used to associate pixels in multiple camera images [52]. Radar heatmaps appear as the reflected signal power, which is the key feature we can leverage to identify and associate targets across the two radars. Moreover, we know that the range (depth) of corresponding targets would be the same. We first list the targets detected by each radar in the order of the reflected signal power for each range. Then, we start associating targets from the highest power to the lowest until the discrepancy becomes larger than 3dB.

Odometry-Aware Temporal Fusion: In order to combat specularity and increase point cloud density, RFconstruct combines a number of consecutive radar frames and accumulates their point clouds.To accurately accumulate point clouds in different frames, we first need to convert the relative reflectors' positions into absolute positions by compensating for the ego-motion of the radar. To obtain the position and orientation of the radar per frame, RFconstruct uses an odometry sensor that is synchronized with the radar. For each frame of 3D point clouds, we first obtain the corresponding position and orientation from the odometry. We then apply the necessary translation and rotation to adjust the point clouds to the same origin. Then, the transformed point clouds are merged to generate a denser 3D point cloud.

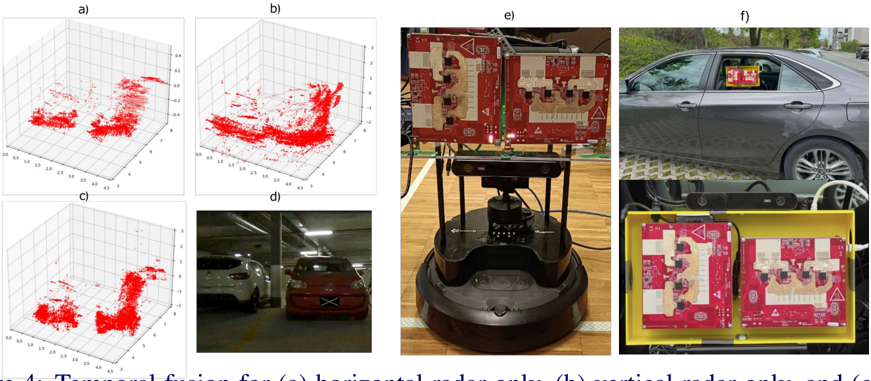

Figure 4: Temporal fusion for (a) horizontal radar only, (b) vertical radar only, and (c) the fused output of both. (d) camera picture of the scene. (e,f) our experimental setup.

3.2 Shape Completion from Radar Point Clouds

RFconstruct takes a data-driven approach to reconstruct complete 3D shapes with missing parts filled in from partial radar point clouds. We focus on four types of object that are most commonly seen and are most important for self-driving cars: cars, bicycles, motorcycles, and pedestrians. Our work builds on the Point-Completion Network (PCN) [55]. However, PCN is designed and trained for significantly different inputs than radar point clouds. It also make restrictive assumption about the object dimensions, centering and orientation which would be impractical for our application.

RFconstruct follows an encoder-decoder architecture similar to that of PCN. However, we introduce two main architectural changes to the neural network as shown in Fig. 3.

(1) Classification-Driven Feature Refinement. A classifier module that classifies the input objects based on the features produced by the encoder. The classifier module is not directly used in our system but it is only used in the computation of the loss function during training. It consists of an MLP layer and helps guide the encoder to better differentiate the four types of targets and produce more class-differentiable features.

(2) Local-Global Context Modeling Decoder. In addition to the global feature vector, we include the point features in the decoder input which we find is necessary for noisy sparse specular input. The decoder is a two-stage decoder. The first stage generates a coarse point cloud that constitutes the general object shape. In the second stage, We measure a nearest-neighbor correspondence between the generated coarse point cloud and the partial input point cloud. This correspondence is used to extract local point feature vectors that are concatenated to the coarse output as well as the global feature vector and fed to a shared MLP to generate a fine-grained point cloud.

Loss Functions: The loss for our network is a weighted combination of a classification loss, coarse loss, and fine loss. The classification loss is a softmax cross-entropy loss that measures the loss between the predicted class and the true class. The coarse loss measures the loss between the output of the coarse stage in the decoder and the ground truth using the Earth Mover's Distance (EMD) metric [12]. Finally, the fine loss is measured between the final output and the ground truth using Chamfer Distance (CD) [12]. *The exact equations of the loss functions can be found in the supplementary material.*

3.3 RFconstruct's Dataset

Because of our unique radar point cloud processing pipeline, there are no available radar datasets that we can use to train and evaluate our model, and collecting data using our hardware data takes a prohibitively long time. Hence, for training, we generate an augmented dataset that closely resembles the behavior and noise of the real-world radar data. This data consists of two parts: the first is simulated mmWave radar data using ray tracing, and the second is a synthetic dataset inspired by the PCN [55] dataset, but is deformed to resemble radar noise patterns and practical motion perspectives. To generate the training data, we use synthetic 3D CAD models from ShapeNet[1] objects dataset as well as CAESAR[53] human dataset. Our final dataset consists of 2,573 unique shapes, including 1,697 cars, 396 bikes and motorbikes, and 480 humans. From each shape, we generate 8 simulated and 8 synthetic data points for a total of 41,168. For the ground truth, we sample the surface of the full 3D shapes uniformly to extract 16384 points, which is equal to the number of output points produced by the network. The real data set includes data collected with the car and the turtlebot shown in Fig. 4. However, only the turtlebot data is used for fine-tuning and quantitative evaluations as we are able to capture the ground truth using a PolyCam with the LiDAR and camera of an iPhone. This data includes 162 cars, 91 bikes, and 52 humans. *A more detailed explanation of how we generate the training and real data sets can be found in the supplementary material.*

4 Implementation and Evaluation Setup

Experimental Hardware Setup: Our experimental setup shown in figure 4 consists of two TI MMWCAS radars [46] and a depth camera ZED 2i [42] with an IMU that is deployed on the same moving platform as the two radars. *The hardware implementation details can be found in the supplementary material.*

Training RFconstruct: RFconstruct as well as all the below baselines are trained with the bounding box priors, such that the partial inputs are positioned in their correct position in the complete shape, and the object orientation is presumed to be fixed. Bounding boxes and the direction of the cars can be obtained from prior work [16, 26]. To avoid the need for bounding boxes, we also train RFconstruct-no-bbox on the true-to-scale and centered data with diverse orientations in the training data. We train both RFconstruct and RFconstruct-no-bbox for 340 epochs. We fine-tune RFconstruct with a portion of the real radar dataset to evaluate improved performance and reduce any simulation biases. We use 80% of the collected data for fine-tuning and 20% for testing. We do not include any objects seen in training for testing. Finally, we also train class-specific RFconstruct models that are trained on only a single class (cars, bikes, or humans).

Baselines: We compare our results against six baselines. (1) *PCN* [55] trained on three classes of interest. (2) *AdaPoinTr* [54] retrained on our augmented dataset. (3) ODGNet [6] trained on our augmented data set. d (4)A class medoid baseline, by selecting the training shape with the lowest average Chamfer Distance to others in the same class. (5) A SAR (synthetic aperture radar) baseline that leverages SAR from vehicle motion to emulate large antenna arrays that improve resolution.

Metrics: We use the following for our evaluation: (1) *Chamfer Distance (CD)* [12] is the minimum distance between a point in one point set and the closest point in the other set. (2)

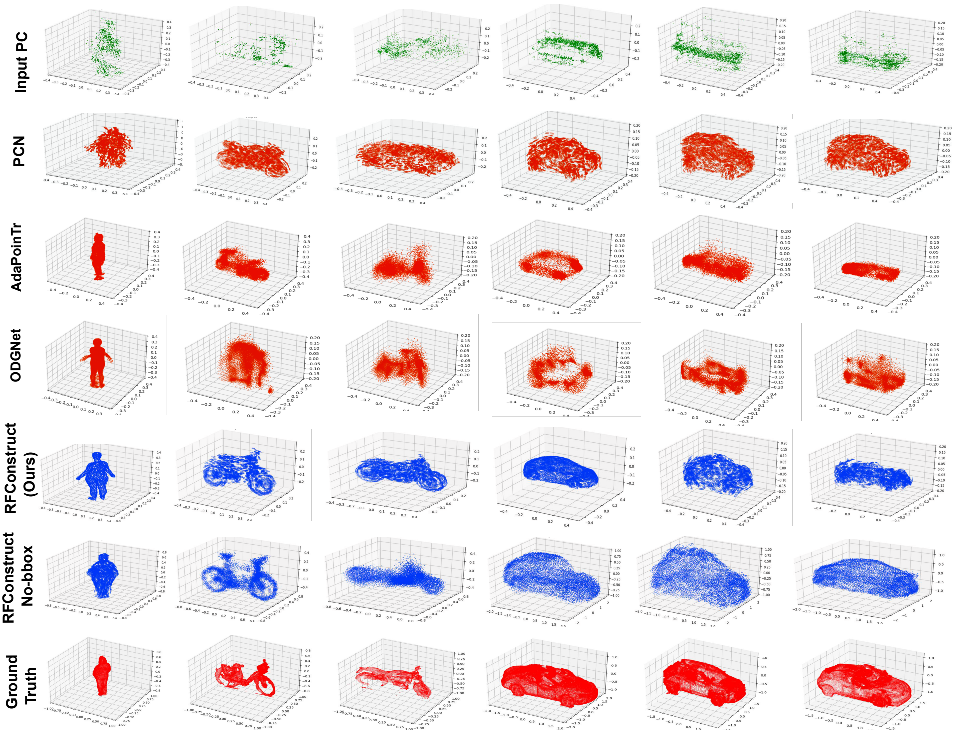

Figure 5: Performance against baselines using 3D radar point clouds without fine-tuning

Earth Mover's Distance (EMD) [12] finds the one-to-one correspondence of points in the two point clouds and calculates the distance between each pair of associated points.

5 Results

Demo: We submitted a video demo in the supplementary material. Note that in this demo, we filter out background buildings and objects and focus on reconstructing foreground objects. Also note that the training data set did not include a human on a bike. Since the reflections from the metal bike are stronger, RFconstruct mainly reconstructs the bike. For moving targets, we use a sliding window to accumulate radar point clouds. However, since RFconstruct was trained on static objects, and since the size of the sliding window should be dynamically adjusted based on the relative speeds of the cars, humans, and bikes, the reconstruction of moving targets appears smeared at times in the video. We evaluate the impact of the vehicle's speed on the reconstruction in the supplementary material. We also elaborate on the limitations of RFconstruct in the supplementary material.

Performance Against Baselines: We show RFconstruct's performance in comparison with the baselines in Fig. 5 as well as in Table 1. RFconstruct accurately reconstructs the 3D shapes of cars, bikes, and humans that closely match the ground truth and outperform the state-of-the-art baselines both quantitatively and qualitatively. The medoid baseline performance is competitive but it assumes perfect class labels and doest not work without bounding box priors. Our fine-tuned model outperforms it, showing better adaptation to individual inputs. Furthermore, note that although RFconstruct is trained primarily on simulated and synthesized data, it could generalize well to real scenes with different backgrounds and visibility conditions. The results also improve significantly with only a small amount of fine-tuning as can be seen in Table 1.

Table 1: Performance against baselines with and without object bounding box knowledge. CD and EMD reported are in cm.

With Bounding-box Priors			**Without Bounding-box Priors**		
	Mean CD	**Mean EMD**		**Mean CD**	**Mean EMD**
ODGNet	3.89	16.07	ODGNet	52.2	108
AdaPointTr	3.42	14.94	AdaPointTr	43	91.46
PCN	4.58	13.03	PCN	42.1	90.8
Medoid Baseline	2.52	9.16	Medoid Baseline	N/A	N/A
RFconstruct	3.52	11.7	RFconstruct	28.8	82.7
Fine-tuned RFconstruct	***2.01***	***6.25***	Fine-tuned RFconstruct	**15.3**	**49.9**

Class-Specific Models. We evaluate RFconstruct trained on all classes against three, class-specific RFconstruct models that are trained on only a single class each. As shown in Table 3, the improvement of a class-specific model over our general model is marginal and inconsistent. This is likely due to the inherent classification module of RFconstruct.

Ablation of the Classifier Module: The classifier module aids the encoder in generating features that are class-separable. This, in turn, improves the decoder performance due to having more representative input features. We analyze the impact of the classifier module by removing it from the pipeline. Table 2 shows that the addition of the classifier module improves the overall CD and EMD.

Ablation of Data Augmentations: RFconstruct is trained on a merged dataset of simulated radar data and synthetic data. Training only on one of these data types results in worse performance, as shown in Table 2. While the simulated data models the noise of radar systems such as low resolution, *sinc* noises, and grating lobes, the geometric perturbations in synthetic data can model signal specularity as well as noise generated by real systems.

Fine-Tuning: Figure 6 shows qualitative results for RFconstruct and Fine-tuned RFconstruct. It also shows the two stages of point generation where the second and fourth columns show the output coarse pointcloud, and the third and fifth columns show the fine pointcloud.

Additional Results: *Further qualitative results for (1) fine-tuning, (2) randomly selected data points, (3) failure cases, (4) the impact of the number of accumulated frames, (5) the impact of relative speeds of the cars, and (5) comparison with SAR can be found in the supplementary material.*

Table 2: Comparison of results without the classifier module and data augmentations.

Classifier module		
Configuration	CD	EMD
With Classifier Module	**3.52**	**11.7**
Without Classifier Module	4.2	12.3
Data augmentations.		
Training Data	CD	EMD
Combined Augmentations	**3.52**	**11.7**
Simulated Radar Data Only	4.62	14.1
Geometric Augmentation Only	4.09	12.53

Table 3: Class-specific models performance

Classes in Evaluation	Classes in Training	CD ↓ (cm)	EMD ↓ (cm)
Human	Human only	2.8	8.34
	Car, Bike, Human	3.1	11.5
Car	Car only	2.65	8.66
	Car, Bike, Human	3.1	10.3
Bike	Bike only	4.3	13.3
	Car, Bike, Human	4.2	13.6

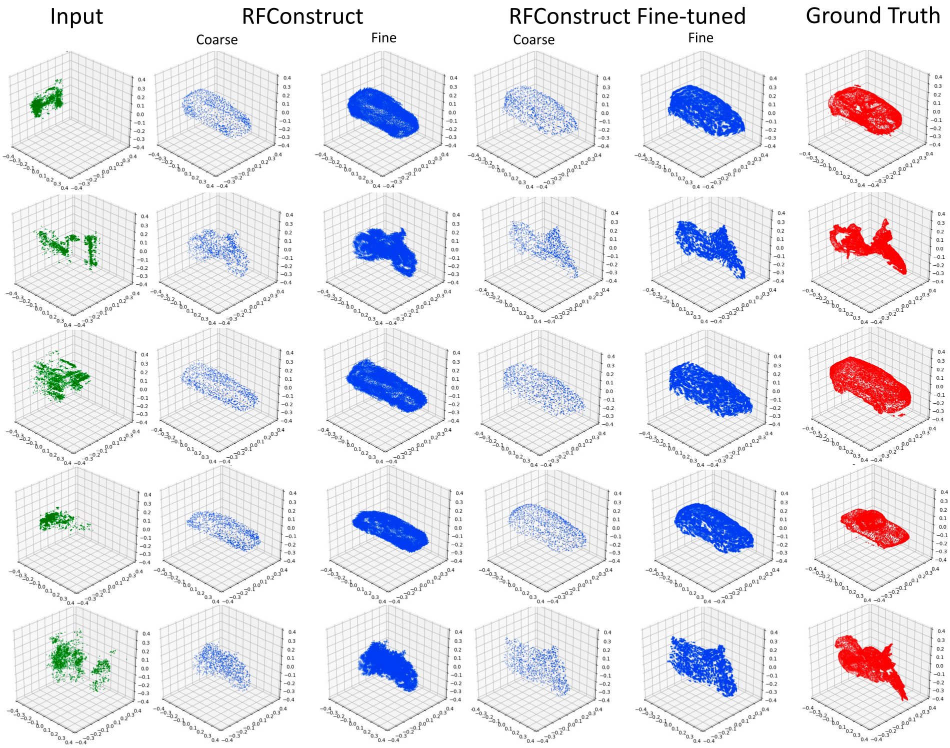

Figure 6: Qualitative results for RFconstruct and RFconstruct fine-tuned on real radar data for randomly selected data points.

6 Conclusion and Limitations

This paper takes major steps towards enhancing autonomous radars by developing a radar system that provides good resolution in both azimuth and elevation and completes 3D shapes of important objects in the context of autonomous driving, such as cars, bikes, and pedestrians, from partial observations. Such capabilities will be essential in enabling self-driving cars to operate in fog and bad weather. RFconstruct was able to tackle challenges in hardware, modality, and practicality to enhance performance. However, significantly more research is needed to enhance the robustness, stability, and generalizability of such systems.

RFconstruct is trained mainly for static objects due to the nature of pointcloud generation. However, this can be expanded for dynamic objects by incorporating dynamic sliding windows for Temporal fusion, as well as incorporating radar doppler information for estimating and compensating for the velocity of dynamic objects in the scene. A preliminary exploration of static sliding windows for dynamic objects can be seen in the demo, which we leave for suture work. A further detailed discussion on RFconstruct's limitations can be found in the supplementary material, which we leave to future work.

References

[1] Shapnet. https://shapenet.org/.

[2] Yasin Almalioglu, Mehmet Turan, Chris Xiaoxuan Lu, Niki Trigoni, and Andrew Markham. Milli-rio: Ego-motion estimation with low-cost millimetre-wave radar. *IEEE Sensors Journal*, 21(3):3314–3323, 2021. doi: 10.1109/JSEN.2020.3023243.

[3] Arbe. Radar Revolution. Delivered. https://arberobotics.com/, 2024. [Online; accessed jul-2-2024].

[4] Kshitiz Bansal, Keshav Rungta, Siyuan Zhu, and Dinesh Bharadia. Pointillism: Accurate 3d bounding box estimation with multi-radars. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, SenSys '20, page 340–353, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450375900. doi: 10.1145/3384419.3430783. URL https://doi.org/10.1145/3384419.3430783.

[5] Pingping Cai and Sanjib Sur. Millipcd: Beyond traditional vision indoor point cloud generation via handheld millimeter-wave devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(4), jan 2023. doi: 10.1145/3569497. URL https://doi.org/10.1145/3569497.

[6] Pingping Cai, Deja Scott, Xiaoguang Li, and Song Wang. Orthogonal dictionary guided shape completion network for point cloud. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024. ISBN 978-1-57735-887-9. doi: 10.1609/aaai.v38i2.27845. URL https://doi.org/10.1609/aaai.v38i2.27845.

[7] Anjun Chen, Xiangyu Wang, Shaohao Zhu, Yanxu Li, Jiming Chen, and Qi Ye. mmbody benchmark: 3d body reconstruction dataset and analysis for millimeter wave radar. In *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, page 3501–3510, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392037. doi: 10.1145/3503161.3548262. URL https://doi.org/10.1145/3503161.3548262.

[8] Weiyan Chen, Fusang Zhang, Tao Gu, Kexing Zhou, Zixuan Huo, and Daqing Zhang. Constructing floor plan through smoke using ultra wideband radar. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(4), dec 2022. doi: 10.1145/3494977. URL https://doi.org/10.1145/3494977.

[9] Andreas Danzer, Thomas Griebel, Martin Bach, and Klaus Dietmayer. 2d car detection in radar data with pointnets. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 61–66, 2019.

[10] Fangqiang Ding, Andras Palffy, Dariu M. Gavrila, and Chris Xiaoxuan Lu. Hidden gems: 4d radar scene flow learning using cross-modal supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9340–9349, June 2023.

[11] Xu Dong, Pengluo Wang, Pengyue Zhang, and Langechuan Liu. Probabilistic oriented object detection in automotive radar. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 102–103, 2020.

[12] Haoqiang Fan, Hao Su, and Leonidas Guibas. A point set generation network for 3d object reconstruction from a single image. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2463–2471, 2017. doi: 10.1109/CVPR.2017.264.

[13] Yiwen Feng, Jiayang Zhao, Chuyu Wang, Lei Xie, and Sanglu Lu. 3d bounding box estimation based on cots mmwave radar via moving scanning. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 8(4), November 2024. doi: 10.1145/3699758. URL https://doi.org/10.1145/3699758.

[14] Xiangyu Gao, Guanbin Xing, Sumit Roy, and Hui Liu. Ramp-cnn: A novel neural network for enhanced automotive radar object recognition. *IEEE Sensors Journal*, 21 (4):5119–5132, Feb 2021.

[15] Junfeng Guan, Sohrab Madani, Suraj Jog, Saurabh Gupta, and Haitham Hassanieh. Through fog high-resolution imaging using millimeter wave radar. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11461–11470, 2020. doi: 10.1109/CVPR42600.2020.01148.

[16] Yiduo Hao, Sohrab Madani, Junfeng Guan, Mohammed Alloulah, Saurabh Gupta, and Haitham Hassanieh. Bootstrapping autonomous driving radars with self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15012–15023, June 2024.

[17] Hao Kong, Xiangyu Xu, Jiadi Yu, Qilin Chen, Chenguang Ma, Yingying Chen, Yi-Chao Chen, and Linghe Kong. m3track: mmwave-based multi-user 3d posture tracking. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, MobiSys '22, page 491–503, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391856. doi: 10.1145/3498361.3538926. URL https://doi.org/10.1145/3498361.3538926.

[18] 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*, ACM MobiCom '24, page 388–403, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400704895. doi: 10.1145/3636534.3649369. URL https://doi.org/10.1145/3636534.3649369.

[19] Shih-Po Lee, Niraj Prakash Kini, Wen-Hsiao Peng, Ching-Wen Ma, and Jenq-Neng Hwang. Hupr: A benchmark for human pose estimation using millimeter wave radar. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 5704–5713, 2023. doi: 10.1109/WACV56688.2023.00567.

[20] Yadong Li, Dongheng Zhang, Jinbo Chen, Jinwei Wan, Dong Zhang, Yang Hu, Qibin Sun, and Yan Chen. Towards domain-independent and real-time gesture recognition using mmwave signal. *IEEE Transactions on Mobile Computing*, 22(12):7355–7369, 2023. doi: 10.1109/TMC.2022.3207570.

[21] Jaime Lien, Nicholas Gillian, M. Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. Soli: ubiquitous gesture sensing with millimeter wave radar. *ACM Trans. Graph.*, 35(4), jul 2016. ISSN 0730-0301. doi: 10.1145/2897824.2925953. URL https://doi.org/10.1145/2897824.2925953.

[22] Jianan Liu, Qiuchi Zhao, Weiyi Xiong, Tao Huang, Qing-Long Han, and Bing Zhu. Smurf: Spatial multi-representation fusion for 3d object detection with 4d imaging radar. *IEEE Transactions on Intelligent Vehicles*, 9(1):799–812, 2024. doi: 10.1109/TIV.2023.3322729.

[23] 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*, MobiSys '20, page 14–27, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450379540. doi: 10.1145/3386901.3388945. URL https://doi.org/10.1145/3386901.3388945.

[24] Chris Xiaoxuan Lu, Muhamad Risqi U. Saputra, Peijun Zhao, Yasin Almalioglu, Pedro P. B. de Gusmao, Changhao Chen, Ke Sun, Niki Trigoni, and Andrew Markham. milliego: single-chip mmwave radar aided egomotion estimation via deep sensor fusion. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, SenSys '20, page 109–122, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450375900. doi: 10.1145/3384419.3430776. URL https://doi.org/10.1145/3384419.3430776.

[25] 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.

[26] Sohrab Madani, Jayden Guan, Waleed Ahmed, Saurabh Gupta, and Haitham Hassanieh. Radatron: Accurate detection using multi-resolution cascaded mimo radar. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIX*, page 160–178, Berlin, Heidelberg, 2022. Springer-Verlag. ISBN 978-3-031-19841-0. doi: 10.1007/978-3-031-19842-7_10. URL https://doi.org/10.1007/978-3-031-19842-7_10.

[27] Bence Major, Daniel Fontijne, Amin Ansari, Ravi Teja Sukhavasi, Radhika Gowaikar, Michael Hamilton, Sean Lee, Slawomir Grzechnik, and Sundar Subramanian. Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 924–932, 2019.

[28] Michael Meyer, Georg Kuschk, and Sven Tomforde. Graph convolutional networks for 3d object detection on radar data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3060–3069, 2021.

[29] Fatemeh Norouzian, Emidio Marchetti, Edward Hoare, Marina Gashinova, Costas Constantinou, Peter Gardner, and Mikhail Cherniakov. Experimental study on

low-thz automotive radar signal attenuation during snowfall. *IET Radar, Sonar & Navigation*, 13(9):1421–1427, 2019. doi: https://doi.org/10.1049/iet-rsn.2018.5644. URL https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-rsn.2018.5644.

[30] Fatemeh Norouzian, Emidio Marchetti, Marina Gashinova, Edward Hoare, Costas Constantinou, Peter Gardner, and Mikhail Cherniakov. Rain attenuation at millimeter wave and low-thz frequencies. *IEEE Transactions on Antennas and Propagation*, 68(1):421–431, 2020. doi: 10.1109/TAP.2019.2938735.

[31] Open Source Robotics Foundation Inc. TutleBot3. https://www.turtlebot.com/turtlebot3/, 2024. [Online; accessed jun-26-2024].

[32] Arthur Ouaknine, Alasdair Newson, Patrick Perez, Florence Tupin, and Julien Rebut. Multi-view radar semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15671–15680, October 2021.

[33] Dong-Hee Paek, Seung-Hyun Kong, and Kevin Tirta Wijaya. K-radar: 4d radar object detection for autonomous driving in various weather conditions. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022. URL https://openreview.net/forum?id=W_bsDmzwaZ7.

[34] Akarsh Prabhakara, Tao Jin, Arnav Das, Gantavya Bhatt, Lilly Kumari, Elahe Soltanaghai, Jeff Bilmes, Swarun Kumar, and Anthony Rowe. High resolution point clouds from mmwave radar. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4135–4142, 2023. doi: 10.1109/ICRA48891.2023.10161429.

[35] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[36] Xie Qian, Deng Qianyi, Cheng Ta-Ying, Zhao Peijun, Patel Amir, Trigoni Niki, and Markham Andrew. mmpoint: Dense human point cloud generation from mmwave. In *The British Machine Vision Conference (BMVC)*, 2023.

[37] Julien Rebut, Arthur Ouaknine, Waqas Malik, and Patrick Pérez. Raw high-definition radar for multi-task learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17021–17030, June 2022.

[38] Ole Schumann, Christian Wöhler, Markus Hahn, and Jürgen Dickmann. Comparison of random forest and long short-term memory network performances in classification tasks using radar. In *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pages 1–6, 2017. doi: 10.1109/SDF.2017.8126350.

[39] Ole Schumann, Markus Hahn, Jürgen Dickmann, and Christian Wöhler. Semantic segmentation on radar point clouds. In *2018 21st International Conference on Information Fusion (FUSION)*, pages 2179–2186, 2018.

[40] Smart Radar System Inc. Radar Professionals: Smart Radar System. https://www.smartradarsystem.com/en/index.html, 2024. [Online; accessed jul-2-2024].

[41] s.m.s, smart microwave sensors GmbH. SMARTMICRO AUTOMOTIVE RADAR. https://www.smartmicro.com/, 2024. [Online; accessed jul-2-2024].

[42] Stereolabs Inc. ZED 2. https://www.stereolabs.com/products/zed-2, 2022. [Online; accessed jun-20-2024].

[43] 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, 2021. doi: 10.1109/IPCCC51483.2021.9679394.

[44] 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, 2022. doi: 10.1109/MASS56207.2022.00075.

[45] Yue Sun, Honggang Zhang, Zhuoming Huang, and Benyuan Liu. R2p: A deep learning model from mmwave radar to point cloud. In *Artificial Neural Networks and Machine Learning – ICANN 2022: 31st International Conference on Artificial Neural Networks, Bristol, UK, September 6–9, 2022, Proceedings, Part I*, page 329–341, Berlin, Heidelberg, 2022. Springer-Verlag. ISBN 978-3-031-15918-3. doi: 10.1007/978-3-031-15919-0_28. URL https://doi.org/10.1007/978-3-031-15919-0_28.

[46] Texas Instruments Inc. MMWCAS-RF-EVM mmwave cascade imaging radar rf evaluation module. https://www.ti.com/tool/MMWCAS-RF-EVM, 2024. [Online; accessed jun-24-2024].

[47] Vayyar. Vayyar's 4D imaging radar platform: Redefining 79 GHz automotive radar. https://vayyar.com/, 2024. [Online; accessed jul-2-2024].

[48] Xiaogang Wang, Marcelo H Ang, and Gim Hee Lee. Voxel-based network for shape completion by leveraging edge generation. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13169–13178, 2021. doi: 10.1109/ICCV48922.2021.01294.

[49] Waymo. A fog blog. https://blog.waymo.com/2021/11/a-fog-blog.html, 2021.

[50] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1912–1920, 2015. doi: 10.1109/CVPR.2015.7298801.

[51] 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*, MobiSys '21, page 269–282, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450384438. doi: 10.1145/3458864.3467679. URL https://doi.org/10.1145/3458864.3467679.

[52] Senquan Yang, Fan Ding, Pu Li, and Songxi Hu. Distributed multi-camera multi-target association for real-time tracking. *Scientific Reports*, 12(1):11052, 2022.

[53] Yipin Yang, Yao Yu, Yu Zhou, Sidan Du, James Davis, and Ruigang Yang. Semantic parametric reshaping of human body models. In *2014 2nd International Conference on 3D Vision*, volume 2, pages 41–48, 2014. doi: 10.1109/3DV.2014.47.

[54] X. Yu, Y. Rao, Z. Wang, J. Lu, and J. Zhou. Adapointr: Diverse point cloud completion with adaptive geometry-aware transformers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12):14114–14130, dec 2023. ISSN 1939-3539. doi: 10.1109/TPAMI.2023.3309253.

[55] 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, 2018.

[56] Shizhe Zang, Ming Ding, David Smith, Paul Tyler, Thierry Rakotoarivelo, and Mohamed Ali Kaafar. The impact of adverse weather conditions on autonomous vehicles: How rain, snow, fog, and hail affect the performance of a self-driving car. *IEEE Vehicular Technology Magazine*, 14(2):103–111, 2019. doi: 10.1109/MVT.2019.2892497.

[57] Zendar Inc. Zendar is accelerating ADAS performance with high resolution automotive radar. https://www.zendar.io/, 2024. [Online; accessed jul-2-2024].

[58] A. Zhang, F. Nowruzi, and R. Laganiere. Raddet: Range-azimuth-doppler based radar object detection for dynamic road users. In *2021 18th Conference on Robots and Vision (CRV)*, pages 95–102, Los Alamitos, CA, USA, may 2021. IEEE Computer Society. doi: 10.1109/CRV52889.2021.00021. URL https://doi.ieeecomputersociety.org/10.1109/CRV52889.2021.00021.

[59] Mingmin Zhao, Tianhong Li, Mohammad Abu Alsheikh, Yonglong Tian, Hang Zhao, Antonio Torralba, and Dina Katabi. Through-wall human pose estimation using radio signals. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7356–7365, 2018. doi: 10.1109/CVPR.2018.00768.

[60] Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. Rf-based 3d skeletons. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, SIGCOMM '18, page 267–281, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450355674. doi: 10.1145/3230543.3230579. URL https://doi.org/10.1145/3230543.3230579.

[61] Mingmin Zhao, Yingcheng Liu, Aniruddh Raghu, Hang Zhao, Tianhong Li, Antonio Torralba, and Dina Katabi. Through-wall human mesh recovery using radio signals. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10112–10121, 2019. doi: 10.1109/ICCV.2019.01021.