ACCURATE DETECTION FOR SELF DRIVING CARS USING MULTI-RESOLUTION MIMO RADAR

BY

WALEED AHMED

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Computer Science
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Adviser:

Professor Haitham Al-Hassanieh

ABSTRACT

Millimeter wave (mmWave) radars are becoming a more popular sensing modality in self-driving cars due to their favorable characteristics in adverse weather. Yet, they currently lack sufficient spatial resolution for semantic scene understanding. In this thesis, we present Radatron, a system capable of accurate object detection using mmWave radar as a stand-alone sensor. To enable Radatron, we introduce a first-of-its-kind, high-resolution automotive radar dataset collected with a cascaded MIMO (Multiple Input Multiple Output) radar. Our radar achieves 5 cm range resolution and 1.2° angular resolution, 10× finer than other publicly available datasets. We also develop a novel hybrid radar processing and deep learning approach to achieve high vehicle detection accuracy. We train and extensively evaluate Radatron to show it achieves 92.6% AP_{50} and 56.3% AP_{75} accuracy in 2D bounding box detection, an 8% and 15.9% improvement over prior art respectively.

To my dearest parents, for their love and support.

ACKNOWLEDGMENTS

The work presented in this thesis would not have been possible without the help and support of a large group of people to whom I owe a lot of gratitude.

First and foremost, I am truly thankful to my advisor, Prof. Haitham Hassanieh, who, for these past two years, has tirelessly supported me and believed in me. He has not only been a great research advisor but also an excellent human being who has cared for me through thick and thin. He has been as patient and supportive through my failures and shortcomings as he has been appreciative and happy during my success. He has always motivated me to put my heart and soul into my career and never settle in life when it is a matter of my ambitions and goals. I will forever be indebted to him for all that he has done and taught me.

I am also really grateful to Prof. Saurabh Gupta who has been like a second advisor to me. He has supported me and guided me through a lot of problems. Every interaction that I've had with him has broadened my perspective on research. I truly admire and respect him.

I would also like to thank my collaborators who contributed to this thesis: Junfeng Guan and Sohrab Madani. I am also grateful to the members of the Systems and Networking Research Group (SyNRG) at UIUC for their support and guidance, especially Ishani, who has been a great friend during my time at UIUC.

Lastly, I cannot express sufficient gratitude towards my loving parents Naila and Sajjad Ahmed, my sisters Ufaq and Shaiza, and my brother Hammad, for their endless love, support and advice. I could not have achieved what I did without their help and I owe all my success to them. No matter what I do, I can never repay them. I am also thankful for all the other members of my family who I love and appreciate tremendously.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION 1

CHAPTER 2 RELATED WORK 4
2.1 Radar-Based Datasets 4
2.2 Learning with Radar Data 4
2.3 Radar-Optical Sensor Fusion 5

CHAPTER 3 BACKGROUND ON MMWAVE MIMO RADAR 6

CHAPTER 4 METHOD 8
4.1 Cascaded MIMO Radar System 8
4.2 Mathematical Formulation of the Motion-Induced Distortion Problem 9
4.3 Motion-Induced Distortion Compensation Algorithm 11
4.4 Radar Signal Processing 13
4.5 Radatron's Network Design 14

CHAPTER 5 RADATRON DATASET 17
5.1 Data Collection Platform 17
5.2 Data Collection 17
5.3 Annotations 17

CHAPTER 6 EVALUATION AND EXPERIMENTS 18
6.1 Evaluation Metrics 18
6.2 Baselines 18
6.3 Radatron Variants 18
6.4 Dataset Split 18
6.5 Test Set Split 18
6.6 Training Details. 19
6.7 Performance Against Baselines 19
6.8 Radatron's Performance 20
6.9 Ablation Studies 21
6.10 Qualitative Results 23
6.11 Failure Cases Analysis 23
6.12 Consistency of Evaluation 26
6.13 Additional Qualitative Results 28

CHAPTER 7 CONCLUSION 30

REFERENCES 31

CHAPTER 1: INTRODUCTION

Recently, there has been a significant amount of work, from both academia [1, 2, 3, 4] and industry [5, 6, 7, 8], on leveraging millimeter wave (mmWave) radars for imaging and object detection in autonomous vehicles. Millimeter wave radars are relatively cheap and can operate in adverse weather conditions such as fog, smog, snowstorms, and sandstorms where today's sensory modalities like cameras and LiDAR fail [9, 10]. Despite that, today's commercial use of mmWave automotive radars remains limited to unidirectional ranging in tasks like adaptive cruise control and parking assistance. This is mainly due to the fact that radar's angular resolution is extremely low, 100× lower than LiDAR as shown in Fig. 1.1(b, c), making it difficult to use radar for object detection. As a result, prior work aiming to gain semantic understanding directly from low resolution radar heatmaps is only able to coarsely localize objects [11, 12, 13] or must fuse radar with LiDAR or cameras to enable object detection [8, 14]. In this thesis, we focus on exploring how well radar performs in object detection tasks and devise techniques to improve its performance.

Improving the angular resolution of conventional radar sensors is challenging. This is because in principle, radar's angular resolution is inversely proportional to the size of the radar antenna aperture [13]. For example, in order to achieve 0.1° angular resolution similar to LiDAR [15], we require a 10 meter-long aperture consisting of an array of 3000 antennas. The cost, power, and large form factor make such a design prohibitively expensive. An alternative cheaper solution is to use a cascaded MIMO (Multiple Input Multiple Output) radar in which multiple radars are combined to emulate a much larger radar aperture [16, 17]. The radars take turns transmitting to avoid interference between the transmitters. Signals from multiple transmitters and receivers are then combined coherently to generate a high resolution image as shown in Fig. 1.1(d) (for primer on radar, see sec. 3). This design, however, cannot work well for dynamic scenes like self-driving cars where the different radar transmitters capture snapshots of the scene at slight timing offsets. In vision, such a problem leads to motion blur which can be addressed using a higher frame rate or deblurring techniques [18, 19]. Radar, on the other hand, uses mmWave RF signals that travel as sine/cosine waves with millimeter scale wavelength. As a result, even a slight motion of few millimeters can completely change the sign of signal across transmitters which can destructively combine to smear, defocus and even eliminate the object especially as the number of radar transmitters increases. Fig. 1.1(d.i) shows this effect: reflections in the moving scene get smeared and appear in different locations than where they really are, which leads to inaccurate bounding boxes prediction.

In this thesis, we present Radatron, a mmWave radar-based object detection system that

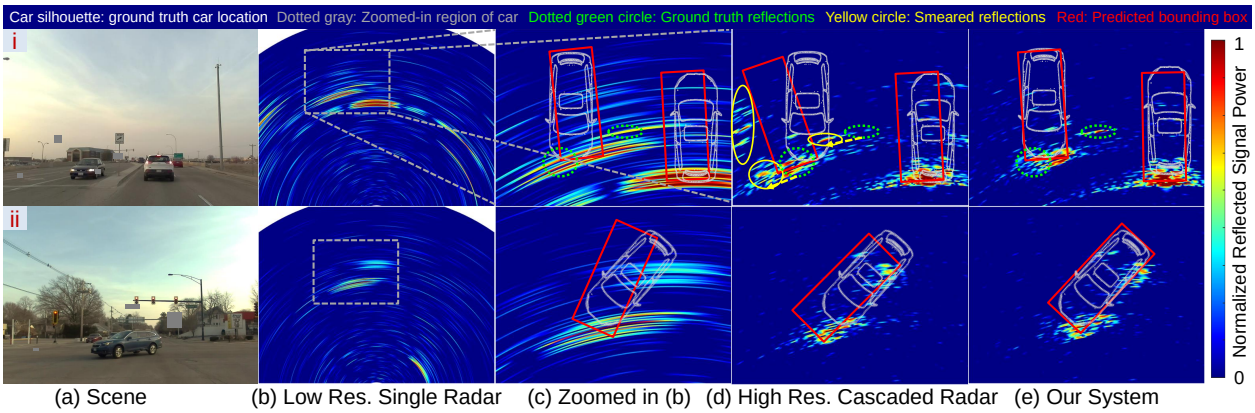

Figure 1.1: The low resolution of millimeter wave radar makes it difficult to perform accurate bounding box detection in (c). High resolution cascaded MIMO radars can improve the resolution but suffer from motion smearing in (d). Radatron delivers accurate detection in (e) by combining motion compensation with a two stream deep learning architecture that takes low and high resolution radar images as input.

can detect precise bounding boxes of vehicles using a cascaded MIMO radar. Radatron overcomes the above challenge by combining a novel radar data pre-processing method with a new deep learning framework. First, we show how to compensate for motion induced errors in pre-processing the raw radar data from a large cascaded MIMO radar. This alleviates most errors, as can be seen by comparing the smeared versions in Fig. 1.1(d) with ones after pre-processing in Fig. 1.1(e). The remaining errors stem from scenarios where the relative speed of the cars is high (e.g. incoming cars, see sec. 4.5). To address these cases, we design a two stream neural network that takes as input both high and low resolution versions of the radar image. Since the low resolution image uses a single radar transmitter, it does not suffer from motion induced errors which allows the network to correct for faulty information like smeared or missed cars that might be mistaken as noise and artifacts.

This thesis also introduces a first-of-its-kind high resolution radar data set collected using a commercial cascaded MIMO radar in urban streets. The data set features radar heatmaps with 10x higher angular resolution than those used in prior work [12, 13, 20], resulting in rich geometric information of objects in the scene, i.e. boundaries and sizes. The data set also includes stereo-camera images which are used for extracting the ground truth and annotating the data. The data set includes 152k frames representing 4.2 hours of driving over 12 days. We also leverage data augmentation to generate significantly more data especially for less common cases (e.g. oriented cars).

We train and extensively evaluate Radatron using our self-collected dataset. Our results show that Radatron improves overall detection accuracy by 8% for AP_{50} and 15.9% for AP_{75} compared to low resolution radars used in prior work [12, 13, 20]. For hard cases like oriented and incoming cars, Radatron improves overall detection accuracy by upto 14.8% for AP_{50}

and 33.1% for AP_{75} compared to low resolution radars, and by upto 13.8% for AP_{50} and 25.2% for AP_{75} compared to a cascaded MIMO Radar without Radatron's pre-processing and two stream network. Besides, we also conducted controlled experiments to qualitatively evaluate Radatron's performance in fog.

Finally, this thesis makes the following contributions. First, we demonstrate the ability of achieving accurate vehicle detection using radar by leveraging the high resolution heatmaps captured by cascaded MIMO radars. Second, we propose a network architecture leveraging multi-resolution radar data along with a motion compensation pre-processing algorithm. Third, we collect a high resolution automotive radar dataset with real-world driving scenarios on urban streets using cascaded MIMO radar which will be released soon.

CHAPTER 2: RELATED WORK

2.1 RADAR-BASED DATASETS

Several radar datasets have recently been introduced using single TI chips [2, 7, 21, 22, 23], the Navtech CTS350-X radar device [3, 4, 11], or other low resolution and 1D radar device [6, 24]. Unlike these datasets, Radatron uses the cascaded MIMO TI radar which provides an angular resolution of 1.18° in azimuth, 18° in elevation and a range resolution of 5 cm enabling accurate object detection. Additional details of our dataset can be found in sec. 5. We summarize and compare our data set to other publicly available datasets in Table 2.1. [3, 4] are the closest in terms of resolution but use a mechanically rotating horn antenna which results in a low frame rate of 4 Hz, motion smearing that cannot be corrected in pre-processing, and inability to compute velocity from Doppler information in the radar signals.

2.2 LEARNING WITH RADAR DATA

Low-cost radar has been used with deep learning in applications such as hand-gesture recognition [25], imaging and tracking of the human body [26, 27, 28, 29], as well as indoor mapping [30]. Our work focuses on using radar for autonomous driving where prior work can be divided into two groups:

2.2.1 Radar Point Clouds

Point clouds are a common interface of commercial automotive radars. Therefore, learning radar data in the format of point clouds is widely studied [20, 33, 34, 35]. [33, 34] demonstrated a semantic segmentation network on radar point clouds while [35] adjusts PointNets [36] for radar data to perform 2D object detection. Pointillism [20] performs 3D bounding box by combining point clouds from multiple spatially separated radars. However, to get point clouds, filtering and thresholding are performed to remove sensor leakage, background clutter, and noise. These hard-coded filtering algorithms lead to the loss of useful information and result in point clouds that are 10 to 100 times sparser than LiDARs [37].

[1]The radar in [2, 7] can provide 3D data with 30° resolution in elevation. However, the data sets provided are 2D.

[2]The radar is mounted on the side of the road rather than on a moving car.

[3]Driving for 280 km which can correspond to 3 to 10 hrs.

[4]Report 260 K objects but only the center is annotated, not the bounding box.

Dataset	Dim.	Resolution			#Total Frames	#Labeled Frames	Frame Rate	Size	Ground Truth	Radar
		Azi.	Ele.	Range						
Nuscenes [31]	1D/2D	N/R	N/A	N/R	1.3 M	40 K	13 fps	5.5 hrs	LiDAR	N/R
CARRADA [7]	2D[1]	15°	N/A	20cm	12.7 K	7.2 K	10 fps	21 mins	Camera	AWR1642
CRUW [2]	2D[1]	15°	N/A	23cm	400 K	N/R[4]	30 fps	3.5 hrs	Camera	AWR1843
OXFORD [3]	2D	1.8°	N/A	17cm	240 K	0	4 fps	280 km[3]	N/A	CTS350-X
RADIATE [4]	2D	1.8°	N/A	17cm	200 K	44 K	4 fps	3 hrs	Camera	CTS350-X
Zendar [6]	2D	30°	N/A	18cm	400 K	11 K	10 fps	11 hrs	LiDAR	N/R
SCORP [22]	3D	15°	30°	12cm	4 K	4 K	10 fps	6.6 mins	Camera	AWR1843
RADDet [23]	3D	15°	30°	20cm	10 K	10 K	N/R	Static[2]	Camera	AWR1843
Radatron	3D	1.2°	18°	5cm	152 K	16 K	10 fps	4.22 hrs	Camera	MMWCAS

Table 2.1: Publicly available radar datasets. We only include publicly available data sets with more than 500 frames that provide 2D and 3D radar heatmaps. Hence, data sets like [1, 5, 8, 11, 32] are not included. N/A: Not Applicable. N/R: Not Reported.

2.2.2 Radar Heatmaps

To avoid loss of information, radar data can be processed as heatmaps with range-angle-Doppler tensors [5, 13, 23, 37, 38]. In order to learn the 3D radar tensors, past methods collapse the 3D radar tensor onto each dimension separately to extract features, and then concatenate the resulting multi-view feature maps for semantic segmentation [38], object classification and center point detection [13], as well as 2D bounding box detection [5]. Other work feeds the 2D BEV range-angle heatmap into the network as an image [11]. Note that while [5, 11] achieved relatively accurate 2D bounding box detection results, their datasets were collected on highways and are not publicly available. Compared to highway driving scenarios, where cars are all moving in the same direction and with similar speeds, our dataset is on urban and suburban streets with more complicated traffic intersections, parked cars on the curbside, and various clutters. In[23], dataset is available but places the radar on the side of the street for traffic monitoring which leads to a poor accuracy of 51.6% AP_{50}. In addition to CNN-based networks, [37] uses graph neural network to achieve a 69% AP_{50} but their data and code are not available.

2.3 RADAR-OPTICAL SENSOR FUSION

Complementary features of multi-sensor data along with the added redundancy has encouraged previous work to combine different sensors. In particular, Radar and LiDAR fusion has been studied in [14, 39, 40] while radar and monocular camera fusion has also been studied in [41, 42, 43, 44, 45, 46]. In this work, we focus on radar as a stand-alone sensor and aim to show the capabilities of high resolution radar in detecting objects with high accuracy, even in urban and dynamic scenarios.

CHAPTER 3: BACKGROUND ON MMWAVE MIMO RADAR

Millimeter wave radars transmit FMCW (Frequency Modulated Continuous Wave) chirps to sense the environment. The chirps emitted from transmitter antenna (TX) reflect off objects in the scene which are then captured by the receiver antenna (RX). By comparing the transmitted and received chirp, we can estimate the round-trip Time-of-Flight (ToF) τ, and hence the ranges of the reflectors $\rho = \tau c/2$ (c denotes the speed of light) in the scene. This is the technique used in today's commercial vehicles that perform radar ranging. Ranging alone, however, is not sufficient to localize objects. One step further is to use a radar with multiple RX antennas that all receive the reflected chirp. The minute ToF differences $\Delta\tau_{ij} = \tau_i - \tau_j$ between these received versions can be exploited to estimate the angle from which the reflections arrive (denoted by ϕ) [47]. The pair (ρ, ϕ) creates a radar heatmap that localizes objects in the 2D polar coordinate.

For this technique to be viable for applications such as semantic scene understanding and object detection, we need to consider the resolution of the radar, which is closely tied to hardware configuration: the range resolution is proportional to the bandwidth of the FMCW chirp, while the angular resolution is proportional to the number of RX antennas. Thanks to the high bandwidth in the mmWave band, mmWave radars achieve cm-level ranging resolution, which is sufficient for most applications. However, reaching an acceptable angular resolution is much more difficult. For instance, to achieve the same angular resolution as a commercial LiDAR, we would need to build a radar with hundreds of RX antennas, which is simply impractical due to the hardware complexity, cost, and power consumption. A much more scalable solution is to use multiple TX as well as multiple RX antennas, a technique referred to as MIMO radar. In MIMO, each of the N TX antennas take turns to transmit one FMCW chirp, which is then received by all M RX antennas, thereby emulating N×M total *virtual* antennas, while using only N+M *physical* antennas [16]. The received chirps from all N·M virtual antennas are then combined to create the (ρ, ϕ) heatmap of the scene.

While MIMO enables higher angular resolution, it comes at the cost of unique challenges. To understand these challenges, we reiterate that in MIMO, TX antennas each transmit one chirp, and all these chirps jointly contribute to the radar heatmap. As TX antennas need to take turns transmitting, there will be a slight time offset δt_{ij} between when the i^{th} and j^{th} chirp are transmitted. For stationary scenes, such time offsets are harmless since they will not affect the ToF difference $\Delta\tau_{ij}$ between different virtual antennas. However, if the scene moves even by as much as 1 mm ($\sim \frac{\lambda}{4}$ at 77 GHz) during the transmitting interval δt_{ij}, the angle estimation and overall radar heatmap can be significantly distorted. This is because the movement of reflections within δt_{ij} contaminates the ToF differences $\Delta\tau_{ij}$

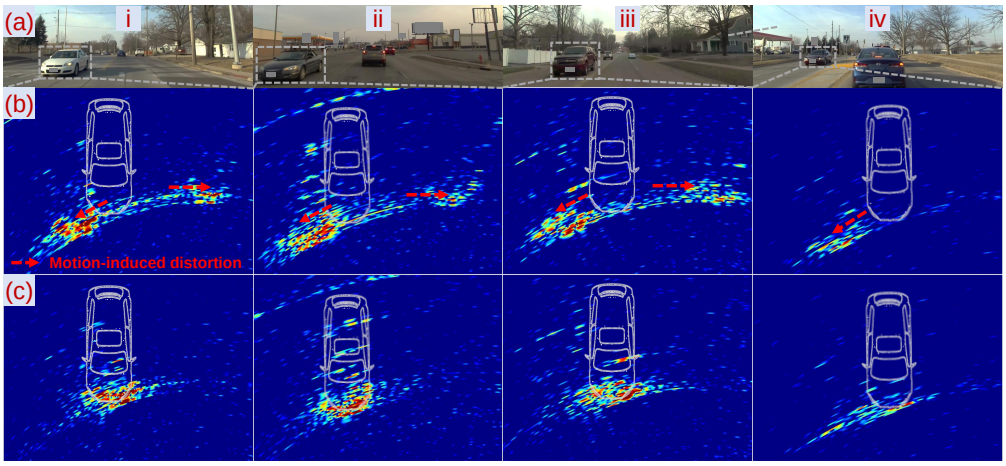

Figure 3.1: Motion-induced distortion and Radatron's compensation algorithm. (a) Original scene. (b) Bird's-eye view radar heatmap under motion-induced distortion. (c) Processed heatmap after applying Radatron's motion compensation algorithm.

between different virtual antennas as follows:

$$\Delta\tau'_{ij} = \tau_i - \tau_j + \delta t_{ij}\frac{2v}{c} = \Delta\tau_{ij} + \delta t_{ij}\frac{2v}{c}, \tag{3.1}$$

where v is the relative speed of the object in the scene, and c is the speed of light. Note that the motion induced ToF change $\delta t_{ij}\frac{2v}{c}$ cannot be isolated from the angle of arrival dependent ToF difference $\Delta\tau_{ij}$. Consequently, object reflections can get smeared in the radar heatmap, moved into another location, or split into multiple less prominent reflections at different angles. We note that the effect of the error term increases with the speed of the object v, making the problem even more severe for high speed objects. We call this effect the *motion-induced distortion* of the MIMO radar. Figure 3.1(b) shows the impact of *motion-induced distortion* in selected range-azimuth radar heatmaps where there is a car moving towards the radar, and we zoom into the region of the incoming car. As one can see, reflections of the car got smeared along ϕ axis, and even split into multiple less prominent reflections appearing at wrong locations away from the car.

CHAPTER 4: METHOD

Our goal is to design a system that can leverage the high resolution cascaded radar as a stand-alone sensor and perform accurate object detection. While the radar heatmaps created using cascaded radar benefit from high angular and range resolution, they come with a set of unique challenges as laid out in sec. 1 and 3. On the one hand, if we cascade multiple TX antennas to emulate a virtual array with more antenna elements, we can maximize the angular resolution and minimize leakages due to sparsity in the antenna array. However, the transmit time offsets between different TX antennas can cause *motion-induced distortion* (sec. 3), and the resulting radar heatmap will be smeared. This issue is particularly severe for automotive radars since both the radar and the scene are moving at high speeds. Radatron overcomes this challenge via a hybrid signal processing and deep learning approach. We will start by explaining our cascaded radar system, the formulation of the motion induced distortion problem and its radar processing solution, and then proceed to describe our network design to tackle this problem.

4.1 CASCADED MIMO RADAR SYSTEM

We collect our own mmWave radar data featuring high angular resolution using TI MMWCAS mmWave cascaded MIMO radar [16]. By cascading four radar system on chips (SoCs), we form a 12 TX and 16 RX MIMO radar system, which can emulate a very large antenna array with up to $16 \times 12 = 192$ elements.

4.1.1 Virtual Antenna Array Emulation

Fig. 4.1 shows the physical positions of the 12 TX antennas, while Fig 4.2 shows the physical positions of the 16 RX antennas. Note that, out of the 12 TX antennas, there are nine TX antennas in the same row (height), whereas the other three antennas located on different rows (heights). These three TX antennas can be used to estimate the elevation angle of the reflections. Although we provide data from these three TX antennas in Radatron's dataset, we do not use them to generate the 2D range-azimuth input heatmap to Radatron's network. We use the other 9 TX antennas along with all 16 RX antennas to emulate an virtual antenna array, the elements of which occupy an 86×1 uniform 1D array as shown in Fig 4.3. We use the radar signal from this uniform 1D array to process the high-resolution input radar heatmaps as we describe in sec. 4 of the thesis. We also use a single TX antenna along with all 16 RX antennas to emulate a sparse 1D array whose topology is the same as the physical RX antenna array. We use this sparse 1D array to process the low-resolution input radar heatmaps as we describe in sec. 4 of the thesis.

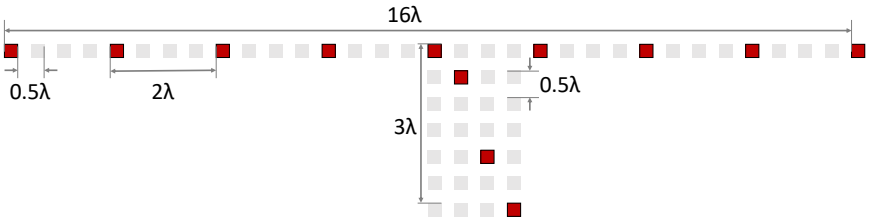

Figure 4.1: Physical TX antenna array of Radatron's cascaded radar.

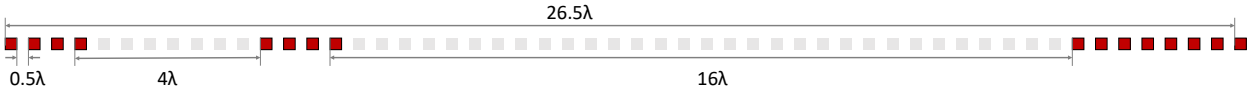

Figure 4.2: Physical RX antenna array of Radatron's cascaded radar.

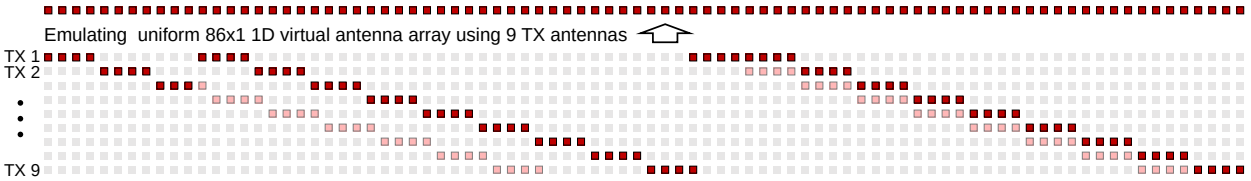

Figure 4.3: Emulating large 1D virtual antenna array using Radatron's cascaded radar. Virtual antenna elements used to emulate large 1D array are marked in red, whereas unused virtual antenna elements are marked in pink.

4.1.2 Radar Configurations

We report our cascaded radar parameters as well as its configuration in our data collection experiments in Tab. 4.1.

4.2 MATHEMATICAL FORMULATION OF THE MOTION-INDUCED DISTORTION PROBLEM

Time-domain shifts of mmWave signals, such as the Time of flight (ToF) differences between antenna elements $\Delta\tau_{ij}$ and motion introduced ToF variance $\delta t_{ij}\frac{2v}{c}$, that we described

Center Frequency	78.5 GHz	Chirp Duration	34.13 us
Bandwidth	3 GHz	# Chirp Loops	64
Range Resolution	5 cm	Chirp Interval	45.62 us
Chirp Slope	88 GHz/ms	Frame Periodicity	40 ms
ADC Sampling Rate	15 MHz	Velocity Resolution	0.054 m/s
# ADC Samples	512	Max Unambiguous Velocity	± 20.85 m/s
Max Range	25.59 m		
Azimuth Aperture	43λ	Elevation Aperture	3.5λ
Azimuth Resolution	∼1.2°	Elevation Resolution	∼18°

Table 4.1: Parameters and experimental configurations of Radatron's mmWave cascaded MIMO radar.

in sec. 3 of the thesis, translate into phase shifts of the electromagnetic waves. To better understand the wave physics under the motion-induced distortion and our compensation algorithms, here we provide more fundamental background on mmWave radar. In the following sections, we explain the challenge of motion-induced distortion and our solution in a more rigorous way.

4.2.1 Underlying Math of mmWave Radar in Phasor Domain

Millimeter wave radars transmit electromagnetic (EM) waves, which are sinusoidal functions that can be represented by *Phasors*. A *Phasors* is a complex number that represents the amplitude (A), frequency (f), and initial phase (θ_0) of a sinusoidal function:

$$\text{Phasor} = Ae^{j(2\pi ft+\theta_0)}, \tag{4.1}$$

where $\theta = 2\pi ft + \theta_0$ is also known as the instantaneous phase of the signal. For FMCW (Frequency Modulated Continuous Wave) radar signals, whose frequency f varies linearly over time, so its phasor representation is:

$$\text{FMCW Radar Waveform} = Ae^{j[2\pi(f_0 t+\frac{\alpha}{2}t^2)+\theta_0]}, \tag{4.2}$$

where f_0 is the starting frequency of the chirp, and α is the chirp slope. The time-delayed reflection signal with round-trip ToF τ can be written as:

$$\text{Reflected Signal} = e^{-j\{2\pi[f_0(t-\tau)+\frac{\alpha}{2}(t-\tau)^2]+\theta_0\}}. \tag{4.3}$$

We compare the received reflection signals against the transmitted the by multiplying with its complex conjugate through a circuit component called frequency mixer. The output signal, which is also known as the beat frequency signal, can be written as:

$$\text{Beat Frequency Signal} \approx e^{j2\pi(\alpha t\tau+f_0\tau)}, \tag{4.4}$$

where a very small phase term $\pi\alpha\tau^2$ has been neglected. As one can see, the instantaneous phase of the beat signal equals the subtraction of the instantaneous phases of TX and RX signals. When we take the standard fast Fourier transform of this time-domain beat signal we get a peak power in the frequency bin corresponding to the beat frequency $\alpha\tau$ and the corresponding phase is $2\pi f_0\tau$.

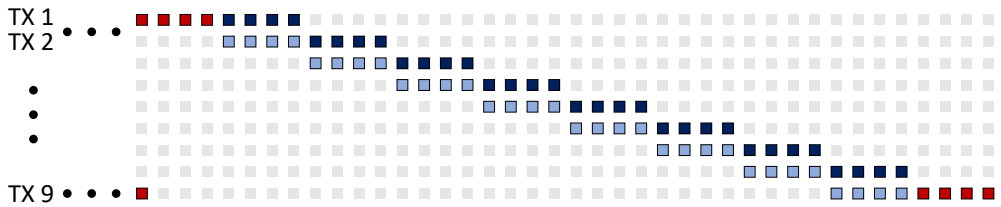

Figure 4.4: Emulated co-located virtual antennas used for motion-induced phase variance estimation. Co-located virtual antenna pairs that are emulated using adjacent TX antennas (time gap equals single chirp interval) are marked in navy and light blue.

4.2.2 Angle of Arrival & Motion-Induced Distortion

After extracting the range information using Fourier transform, we compare beat signals from multiple antennas to estimate the angle from which the reflections arrive (AoA), denoted by ϕ. The pair (ρ, ϕ) creates a radar heatmap in the 2D polar coordinate.

In this step, instead of comparing the minute ToF differences $\Delta\tau_{ij}$ between different antennas, we actually calculate the phase differences $2\pi f_0 \Delta\tau_{ij}$ between antennas. This is because the signal phase is more sensitive to small variances in the round-trip time. Signals coming from different directions lead to different phase differences between adjacent antennas in the antenna array. Specifically, the phase different $\Delta\theta_{ij}$ between element i and j in the linear array will be equal to

$$\Delta\theta_{ij} = 2\pi f_0 \Delta\tau_{ij} = 2\pi f_0(\tau_j - \tau_i) = 2\pi \frac{l \sin(\phi)}{\lambda}(j - i), \tag{4.5}$$

where l is the spacing between adjacent elements, and $\lambda \approx 3.82$ mm is the signal wavelength. For MIMO radars, as TX antennas take turns transmitting, and there is a slight time offset δt between when the i^{th} and j^{th} chirp are transmitted, Eqn. 4.5 becomes:

$$\Delta\theta'_{ij} = 2\pi \frac{l sin(\phi)}{\lambda}(j - i) + 2\pi f_0 \delta t_{ij} \frac{2v}{c} \tag{4.6}$$

As we has discussed in sec. 3 of the thesis, for stationary scenes ($v \approx 0$), the time offsets δt_{ij} will not affect the phase difference $\Delta\theta_{ij}$ between antennas. However, if the scene moves even by as much as 1 mm ($\sim \frac{\lambda}{4}$ at 77 GHz) during the transmitting interval δt_{ij}, the phase different can be significantly off because of $f_0 = 77$ GHz. As a result, the angle estimation and overall radar heatmap can be significantly distorted, especially in sensing highly dynamic environment like self-driving cars.

4.3 MOTION-INDUCED DISTORTION COMPENSATION ALGORITHM

We design a *motion compensation* algorithm as the first step to mitigate the motion induced distortion problem. Our algorithm leverages the *redundancies* in the emulated

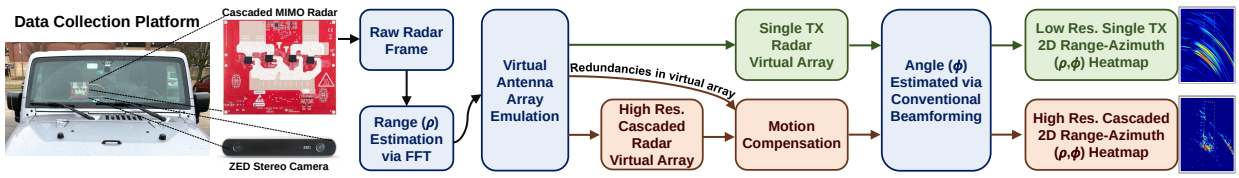

Figure 4.5: Radatron's data pre-processing pipeline.

virtual antenna array, and estimates the motion-induced phase variance.

There are 32 pairs of co-located virtual antennas in the 192 emulated virtual antennas, that are emulated using adjacent physical TX. Therefore, the time interval between each co-located virtual antenna pair i and i' is one chirp interval ΔT. Besides, since virtual antennas i and i' are co-located, there will be no AoA dependent phase differences, and Eq. 4.6 becomes

$$\Delta\theta^{\dagger}_{ii'} = 2\pi f_0 \delta t_{ii'} \frac{2v}{c}. \tag{4.7}$$

Since the only phase difference between these two co-located virtual antennas is the motion-induced phase variance, we can estimate the motion-induced phase variance by measuring $\Delta\theta^{\dagger}_{ii'}$. Therefore, in our radar signal pre-processing pipeline, in addition to the two virtual antenna array formulations, we also group together the 32 pairs of co-located virtual antennas, as shown in Fig. 4.4. We measure the phase differences between each co-located antenna pairs for each range bin, and take an median between the 32 measurements as our final motion-induced phase variance estimation. We then scale the estimated motion-induced phase variance according to the transmitting interval δt for all TX antennas. Finally, we compensate for the motion-induced phase variances for all virtual antennas by multiplying with phasors with opposite phases.

After compensating for the motion-induced phase variances, we then utilize the non-overlapping virtual antennas to extract the angular information of the reflections. Although our algorithm works well in general, as we have shown in this thesis, it does not always work perfectly. It fails in scenes with high-speed incoming car whose relative velocity to the radar is very high.

Besides, although prior work have also noticed the similar motion-induced distortion problem and tried to compensate for it [13, 48], because of their smaller single chip MIMO radar with only two TX antennas, their motion-induced distortion are much less severe. Their compensation technique using multiple chirps from the same TX antenna also cannot work well for our cascaded MIMO radar due to the 6× longer time gap between when the same TX antenna transmits.

4.4 RADAR SIGNAL PROCESSING

On the signal processing end, the complete process of designing the aforementioned *motion compensation* algorithm and integrating it into our radar processing pipeline is shown in Figure 4.5. To summarize, it takes the raw radar signal samples as input, and first applies a standard fast Fourier transform to the time-domain signal, which estimates the reflected power from different ranges. Then, before estimating the angles of reflections to localize the objects, we first compensate for the motion-induced distortion. To do so, we leverage the fact that the emulated virtual antenna array has some *redundancies*; that is, there are some co-located virtual antennas pairs. For the co-located virtual antennas i and i', the estimated ToF difference becomes $\tau_i - \tau_i' + \delta t_{ii'} \frac{2v}{c}$, where $\delta t_{ii'}$ represents the TX interval between co-located virtual antenna pairs. Note that $\tau_i = \tau_i'$ for co-located antennas and they cancel out. Therefore, the measured ToF difference between antenna i and i' is the motion-induced ToF variance: $\delta t_{ii'} \frac{2v}{c}$. As the only unknown in this equation is the speed of the object v, we can estimate v, and therefore the motion-induced variance. We then compensate for the estimated motion-induced variance by adding opposite values to all TX antennas. Figure 3.1(c) shows the intermediate motion compensation results, where the smearing artifacts are mostly corrected, and the reflections overlap well with the ground truth location of the car. After compensating for the motion-induced variance, we then utilize the corrected $\delta\tau$ among non-overlapping virtual antennas to extract the angular information of the reflections. We use the Conventional Beamforming algorithm [49] that outputs a 2D range-azimuth (RA) radar heatmap of the scene in the polar coordinates, where the pixel values represent the reflected signal power. We use this radar signal processing pipeline to create two types of inputs for the network:

4.4.1 High Resolution Cascaded Radar

The radar heatmap is created using a uniform 86×1 virtual antenna array, emulated with multiple TX antennas (sec. 4.1.1) It features the high azimuth resolution achieved using our cascaded MIMO radar.

4.4.2 Low Resolution Single Radar

Instead of using multiple TX antennas, here we only use one TX antenna with all the RX antennas to emulate a non-uniform 16×1 virtual antenna array, so motion compensation is not needed and hence skipped. This processing pipeline approximately reduces the angle resolution by half and introduces leakage artifacts.

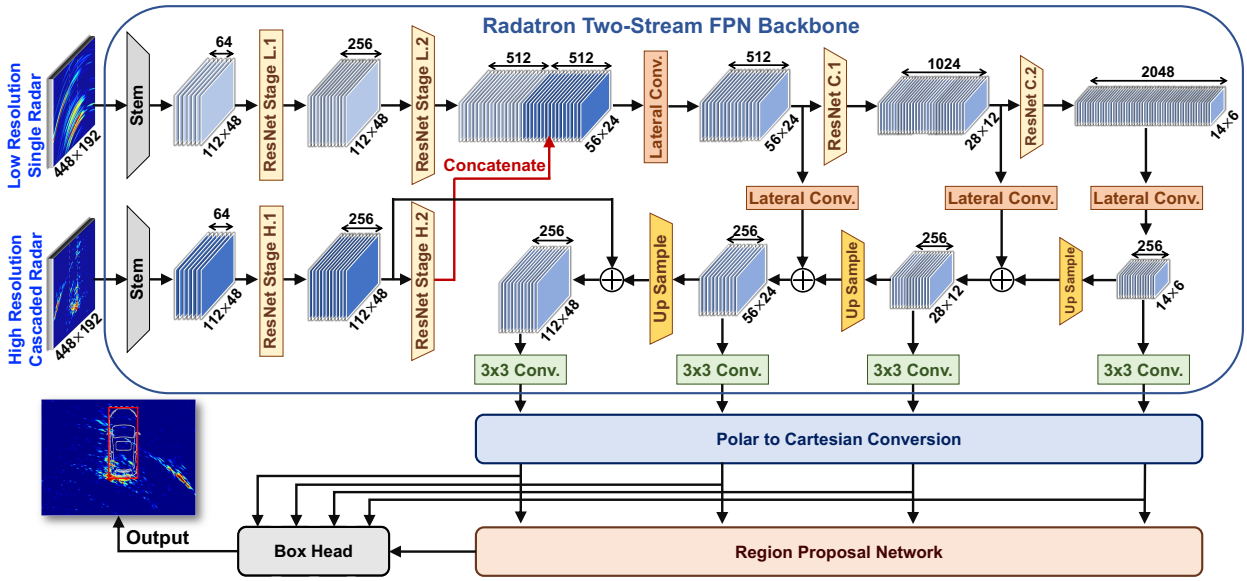

Figure 4.6: Radatron's network architecture. We combine two branches of high resolution and low resolution radar data in an intermediate layer. For each feature map the number of channels and dimensions is indicated above and below it respectively.

4.5 RADATRON'S NETWORK DESIGN

Although our motion compensation algorithm can alleviate the motion-induced distortions to some extent, it is not perfect. Specifically, the algorithm fails in cases of high speed incoming cars, and there will be residual distortions even after applying the motion compensation algorithm. For example, in Fig. 3.1(c.iv), although after compensation the reflection is centered at the location of the car, it's still smeared across a wider range of angles. To deal with these residual distortions, one potential solution would be to cascade M RX antennas with a single TX antenna. As we use only one TX here, the radar heatmap does not suffer from any motion-induced distortion. However, the virtual antennas in the low resolution version are a sparse subset of the complete N·M virtual array. This results in a heatmap with lower resolution and more leakages, as shown in Fig. 4.5. Using this heatmap alone as a solution is therefore not sufficient.

In order to get the best of both worlds, Radatron combines the high resolution with the low resolution solution. Specifically, we leverage the high angular resolution nature of former and the distortion-free nature of latter, by fusing these two versions of radar heatmaps in Radatron's network model. We adapt the Faster R-CNN FPN architecture [50] which has been shown effective previously in [14, 37] for radar data. Fig. 4.6 shows Radatron's network architecture. It takes the two versions of radar heatmaps as input into two parallel branches: The first branch uses the low resolution single radar heatmap, which is free of motion-smearing and hence effective in detecting highly dynamic objects such as incoming

vehicles; the second branch uses the high resolution cascaded radar heatmap and excels in accurately capturing vehicle outlines. Radatron processes these two parallel branches to bring them into a common feature space and then deep-fuses them at an intermediate layer of the backbone network as shown in Fig. 4.6. At the end of the backbone, the feature maps are then converted from the polar to Cartesian coordinates before being fed to the Region Proposal Network and the ROI heads. The output of the network will be 2D vehicle bounding boxes. We will now explain each part of Radatron's network in more detail.

4.5.1 Radatron's Backbone

For the backbone, we adapt an FPN-based architecture. We process the two input heatmaps to have the same dimension, and feed them into two identical branches. Each of the two branches first goes through a stem layer which consists of a 7×7 Conv. layer, ReLU non-linearity [51] and BatchNorm [52]. Each branch then goes through two ResNet stages, which are the same ones used as the building blocks of ResNet50 [53]. We then combine the two branches by concatenating their feature maps of the same dimension across channels, and fuse them by applying a 3×3 Conv. layer. We further encode the feature maps by passing them through ResNet stages, and combine them to create the feature maps similar to [50].

4.5.2 Coordinate Conversion:

Compared to the Cartesian coordinate, the polar coordinate is more natural to radar data as radar has uniform resolution across range and angle. It is also easier for a convolutional network to learn radar artifacts like side lobe leakages in the polar coordinates as they appear parallel to the range and angle coordinates, but extend in a circular fashion in the Cartesian coordinates. On the other hand, bounding boxes work naturally with Cartesian coordinates. We therefore feed in the radar data in the polar coordinates to Radatron's backbone network, and at the end of the backbone explicitly map the features from polar to Cartesian coordinates using bilinear interpolation and before feeding it to the RPN and ROI heads.

4.5.3 RPN and ROI head:

As described earlier, the output feature maps of the backbone are converted from polar to Cartesian coordinates before being fed into the network. We adopt the RPN and ROI architecture in [50] and add oriented boxes. Implementation details can be found in sec. 6.6.

4.5.4 Data Augmentation:

We applied two forms of data augmentations in training:

A. Flipping in Angle. The input heatmap is flipped along the angle axis. In normal driving

scenarios, most incoming cars appear on only one side of the ego vehicle, and flipping azimuth angles eliminates such inherent bias in the dataset.

B. Translation in Angle. We translate the input heatmap along the angle axis. This transformation is similar to one in [13], with the difference that we perform circular shift in angle; i.e., the angles outside the field of view wrap around and fill in the resulting blank space after translation. As most other vehicles appear straight with respect to the ego vehicle, this helps create more oriented cars.

CHAPTER 5: RADATRON DATASET

5.1 DATA COLLECTION PLATFORM

Our data collection platform consists of a TI-MMWCAS cascaded MIMO radar [16] and a ZED stereo camera [54] as shown in Fig. 4.5. Our radar data features high resolution in both range and angle. Our hardware cascades four TI radar chips, with 3 TX and 4 RX antennas each similar to the ones used in prior work [12, 13, 20], into a 12 TX and 16 RX MIMO radar system. This cascaded MIMO radar can emulate a large virtual antenna array with up to 192 antenna elements, which provides us with 1.2° azimuth resolution and and 18° elevation resolution. We transmit FMCW radar signals at 77 GHz with 3 GHz bandwidth, yielding a range resolution of 5 cm.

5.2 DATA COLLECTION

We drove with our data collection platform in diverse scenarios including campus road, our local urban streets, and downtown area of a nearby major city over 12 days. Each day, we conducted four 20-minute data collection sessions, during which we streamed data with a frame rate of 10 FPS. Then we further refined the data and filtered out empty frames with no objects. Our final dataset consists of 152K frames translating into a duration of 4.2 hrs. Note that although Radatron's network only takes 2D range-azimuth heatmap as the input, the raw radar data in our dataset also contains elevation and Doppler information. For operator safety and numerical evaluation need, our dataset was collected in clear weather, but we expect the results to hold in tough weather, as vast prior work have shown that radar works well in fog, rain, and snow [3, 55, 56]. As a initial verification, we conducted controlled fog experiments to qualitatively evaluate Radatron's performance in fog.

5.3 ANNOTATIONS

We manually annotated 2D bird's-eye view (BEV) bounding boxes on our radar data using stereo camera point clouds and RGB camera images as references. We synchronized the radar and stereo camera frames using their own time stamps after aligning the starting time of both sensors. We also calibrated for the coordinate system offsets between the two sensors by applying a rigid motion transformation on anchor points following [57].

CHAPTER 6: EVALUATION AND EXPERIMENTS

6.1 EVALUATION METRICS

. We use Average Precision (AP) as our main metric to evaluate Radatron's detection performance, following recent work [14, 37] in radar object detection, using Intersection over Union (IoU) thresholds values of 0.5, and 0.75. We also use the mean AP (mAP) of IOU values from 0.5 to 0.95 with 0.05 steps. We follow the COCO framework [58] to evaluate Radatron.

6.2 BASELINES

We compare with the following baselines:

A. Radar used in prior work: We implement a virtual array equivalent to the radar used in recent radar datasets [2, 4, 7, 13, 21, 22, 59].

B. Stand-alone single radar TX: We trim Radatron's network to parse one TX antenna only, which is equivalent to having stand-alone top stream in Fig. 4.6.

C. Stand-alone cascaded radar: We process the Cascaded radar data with high resolution but bypass our motion compensation algorithm, and feed it into stand-alone bottom stream in Fig. 4.6.

6.3 RADATRON VARIANTS

We implement three different variants of Radatron:

A. Radatron (No Compensation): We remove the motion compensation algorithm (4.4) from the signal processing pipeline.

B. Radatron (High-res Only): We remove the *top branch* from Fig. 4.6 and only feed in the high-resolution processed radar data through the bottom branch.

C. Radatron(Multi-res): We perform the motion compensation algorithm and use both branches with high- and low-resolution processed radar data in Fig 4.6.

6.4 DATASET SPLIT

Out of 152K overall frames, we manually annotate 16K frames following sec. 5.3. We split the dataset into train and test sets by a 3 to 1 ratio. The set of days from which train and test frames were chosen were disjoint.

6.5 TEST SET SPLIT

To show Radatron's performance under different difficulty scenarios following secs. 4.4 and 4.5, we split vehicles of the test set into 3 categories:

1. *straight*: Any vehicle on the same lane with an orientation within $\pm 5^{\circ}$.

Eval Metric		AP 50				AP 75				mAP			
Model	Split	str.	ori.	inc.	overall	str.	ori.	inc.	overall	str.	ori.	inc.	overall
Radar in Prior work		88.6%	73.9%	69.4%	84.6%	45.0%	24.0%	24.6%	40.4%	47.3%	34.4%	31.2%	44.2%
Stand-alone single-TX		92.4%	77.6%	74.3%	88.9%	50.2%	31.6%	33.6%	46.4%	51.4%	36.6%	37.6%	48.4%
Stand-alone cascaded		87.7%	80.9%	65.9%	84.6%	42.9%	31.9%	26.2%	39.8%	45.5%	38.1%	30.9%	43.2%
Radatron (multi-res)		**95.6%**	**88.7%**	**79.7%**	**92.6%**	**56.3%**	**57.1%**	**38.2%**	**56.3%**	**53.8%**	**53.1%**	**41.4%**	**53.8%**

Table 6.1: Performance against baselines. Best performing model is boldfaced. Str. stands for straight. Ori. stands for oriented. Inc. stands for incoming.

2. *oriented*: Any vehicle whose orientation is out of the $\pm 5^\circ$ range.
3. *incoming*: Any vehicle on the opposite lane, moving towards the ego vehicle.

The *straight* vehicles are relatively easy to detect even using low resolution radars. However, for *oriented* vehicles, high resolution radar is required to accurately detect their angle with respect to the ego vehicle. Finally, *incoming* vehicles tend to get missed by the high resolution heatmap due to the motion induced distortions, as explained in sec 4.2. Instead, our partial cascade radar will pick up the incoming cars when the high resolution heatmap fails. Our test set includes 2854 straight, 327 oriented, and 512 incoming cars.

6.6 TRAINING DETAILS.

We summarize our training details:

6.6.1 Input

The input dimensions to our network are both 448×192 in the polar (ρ, ϕ) coordinates, with range going from 2m to 22.4m and 5cm resolution, and the azimuth angle in $[0^\circ, 180^\circ]$, with 0.94° resolution. The output after conversion to Cartesian (sec. 4.5) is of size 256×320, with the x-axis from -16 to 16m and y-axis from 0 to 25.6m, both with 0.1m resolution. We zero-pad the unmatched areas between the two representations.

6.6.2 Anchor Boxes

We choose two anchor sizes of 28 and 35 pixels (geometric mean of dimensions) according to the average sizes of the cars in our dataset and our output grid resolution. We choose the aspect ratio of the anchors to be 2.5 which is typical for most vehicles, and anchor orientation angles of $-90^\circ, \pm 45^\circ$, and 0°.

6.6.3 Train Parameters.

We train for 25K iterations with SGD Optimizer. The learning rate starts at 0.01, decays by 0.2 after 15K and again after 20K iterations.

6.7 PERFORMANCE AGAINST BASELINES

We first compare Radatron with the prior work radar baseline which uses radar heatmaps used by previous art. As seen in Table 6.1 , Radatron outperforms the prior work radar

Eval Metric		AP 50				AP 75				mAP			
Model	Split	str.	ori.	inc.	overall	str.	ori.	inc.	overall	str.	ori.	inc.	overall
Radatron (no comp.)		93.3%	84.6%	78.9%	91.1%	49.9%	40.4%	37.3%	46.9%	51.3%	43.9%	40.6%	49.1%
Radatron (high-res only)		94.7%	**90.7%**	73.1%	92.4%	**61.4%**	56.3%	34.6%	**57.1%**	**56.6%**	52.3%	37.6%	**53.9%**
Radatron (multi-res)		**95.6%**	88.7%	**79.7%**	**92.6%**	56.3%	**57.1%**	**38.2%**	56.3%	53.8%	**53.1%**	**41.4%**	53.8%

Table 6.2: Performance of Radatron's variants. Best performing model is boldfaced. Str. stands for straight. Ori. stands for oriented. Inc. stands for incoming.

baseline consistently across all evaluation metrics. This proves empirically that the higher angular resolution of our radar data indeed improves the vehicle detection task. We highlight that while their difference in the overall AP_{50} is around 8%, for the harder cases of oriented cars, Radatron outperforms the baseline by as much as 14.8% in the AP_{50} metric. The gap in performance becomes even more prominent for AP_{75}, where Radatron outperforms the prior work radar baseline by as much as 15.9% overall and 33.1% for oriented cars. The same trend is also seen using the mAP metric. We attribute this performance gap to our motion compensation algorithm, multi-resolution network, and high angular resolution of our dataset. For example, as shown in Fig. 1.1, one can visually make out the outline of a vehicle by only looking at the radar heatmaps of Radatron, while the prior work radar baseline only roughly localizes the car. This also explains increased performance gap for the harder cases of oriented cars, and for the higher IoU thresholds.

We next compare Radatron with the other two baselines to show the impact of the our compensation algorithm (sec. 4.4) as well as our fusion network (sec. 4.5) on Radatron's performance. We state few points. First, in AP_{50}, Radatron outperforms the single-TX and cascaded baseline baselines by 3.7% and 8% respectively. For AP_{75}, the margin jumps to 9.9% and 16.5% respectively. This indicates that Radatron is better able to capture the harder cases compared to the two baselines. Second, Radatron outperforms the single-TX baseline in the oriented cars significantly, by 11.1% and 25.5% in AP_{50} and AP_{75} respectively. This is in line with our expectation from sec. 4.5, as the low-resolution and high leakage of single-TX makes it difficult to find the vehicle orientation. Finally, for the incoming cars, Radatron outperforms the cascaded baseline by large margins of 13.8% and 12% for AP_{50} and AP_{75} respectively. This confirms our hypothesis in sec. 4.5 and 4.4, as the lack of motion compensation algorithm severely distorts the cascaded baseline, as shown in Fig. 3.1(b).

6.8 RADATRON'S PERFORMANCE

We now analyze the performance of three different variants of Radatron defined earlier in this section. The results are shown in Table 6.2. The *multi-resolution* model outperforms the *no compensation* model by 1.5% and 9.4% in AP_{50} and AP_{75} respectively, which means that the multi-res architecture alone without the motion compensation algorithm will not perform well enough, especially for the harder cases, like high-speed incoming cars. On the

Eval Metric		AP 50			AP 75		
Ablation	Split	str.	ori.	inc.	str.	ori.	inc.
Cartesian input		91.8%	86.3%	66.5%	49.1%	53.5%	23.8%
Learned conversion		86.5%	55.4%	45.4%	42.7%	9.0%	8.7%
No augmentation		90.6%	77.7%	65.9%	53.2%	29.6%	21.3%
Radatron (multi-res)		**95.6%**	**88.7%**	**79.7%**	**56.3%**	**57.1%**	**38.2%**

Table 6.3: Ablation results. Best performing model is boldfaced.

other hand, the *multi-resolution* model also outperforms *high-resolution only* for incoming cars by 6.6% and 3.6% respectively, which further shows that the motion compensation algorithm alone is not sufficient and can be improved upon using the multi-res network. We note, however, that *multi-resolution*'s performance improvement for the high speed incoming vehicles comes with a slight decrease in performance for oriented cars compared to the *high-resolution only* network. We envision that one could come up with smart combination of high-res and multi-res variants of Radatronto improve the results on all metrics.

6.9 ABLATION STUDIES

6.9.1 Impact of Data Augmentation:

To study the impact of the two forms of data augmentations applied (discussed in sec. 4.5.4) on Radatron's performance, we remove the data augmentations while keeping the rest of Radatron's pipeline the same. As the results in Table 6.3 show, the augmentations consistently improve the performance across all metrics. The 16.9% AP_{75} improvement over incoming cars confirms our assumption on the horizontal flipping augmentation (sec. 4.5.4), while the 27.5% AP_{75} improvement for oriented cars shows affirms that angular shift can help with oriented vehicle predictions.

6.9.2 Impact of Coordinate System:

Here we wish to study the impact of different possible choices for input coordinates. To do so, we consider two alternatives to our design. In the first version, *Cartesian input*, we feed in Cartesian coordinates to the network from the beginning by converting the input radar tensors from polar to Cartesian. In the second version, *learned conversion*, we remove the conversion and let the network implicitly learn to convert from the polar input to Cartesian bounding boxes at the output. As the results in Table 6.3 show, Radatron's original coordinate conversion outperforms *Cartesian input* by 3.8% in AP_{50} and 7.2% in AP_{75} for straight cases. A similar trend is seen for oriented and incoming cars. This confirms our hypothesis in sec. 4.5.2 that it is easier for the network to learn the radar artifacts and suppress them in polar coordinates compared to Cartesian. Radatron also outperforms *learned conversion* by 9.1% in AP_{50} and 13.6% in AP_{75} for straight cars and even larger margins for other cases. Hence explicit conversion of the coordinates rather than letting the network learn the

conversion improves the performance.

6.9.3 Fusion at Different Stages:

In sec. 4.5, we proposed a fusion based approach for Radatron to leverage the high resolution of the cascaded radar input and the distortion-free nature of the single radar input. We pass the two inputs through identical streams and concatenate them after the second ResNet block. The decision of where to fuse the two input streams is a key design choice that affects the performance of Radatron. We show this ablation study in Table 6.4 where we compare Radatron with its two other implementations: one where we fuse the two inputs at the beginning and pass them through a single stream network, second where we fuse the two streams after passing them individually through all the ResNet blocks.

Looking at the results, it's evident that fusing the low resolution and high resolution inputs before feeding them into the network gives worse performance as compared to our proposed implementation. While Radatron is outperformed by its late fusion implementation for straight cars for AP_{75}, it still hold significant advantage over the late fusion implementation for the harder cases like incoming cars with improvements of 2.3%, 1.8% and 1.4% in the AP_{50}, AP_{75} and mAP metrics respectively. It also beats the other two fusion strategies by significant margins for the oriented car case. One possible reason for this is that the number of learnable parameters increase exponentially for the late fusion implementation and the network does not see enough of these rare hard examples to learn so many parameters optimally.

6.9.4 Doppler

As we mentioned in sec. 5.2 of the thesis, our cascaded radar also provides Doppler information, and we have conducted some initial experiments on leveraging this Doppler information. In these experiments, we extract the Doppler information for the singe radar TX and concatenate it as a second channel to the single radar TX input of our network. Here, we show a comparison of Radatron with and without Doppler in Table 6.5, while our Doppler pre-processing algorithm is described in appendix G. It can be observed that concatenating Doppler information as a second channel to the single radar TX input does not provide any notable improvement. The intuition behind this is that although Doppler can provide useful information for separating out closely spaced cars based on their different velocities, such uncommon scenarios are not the major source of error in our results. In particular, Doppler does not help with orientation or motion-induced distortion which are our major challenges.

However, we believe that the Doppler information can be extremely useful if we extend Radatron to not only detecting bounding boxes of vehicles but also estimating the moving directions and speeds of vehicles. We leave leveraging the Doppler information for other

Eval Metric		AP 50				AP 75				mAP			
Model	Split	str.	ori.	inc.	overall	str.	ori.	inc.	overall	str.	ori.	inc.	overall
Radatron (Early Fusion)		93.0%	88.0%	77.1%	91.1%	53.5%	53.6%	36.1%	50.7%	52.8%	50.1%	38.4%	51.5%
Radatron (Late Fusion)		94.5%	88.1%	77.4%	92.2%	**56.6%**	52.0%	36.4%	53.1%	**54.9%**	49.7%	40.0%	52.6%
Radatron (multi-res)		**95.6%**	**88.7%**	**79.7%**	**92.6%**	56.3%	**57.1%**	**38.2%**	**56.3%**	53.8%	**53.1%**	**41.4%**	**53.8%**

Table 6.4: Additional ablation study on fusion at different stages. Best performing model is boldfaced.

Eval Metric		AP 50				AP 75				mAP			
Model	Split	str.	ori.	inc.	overall	str.	ori.	inc.	overall	str.	ori.	inc.	overall
Radatron (With Doppler)		94.1%	86.2%	77.2%	91.1%	51.7%	52.6%	**40.3%**	49.8%	52.4%	50.2%	41.2%	50.6%
Radatron (multi-res)		**95.6%**	**88.7%**	**79.7%**	**92.6%**	**56.3%**	**57.1%**	38.2%	**56.3%**	**53.8%**	**53.1%**	**41.4%**	**53.8%**

Table 6.5: Additional ablation study on Doppler input. Best performing model is boldfaced.

tasks such as speed estimation for future work.

6.10 QUALITATIVE RESULTS

We show example qualitative results from our test set in Fig. 6.4, by overlaying the predictions (in solid red line) and ground truth bounding boxes (dotted green line) on top of Radatron's high-resolution input radar heatmaps in row (b). We also compare Radatron's performance against other baselines, and summarize our observations as follows. As the resolution of the radar heatmap improves, the predictions also become more accurate especially for oriented cars. However, even with the same resolution as Radatron's heatmap, the cascaded baseline suffers when the targets are moving with a high relative speed to the radar, e.g. the incoming cars in Fig. 6.4(c.iii-vi), due to motion-induced distortion as we described in chapter 3. Through distortion compensation and fusion network, Radatron is able to overcome this challenge and accurate predict incoming cars. We also noticed some typical failure cases for Radatron, which we show in Fig. 6.4(b.vi-vii). These cases are likely caused by the fusion network falsely trusting the low-resolution branch and trying to resolve non-existing motion distortion.

6.10.1 Controlled Fog Experiment

Figure 6.2 shows Radatron's performance in realistic fog emulated using a fog machine with high-density water-based fog fluid, following past work [55, 60]. As depicted in the figure, while the cars are not be visible in the RGB image, Radatron can accurately detect cars in the scene.

6.11 FAILURE CASES ANALYSIS

Here we summarize a few typical failure examples of Radatron, and we analyze the possible reason for the prediction errors.

1. *Occlusion.* The first type of failure cases we notice is when the line of sight path to a

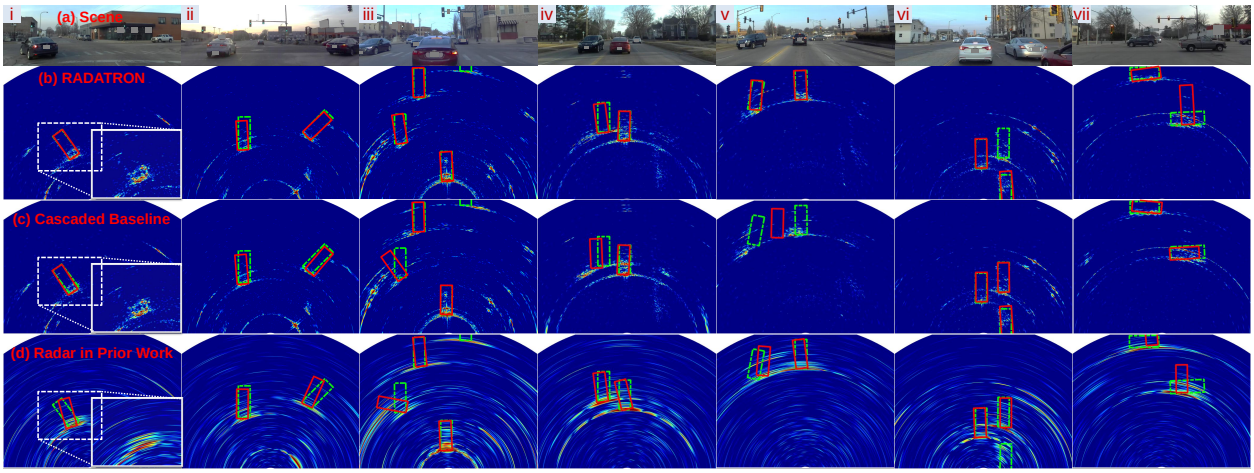

Figure 6.1: Examples from our test set. Ground truth marked in green and predictions in red. (a) Original scene. Row (b) shows Radatron's performance overlaid on distortion compensated radar heatmaps. Row (c) and (d) show the performances of stand-alone cascaded and radar in prior work baselines along with their input heatmaps respectively.

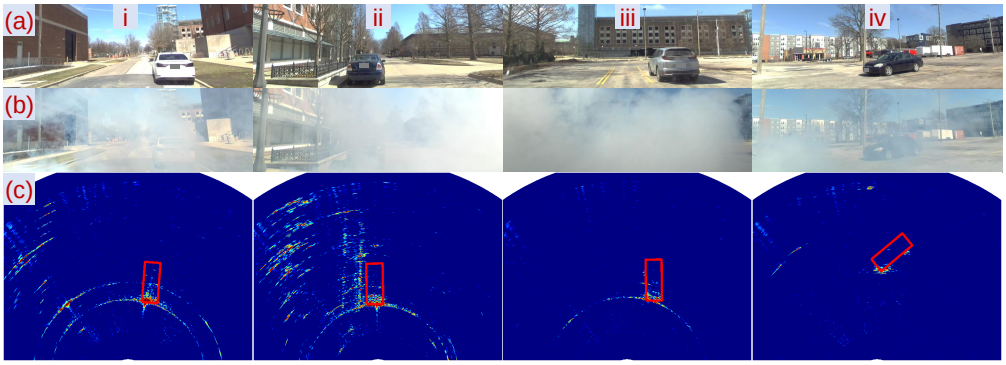

Figure 6.2: Controlled Fog Experiment. (a) Original scene. (b) Scene in fog. (c) Prediction overlaid on radar heatmap captured in fog.

car is partially blocked by another car. In these scenarios, Radatron can either miss the occluded car, e.g. Fig. 6.3(1), or predict misplaced bounding boxes, e.g. Fig. 6.3(2). This is because the metallic bodies of vehicles block mmWave signals, such that the radar signals cannot reach the occluded parts of cars. Therefore, these parts become invisible in the radar heatmap, and in some cases the incomplete reflections provide too little information for Radatron to detect the partially occluded cars.

2. *Specular reflection.* We also noticed that some predicted bounding boxes suffer from low intersection over union (IoU), either because of incorrect car size, e.g. Fig. 6.3(3,4), or inaccurate orientation, e.g. Fig. 6.3(5). Such errors are likely caused by the specular nature of mmWave radar reflections. Millimeter-Wave signals exhibit mirror-like reflections on the smooth metallic surfaces of cars [61], as a result, even if the car is not occluded, reflections from some parts of the car cannot propagate back to the

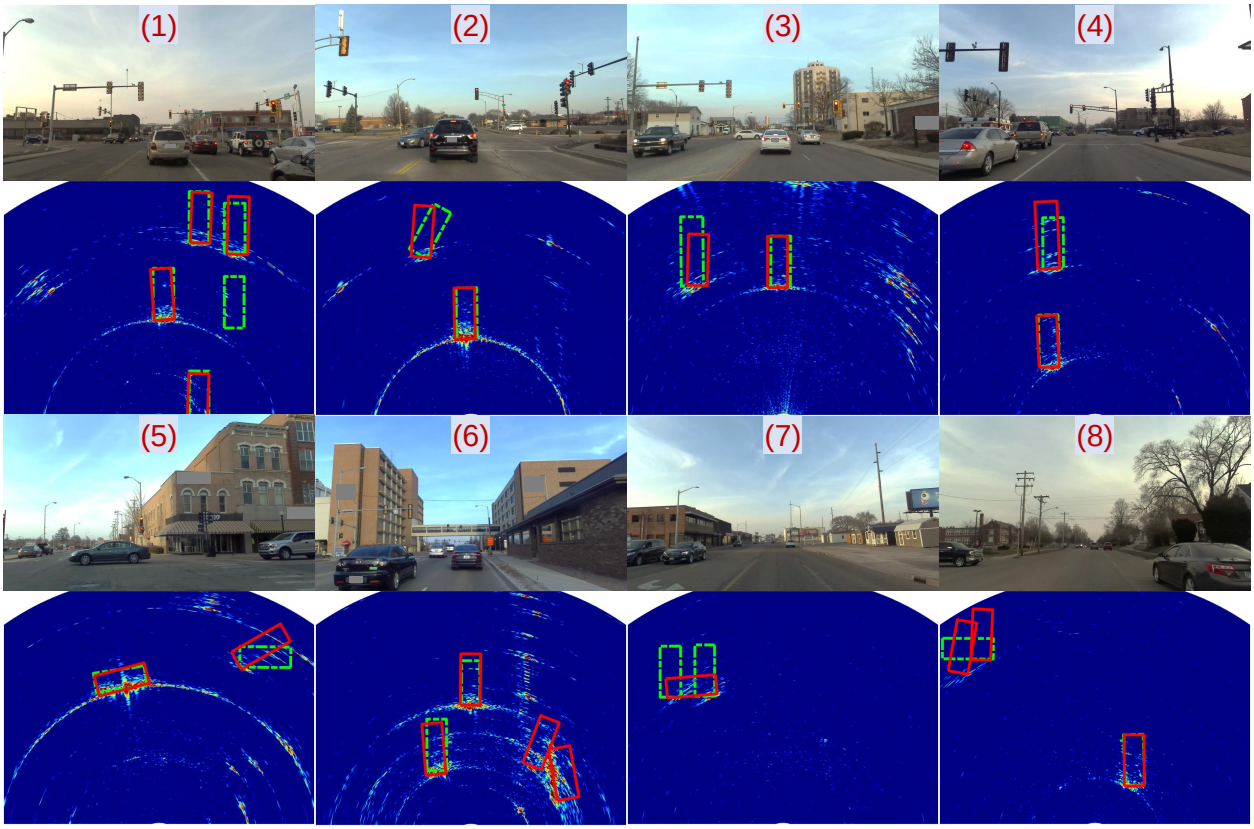

Figure 6.3: Typical prediction errors in our test set. Ground truth is marked in green and predictions are marked in red. Top row of each example shows the original scene and the bottom row shows Radatron's predictions and ground truth bounding boxes overlaid on the input radar heatmaps.

radar receiver, rendering these parts invisible in the heatmap. Radatron tries to learn the specularity effect in radar reflections and infer the complete car bounding boxes. However, due to severe specularity in some scenarios, e.g. the side of the incoming pickup truck in Fig. 6.3(3), predictions can be off in size and orientation.

3. *False alarm due to background reflections.* Although in most cases Radatron correctly identifies foreground objects from the background, it sometimes confuses background reflections for cars. For example, in Fig. 6.3(6), the strong reflections from the building structures very close to the road is incorrectly detected as cars.

4. *Two adjacent cars.* Another tricky scenario for Radatron is when two cars are very close to each other as shown in Fig. 6.3(7). Radatron sometimes mistakes the two clusters of reflections from the two nearby cars as the specular reflection from a horizontal car, so it draws a single bounding box across the two cars. Interestingly, we have also seen the reverse case where Radatron predicts two vertical bounding boxes for a single horizontal car as shown in Fig. 6.3(8). Fortunately, as we discussed in sec. 6.9.4, we can

leverage Doppler information to better distinguish two cars very close to each other versus a single horizontal car.

5. *Lower spatial resolution on the edges of the field of view.* Finally, compared to the center of the scene, Radatron tends to make more mistakes on the edges of the radar field of view, e.g. Fig. 6.3(5,8). This is potentially due to the lower spatial resolution on the edges compared to the center. Note that radar heatmaps do not have uniform spatial resolution across the entire field of view. The radar angular resolution decreases towards the left and right boundaries of the field of view. Besides, for the farther away distances, the same angular resolution translates into a lower spatial resolution. Finally, the transmitter and receiver antennas of the radar also have lower gain away from the center. As a result, prediction errors caused by the above mentioned sources are more commonly seen on the edges of the heatmap due to relatively lower spatial resolution. On the other hand, the reduced detection accuracy in the lower resolution regions also proves the importance of improving the spatial resolution of radar in achieving accurate object detection.

6.12 CONSISTENCY OF EVALUATION

As mentioned in sec. 6.4 of the thesis, the set of days from which the frames for training and testing are chosen are disjoint. In total, our annotated dataset spans four days and the distribution of frames (*straight, oriented, incoming* following sec. 6) across each day is different as shown in Table 6.6. For the results shown in sec. 6.7- 6.8, we chose *Day 1* for testing and all other days for training to make the dataset follow an approximate 3:1 train-test split.

To show that Radatron's improvement over other baselines is consistent across different train-test splits, we repeat all experiments while choosing different days as the test set. Table 6.7 and Table 6.8 show the results when we use *Day 2* and *Day 3* respectively as the test set, while using all other days for training. For both cases, following the trends reported in sec. 6.7, all three implementations of Radatron consistently outperform the three baselines. Next, we discuss the results for each day in detail.

6.12.1 Day 2

Compared to the prior work radar baseline, Radatron achieves a 12.8% improvement overall, a massive 44% improvement for oriented cars and a 14.2% improvement for incoming cars in the AP_{50} metric. Similarly, in the AP_{75} metric, Radatron outperforms the prior work radar baseline by as much as 10.4% overall, 21.3% for oriented cars and 5.9% for incoming cars. The same trend can be seen in the mAP metric. The notable improvements of 44%

and 21.3% in the AP_{50} and AP_{75} metrics respectively, for oriented cars stem from the fact that *Day 2* has significantly more oriented cars as shown in Table 6.6. By using this day for testing, the network misses out on learning from a large number of frames with oriented cars during training. The effect of this is evident in the absolute AP values for all experiments, but is especially amplified for prior work radar baseline since it's already low resolution, and needs a lot more frames to learn the embeddings for oriented cars.

Next, we compare Radatron to the other two baselines. Similar to sec. 6.1, Radatron betters the single-TX and cascaded baselines by 9.7% and 9.1% respectively overall for AP_{50}. The margin becomes 8.6% and 9.3% respectively for AP_{75}. Similar to the prior work baseline, Radatron surpasses the single-TX and cascaded baselines by as much as 35.6% and 25.4% respectively for oriented cars, and by 7% and 10.3% respectively for incoming cars in the AP_{50} metric. For AP_{75}, the margins jump to 16.3% and 18.5% respectively for oriented cars, and to 3.8% and 4.7% respectively for incoming cars. The mAP values follow a similar trend.

6.12.2 Day 3

The trends seen for *Day 1* and *Day 2* more or less follow in case of *Day 3* as well. Radatron betters the prior work baseline by 5% overall, by 11.8% for oriented cars and by 5% for incoming cars in the AP_{50} metric. In the AP_{75} metric, Radatron achieves an improvement of as much as 15.2% overall, 18.1% for oriented cars and 6.1% for incoming cars. The mAP metric follows a similar trend.

Next, Radatron outperforms the single-TX and cascaded baselines by 4.8% and 6.6% respectively overall for AP_{50}. The gap jumps to 7.5% and 5.6% respectively for AP_{75}. For oriented cars, Radatron betters the single-TX and cascaded baselines by 9.7% and 9.8% respectively in the AP_{50} metric, and by 18.1% and 16.2% respectively in the AP_{75} metric. For incoming cars, Radatron outperforms the single-TX and cascaded baselines by as much as 5.9% and 19.7% respectively in the AP_{50} metric, and by 6.1% and 15.6% respectively in the AP_{75} metric. A similar trend can be seen for the mAP values.

Note. We note that for evaluation on *Day 2* and *Day 3*, the absolute AP values for oriented

Day	Total frames	Total cars	Straight cars	Oriented cars	Incoming cars
Day1	2950	4107	3207	327	573
Day2	4171	6186	3890	1014	1282
Day3	8376	13032	10975	509	1548
Day4	720	1029	812	132	85

Table 6.6: Distribution of different categories across all days

Eval Metric		AP 50				AP 75				mAP			
Model	Split	str.	ori.	inc.	overall	str.	ori.	inc.	overall	str.	ori.	inc.	overall
Radar in Prior work		86.4%	26.4%	48.9%	71.5%	49.0%	4.4%	12.5%	35.5%	47.8%	9.5%	19.5%	37.1%
Stand-alone Single-TX		90.5%	34.8%	56.1%	74.6%	52.2%	9.4%	14.6%	37.3%	51.6%	13.8%	22.4%	39.2%
Stand-alone cascaded		88.7%	45%	52.8%	75.2%	52.5%	7.2%	13.7%	36.6%	51.2%	16.6%	21.2%	39.4%
Radatron (No comp.)		90.7%	49.9%	51.7%	77.6%	50.0%	10.4%	14.5%	36.8%	50.2%	19.2%	20.8%	39.7%
Radatron (high-res only)		**94.8**%	68.0%	59.3%	84.2%	**63.7%**	**27.3%**	16.6%	**48.2%**	**57.9%**	**32.9%**	24.3%	**47.2%**
Radatron(multi-res)		93.8%	**70.4%**	**63.1%**	**84.3%**	59.8%	25.7%	**18.4%**	45.9%	56.3%	32.8%	**26.8%**	46.6%

Table 6.7: Quantitative results on Day #2. Best performing model is boldfaced.

Eval Metric		AP 50				AP 75				mAP			
Model	Split	str.	ori.	inc.	overall	str.	ori.	inc.	overall	str.	ori.	inc.	overall
Radar in Prior work		89.9%	60.7%	67.7%	86.8%	37.0%	19.3%	20.8%	34.9%	43.9%	26.6%	29.1%	41.9%
Stand-alone Single-TX		90.2%	62.8%	66.8%	87.0%	45.5%	19.3%	20.8%	42.6%	47.8%	26.6%	29.1%	45.5%
Stand-alone cascaded		90.3%	62.7%	53%	85.2%	49.7%	21.2%	11.3%	44.5%	48.9%	28.8%	19.2%	45.1%
Radatron (No comp.)		94.0%	68.7%	72.4%	90.8%	47.9%	26.0%	25.7%	44.3%	50.2%	33.5%	33.4%	47.8%
Radatron (high-res only)		**96.3**%	72.0%	61.9%	**92.0**%	53.5%	**38.3%**	20.1%	49.5%	**52.7%**	**39.7%**	27.0%	49.7%
Radatron(multi-res)		94.2%	**72.5%**	**72.7%**	91.8%	**54.3%**	37.4%	**26.9%**	**50.1%**	52.5%	38.4%	**33.4%**	**49.9%**

Table 6.8: Quantitative results on Day #3. Best performing model is boldfaced.

and incoming cars are lower than those reported for *Day 1* in sec. 6.1. This is because *Day 2* and *Day 3* both have significantly more number of frames with these hard cases, and the network does not see enough of them during training. This results in lower AP values for these two categories across all experiments. However, the improvement trends still hold as discussed before and all implementations of Radatron still outperform the three baselines in all categories consistently across all days.

6.13 ADDITIONAL QUALITATIVE RESULTS

We show additional *randomly sampled* qualitative results samples from our test set in Fig. 6.4. We also compare Radatron's performance against baselines using stand-alone cascaded radar without our motion-induced distortion compensation algorithm and single chip radar similar to the ones used in recent radar datasets [2, 21, 22, 59].

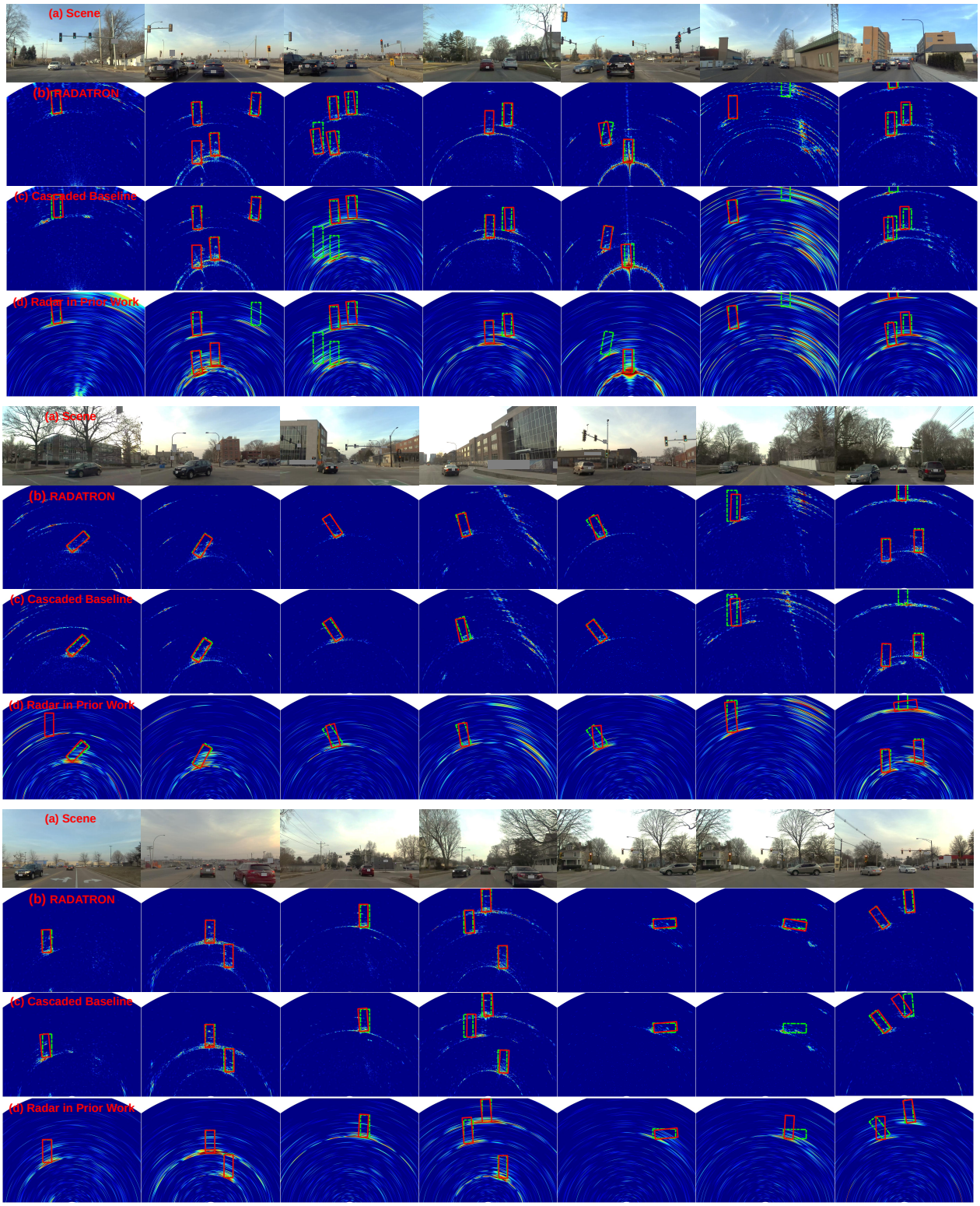

Figure 6.4: Randomly sampled examples from our test set. Ground truth is marked in green and predictions in red. Row (a) shows the original scene. Row (b) shows Radatron's performance overlaid on distortion compensated radar heatmaps. Row (c) and (d) show the performances of our baselines with stand-alone cascaded radar and the radar used in prior work along with their input radar heatmaps respectively.

CHAPTER 7: CONCLUSION

This thesis presents Radatron, a mmWave radar-based object detection system capable of working in adverse weather conditions, and also introduces the first-of-its-kind high-resolution automotive radar dataset. Radatron achieves accurate bounding box detection for cars using a hybrid signal processing and deep learning solution. The work is aimed at solving the prediction and perception challenges faced by the state-of-the-art in the self-driving cars regime, where cameras and LiDARs are still the go-to sensors despite their failure to function well in bad weather scenarios like snow, fog, rain etc. Despite significant improvement in detection performance over prior work, there are still some limitations which need to be addressed. First, the maximum range of Radatron's radar was configured to 25m to match that of our stereo camera [54]. Hence, our dataset does not include cars beyond 25m. Second, Radatron does not leverage the 3D nature of its high resolution datasets, which could potentially be used to detect 3D bounding boxes. Third, Radatron was trained and tested using data collected in the same country and may not work as well in other locations. Finally, Radatron currently only detecs vehicles but could be expanded to more objects like pedestrians and bikes by annotating these classes. Addressing these limitations is left for future work.

REFERENCES

[1] J. Guan, S. Madani, S. Jog, S. Gupta, and H. Hassanieh, "Through fog high-resolution imaging using millimeter wave radar," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[2] Y. Wang, G. Wang, H.-M. Hsu, H. Liu, and J.-N. Hwang, "Rethinking of radar's role: A camera-radar dataset and systematic annotator via coordinate alignment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 2815–2824.

[3] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 6433–6438.

[4] M. Sheeny, E. De Pellegrin, S. Mukherjee, A. Ahrabian, S. Wang, and A. Wallace, "Radiate: A radar dataset for automotive perception," *arXiv preprint arXiv:2010.09076*, vol. 3, no. 4, p. 7, 2020.

[5] B. Major, D. Fontijne, A. Ansari, R. T. Sukhavasi, R. Gowaikar, M. Hamilton, S. Lee, S. Grzechnik, and S. 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)*, 2019, pp. 924–932.

[6] M. Mostajabi, C. M. Wang, D. Ranjan, and G. Hsyu, "High-resolution radar dataset for semi-supervised learning of dynamic objects," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 100–101.

[7] A. Ouaknine, A. Newson, J. Rebut, F. Tupin, and P. Pérez, "Carrada dataset: camera and automotive radar with range-angle-doppler annotations," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 5068–5075.

[8] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, and F. Heide, "Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 682–11 692.

[9] G. Satat, M. Tancik, and R. Raskar, "Towards photography through realistic fog," in *2018 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2018, pp. 1–10.

[10] N. Y. Times, "5 things that give self-driving cars headaches," https://www.nytimes.com/interactive/2016/06/06/automobiles/autonomous-cars-problems.html, 2016.

[11] X. Dong, P. Wang, P. Zhang, and L. Liu, "Probabilistic oriented object detection in automotive radar," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 102–103.

[12] Y. Wang, Z. Jiang, X. Gao, J.-N. Hwang, G. Xing, and H. Liu, "Rodnet: Radar object detection using cross-modal supervision," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2021, pp. 504–513.

[13] X. Gao, G. Xing, S. Roy, and H. Liu, "Ramp-cnn: A novel neural network for enhanced automotive radar object recognition," *IEEE Sensors Journal*, vol. 21, no. 4, p. 5119–5132, Feb 2021.

[14] K. Qian, S. Zhu, X. Zhang, and L. E. Li, "Robust multimodal vehicle detection in foggy weather using complementary lidar and radar signals," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 444–453.

[15] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3354–3361.

[16] Texas Instruments Inc., "mmWave cascade imaging radar RF evaluation module," https://www.ti.com/tool/MMWCAS-RF-EVM, 2022, [Online; accessed mar-7-2022].

[17] Uhnder Inc., "Uhnder - Digital Automotive Radar," https://www.uhnder.com/, 2022, [Online; accessed mar-7-2022].

[18] S. Cho and S. Lee, "Fast motion deblurring," in *ACM SIGGRAPH Asia 2009 papers*, 2009, pp. 1–8.

[19] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *Acm transactions on graphics (tog)*, vol. 27, no. 3, pp. 1–10, 2008.

[20] K. Bansal, K. Rungta, S. Zhu, and D. Bharadia, "Pointillism: Accurate 3d bounding box estimation with multi-radars," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, ser. SenSys '20, 2020, p. 340–353.

[21] X. Gao, G. Xing, S. Roy, and H. Liu, "Experiments with mmwave automotive radar test-bed," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2019, pp. 1–6.

[22] F. E. Nowruzi, D. Kolhatkar, P. Kapoor, F. Al Hassanat, E. J. Heravi, R. Laganiere, J. Rebut, and W. Malik, "Deep open space segmentation using automotive radar," in *2020 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*. IEEE, 2020, pp. 1–4.

[23] A. Zhang, F. E. Nowruzi, and R. Laganiere, "Raddet: Range-azimuth-doppler based radar object detection for dynamic road users," *arXiv preprint arXiv:2105.00363*, 2021.

[24] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck, and K. Dietmayer, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1341–1360, 2020.

[25] Z. Zhang, Z. Tian, and M. Zhou, "Latern: Dynamic continuous hand gesture recognition using fmcw radar sensor," *IEEE Sensors Journal*, vol. 18, no. 8, pp. 3278–3289, 2018.

[26] M. Zhao, T. Li, M. A. Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, "Through-wall human pose estimation using radio signals," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7356–7365.

[27] M. Zhao, Y. Tian, H. Zhao, M. A. Alsheikh, T. Li, R. Hristov, Z. Kabelac, D. Katabi, and A. Torralba, "Rf-based 3d skeletons," in *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, ser. SIGCOMM '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 267–281.

[28] M. Zhao, Y. Liu, A. Raghu, H. Zhao, T. Li, A. Torralba, and D. Katabi, "Through-wall human mesh recovery using radio signals," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 10 112–10 121.

[29] T. Li, L. Fan, M. Zhao, Y. Liu, and D. Katabi, "Making the invisible visible: Action recognition through walls and occlusions," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 872–881.

[30] C. X. Lu, S. Rosa, P. Zhao, B. Wang, C. Chen, J. A. Stankovic, N. Trigoni, and A. 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*, ser. MobiSys '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 14–27.

[31] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 621–11 631.

[32] M. Meyer and G. Kuschk, "Automotive radar dataset for deep learning based 3d object detection," in *2019 16th European Radar Conference (EuRAD)*. IEEE, 2019, pp. 129–132.

[33] O. Schumann, C. Wöhler, M. Hahn, and J. 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)*, 2017, pp. 1–6.

[34] O. Schumann, M. Hahn, J. Dickmann, and C. Wöhler, "Semantic segmentation on radar point clouds," in *2018 21st International Conference on Information Fusion (FUSION)*, 2018, pp. 2179–2186.

[35] A. Danzer, T. Griebel, M. Bach, and K. Dietmayer, "2d car detection in radar data with pointnets," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 61–66.

[36] C. R. Qi, H. Su, K. Mo, and L. 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*, 2017, pp. 652–660.

[37] M. Meyer, G. Kuschk, and S. Tomforde, "Graph convolutional networks for 3d object detection on radar data," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3060–3069.

[38] A. Ouaknine, A. Newson, P. Perez, F. Tupin, and J. Rebut, "Multi-view radar semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 15 671–15 680.

[39] M. Shah, Z. Huang, A. Laddha, M. Langford, B. Barber, S. Zhang, C. Vallespi-Gonzalez, and R. Urtasun, "Liranet: End-to-end trajectory prediction using spatio-temporal radar fusion," 2020.

[40] B. Yang, R. Guo, M. Liang, S. Casas, and R. Urtasun, "Radarnet: Exploiting radar for robust perception of dynamic objects," in *European Conference on Computer Vision*. Springer, 2020, pp. 496–512.

[41] T.-Y. Lim, A. Ansari, B. Major, D. Fontijne, M. Hamilton, R. Gowaikar, and S. Subramanian, "Radar and camera early fusion for vehicle detection in advanced driver assistance systems," *NeurIPS Machine Learning for Autonomous Driving Workshop*, 2019.

[42] J. Kim, Y. Kim, and D. Kum, "Low-level sensor fusion network for 3d vehicle detection using radar range-azimuth heatmap and monocular image," in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, November 2020.

[43] S. Chadwick, W. Maddern, and P. Newman, "Distant vehicle detection using radar and vision," 2019.

[44] Y. Long, D. Morris, X. Liu, M. Castro, P. Chakravarty, and P. Narayanan, "Radar-camera pixel depth association for depth completion," 2021.

[45] R. Nabati and H. Qi, "Centerfusion: Center-based radar and camera fusion for 3d object detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2021, pp. 1527–1536.

[46] Y. Kim, J. W. Choi, and D. Kum, "Grif net: Gated region of interest fusion network for robust 3d object detection from radar point cloud and monocular image," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 10 857–10 864.

[47] C. Iovescu and S. Rao, "The fundamentals of millimeter wave sensors," *Texas Instruments*, pp. 1–8, 2017.

[48] J. Bechter, F. Roos, and C. Waldschmidt, "Compensation of motion-induced phase errors in tdm mimo radars," *IEEE Microwave and Wireless Components Letters*, vol. 27, no. 12, pp. 1164–1166, 2017.

[49] A. Manikas, *Beamforming: Sensor Signal Processing for Defence Applications*. World Scientific, 2015, vol. 5.

[50] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," https://github.com/facebookresearch/detectron2, 2019.

[51] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Icml*, 2010.

[52] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. PMLR, 2015, pp. 448–456.

[53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[54] Stereolabs Inc., "Zed Stereo Camera," https://www.stereolabs.com/zed/, 2022, [Online; accessed mar-7-2022].

[55] Y. Golovachev, et. al., "Millimeter wave high resolution radar accuracy in fog conditions-theory and experimental verification," *Sensors*, vol. 18, no. 7, p. 2148, 2018.

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

[57] O. Sorkine, "Least-squares rigid motion using svd," *Technical notes*, vol. 120, no. 3, p. 52, 2009.

[58] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[59] T.-Y. Lim, S. A. Markowitz, and M. N. Do, "Radical: A synchronized fmcw radar, depth, imu and rgb camera data dataset with low-level fmcw radar signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 941–953, 2021.

[60] Y. Golovachev, A. Etinger, G. Pinhasi, and Y. Pinhasi, "Propagation properties of sub-millimeter waves in foggy conditions," *Journal of Applied Physics*, vol. 125, no. 15, p. 151612, 2019.

[61] J. S. Lu, P. Cabrol, D. Steinbach, and R. V. Pragada, "Measurement and characterization of various outdoor 60 ghz diffracted and scattered paths," in *2013 IEEE Military Communications Conference*, Nov. 2013, pp. 1238–1243.