Radatron: Accurate Detection Using Multi-Resolution Cascaded MIMO Radar

Sohrab Madani*[1], Junfeng Guan*[1], Waleed Ahmed*[1], Saurabh Gupta[1], and Haitham Hassanieh[2]

[1] University of Illinois Urbana-Champaign
[2] EPFL
* indicates equal contribution.

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 paper, 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. Code and dataset is available on `https://jguan.page/Radatron/`.

1 Introduction

Recently, there has been a significant amount of work, from both academia [14, 49, 2, 43] and industry [26, 30, 35, 3], 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 [38, 46]. 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(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 [8, 48, 11] or must fuse radar with LiDAR or cameras to enable object detection [37, 3]. In this paper, we focus on exploring how well radar performs in object detection tasks and devise techniques to improve its performance.

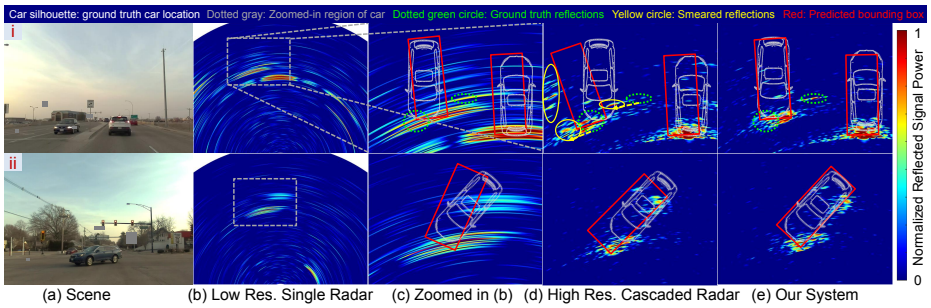

Fig. 1: The low resolution of millimeter wave radar makes it difficult to perform accurate bounding box detection in (c). High resolution cascaded MIMO radars can improve the resolution but suffer from motion smearing in (d). Radatron delivers accurate detection in (e) by combining motion compensation with a two stream deep learning architecture that takes low and high resolution radar images as input.

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 [11]. For example, in order to achieve 0.1° angular resolution similar to LiDAR [12], 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 [45, 47]. 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(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 [6, 42]. 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(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 paper, we present Radatron, a mmWave radar-based object detection system that 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(d) with ones after pre-processing in Fig. 1(e). The remaining errors stem from scenarios where the relative speed of the cars is high (e.g. incoming cars, see sec. 4.2). 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.

The paper 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 [48, 11, 1], 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 [48, 1, 11]. 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 paper 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 we plan to release once the paper is accepted.

2 Related Work

A. Radar-based Datasets. Several radar datasets have recently been introduced using single TI chips [35, 10, 33, 49, 54], the Navtech CTS350-X radar device [43, 8, 2], or other low resolution and 1D radar device [30, 9]. 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 1. [43, 2] are the closest in terms of resolution but use a mechanically rotating horn antenna which results in a low frame rate of

Dataset	Dim.	Resolution			#Total Frames	#Labeled Frames	Frame Rate	Size	Ground Truth	Radar
		Azi.	Ele.	Range						
Nuscenes [4]	1D/2D	N/R	N/A	N/R	1.3 M	40 K	13 fps	5.5 hrs	LiDAR	N/R
CARRADA [35]	2D[1]	15°	N/A	20cm	12.7 K	7.2 K	10 fps	21 mins	Camera	AWR1642
CRUW [49]	2D[1]	15°	N/A	23cm	400 K	N/R[4]	30 fps	3.5 hrs	Camera	AWR1843
OXFORD [2]	2D	1.8°	N/A	17cm	240 K	0	4 fps	280 km[3]	N/A	CTS350-X
RADIATE [43]	2D	1.8°	N/A	17cm	200 K	44 K	4 fps	3 hrs	Camera	CTS350-X
Zendar [30]	2D	30°	N/A	18cm	400 K	11 K	10 fps	11 hrs	LiDAR	N/R
SCORP [33]	3D	15°	30°	12cm	4 K	4 K	10 fps	6.6 mins	Camera	AWR1843
RADDet [54]	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 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 [3, 14, 26, 8, 28] are not included. N/A: Not Applicable. N/R: Not Reported.

4 Hz, motion smearing that cannot be corrected in pre-processing, and inability to compute velocity from Doppler information in the radar signals.

Low-cost radar has been used with deep learning in applications such as hand-gesture recognition [55], imaging and tracking of the human body [56, 58, 57, 20], as well as indoor mapping [25]. We focus on using radar for autonomous driving where prior work comprises two groups:

1. Radar Point Clouds: Learning radar data in the format of point clouds is widely studied [40, 39, 7, 1]. [40, 39] demonstrated a semantic segmentation network on radar point clouds while [7] adjusts PointNets [36] for radar data to perform 2D object detection. Pointillism [1] 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 [29].

2. Radar Heatmaps: To avoid loss of information, radar data can be processed as heatmaps with range-angle-Doppler tensors [11, 34, 26, 29, 54]. 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 [34], object classification and center point detection [11], as well as 2D bounding box detection [26]. Other work feeds the 2D BEV range-angle heatmap into the network as an image [8]. Note that while [26, 8] 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

[1] The radar in [35, 49] 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.

with more complicated traffic intersections, parked cars on the curbside, and various clutters. In[54], 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, [29] uses graph neural network to achieve a 69% AP_{50} but their data and code are not available. 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 [41, 37, 53] while radar and monocular camera fusion has also been studied in [21, 18, 5, 24, 31, 19]. 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.

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) [17]. 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 [45]. 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

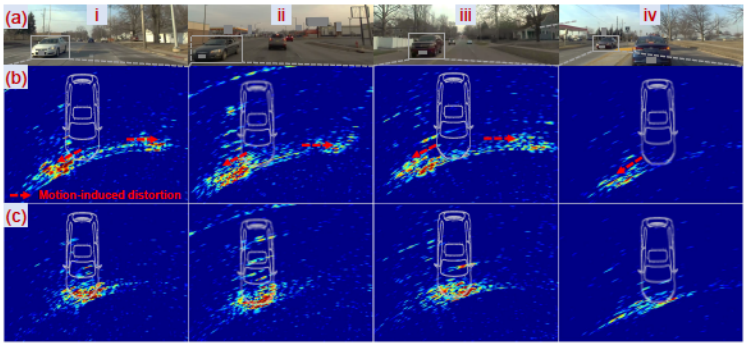

Fig. 2: Motion-induced distortion and Radatron's compensation algorithm. (a) Original scene. (b) Bird's-eye view radar heatmap under motion-induced distortion. (c) Processed heatmap after applying Radatron's motion compensation algorithm.

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 $\Delta\tau_{ij}$ between different virtual antennas as follows:

$$\Delta\tau'_{ij} = \tau_i - \tau_j + \delta t_{ij}\frac{2v}{c} = \Delta\tau_{ij} + \delta t_{ij}\frac{2v}{c}, \tag{1}$$

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 2(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.

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

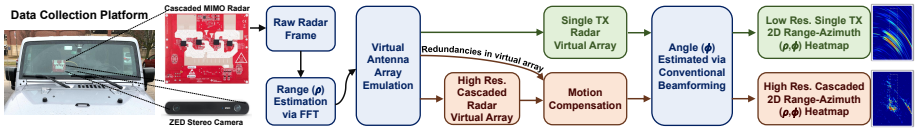

Fig. 3: Radatron's data pre-processing pipeline.

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 radar processing solution, and then proceed to describe our network design to tackle this problem.

4.1 Radar Signal Processing

On the signal processing end, we design a *motion compensation* algorithm and integrate it into our radar processing pipeline as shown in Figure 3. 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. We explain our algorithm in more detail in the supplementary material. Figure 2(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 [27] 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:

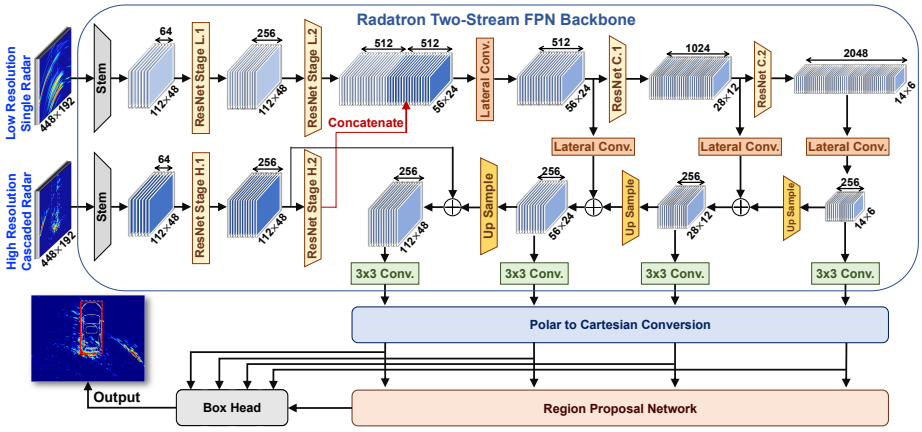

Fig. 4: Radatron's network architecture. We combine two branches of high resolution and low resolution radar data in an intermediate layer. For each feature map the number of channels and dimensions is indicated above and below it respectively.

High resolution cascaded radar: The radar heatmap is created using a uniform 86×1 virtual antenna array, emulated with multiple TX antennas.[3] It features the high azimuth resolution achieved using our cascaded MIMO radar.
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.

4.2 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. 2(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. 3. 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

[3] We describe the virtual antenna array emulation in the supp. material.

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 [51] which has been shown effective previously in [37, 29] for radar data. Fig. 4 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. 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.

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 [32] and BatchNorm [16]. Each branch then goes through two ResNet stages, which are the same ones used as the building blocks of ResNet50 [15]. 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 [51].

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.

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 [51] and add oriented boxes. Implementation details can be found in sec. 6.

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 [11], 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.

5 Radatron Dataset

Data Collection Platform: Our data collection platform consists of a TI-MMWCAS cascaded MIMO radar [45] and a ZED stereo camera [44] as shown in Fig. 3. 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 [48, 1, 11], 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. We show more details on our radar hardware in the supplementary material.

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 [52, 2, 50]. As a initial verification, we conducted controlled fog experiments to qualitatively evaluate Radatron's performance in fog.

6 Evaluation and Experiments

Evaluation Metrics. We use Average Precision (AP) as our main metric to evaluate Radatron's detection performance, following recent work [37, 29] 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 [23] to evaluate Radatron.
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 [35, 10, 49, 33, 43, 22, 11].
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.
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.

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

D. MVDNet: We train the radar-only version of MVDNet [37] using our dataset. We implement this baseline once with and once again without our compensation algorithm. As MVDNet only accepts one input, we compare it with the "high-res only" version of Radatron.

Radatron Variants. We implement three different variants of Radatron:

A. Radatron (No Compensation): We remove the motion compensation algorithm (4.1) from the signal processing pipeline.

B. Radatron (High-res Only): We remove the *top branch* from Fig. 4 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.

Dataset split. Out of 152K overall frames, we manually annotate 16K frames following sec. 5. 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.

Test set split. To show Radatron's performance under different difficulty scenarios following secs. 4.1 and 4.2, we split vehicles of the test set into 3 categories:

1. *straight*: Any vehicle on the same lane with an orientation within $\pm 5°$.
2. *oriented*: Any vehicle whose orientation is out of the $\pm 5°$ 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. 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.

6.1 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 2 , Radatron outperforms the prior work radar 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

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

Eval Metric		AP 50 (%)			AP 75 (%)		
Ablation	Split	str.	ori.	inc.	str.	ori.	inc.
MVDNet (w/o our comp.)		62.5	62.9	3.7	18.1	22.0	0.6
MVDNet (w/ our comp.)		90.5	81.2	41.9	43.8	39.1	8.0
Radatron (high-res only)		**94.7**	**90.7**	**73.1**	**61.4**	**56.3**	**34.6**

Table 4: Comparison with prior work with best results boldfaced.

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, 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. 4.1) as well as our fusion network (sec. 4.2) 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. 4.2, 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. 4.2 and 4.1, as the lack of motion compensation algorithm severely distorts the cascaded baseline, as shown in Fig. 2(b).

Finally, we compare Radatron with a radar-only version of MVDNet in Table 4. As seen, Radatron outperforms MVDNet by large margins. We note, however, that our motion compensation algorithm helps improve MVDNet's performance, especially for incoming cars by 38.2%. However, Radatron still outperforms MVDNet by 4.7%, 9.5%, and 31.2% for the three categories respectively in AP50. We attribute this improvement to the use of the ResNet FPN backbone and the polar to cartesian conversion in Radatron.

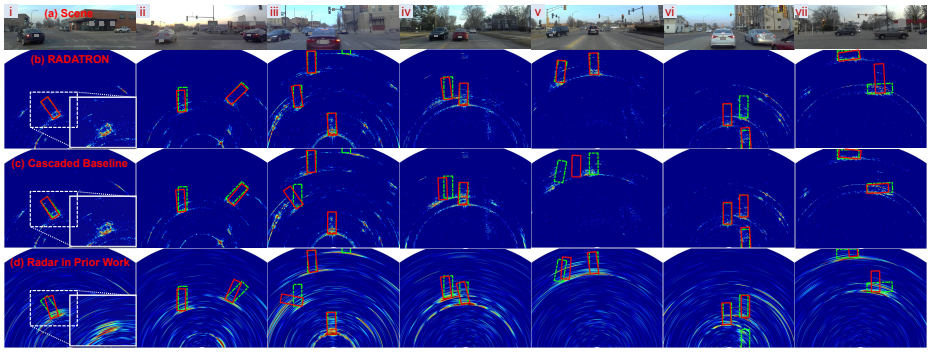

Fig. 5: Examples from our test set. Ground truth marked in green and predictions in red. (a) Original scene. Row (b) shows Radatron's performance overlaid on distortion compensated radar heatmaps. Row (c) and (d) show the performances of stand-alone cascaded and radar in prior work baselines along with their input heatmaps respectively.

6.2 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 3. 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 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 Radatron to improve the results on all metrics.

Ablation studies: We also perform extensive ablation studies to better study each component of Radatron. Specifically, we perform ablation studies on the impact of data augmentation, using different coordinate systems, fusing the high- and low-res versions at different stages, and using the Doppler information. The detailed results are presentend in the supp. material.

6.3 Qualitative Results

We show example qualitative results from our test set in Fig. 5, 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

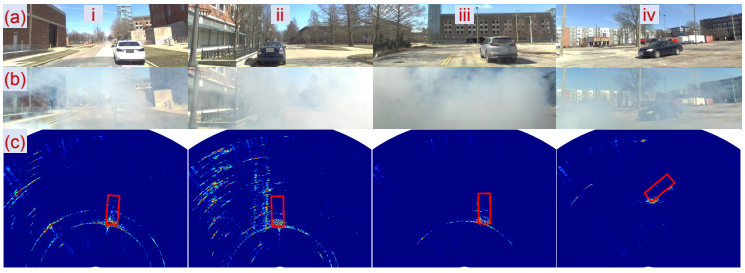

Fig. 6: Controlled Fog Experiment. (a) Original scene. (b) Scene in fog. (c) Prediction overlaid on radar heatmap captured in fog.

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. 5(c.iii-vi), due to motion-induced distortion as we described in sec. 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. 5(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. We show more results and failure mode analysis in supplementary material.

Controlled Fog Experiment. Figure 6 shows Radatron's performance in realistic fog emulated using a fog machine with high-density water-based fog fluid, following past work [52, 13]. As depicted in the figure, while the cars are not be visible in the RGB image, Radatron can accurately detect cars in the scene.

7 Limitations and Discussion

First, the maximum range of Radatron's radar was configured to 25m to match that of our stereo camera [44]. 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. Besides, our cascaded radar also provides Doppler information. We have conducted some initial experiments on leveraging this Doppler information. The implementation and results are presented in the supplementary material.

References

1. Bansal, K., Rungta, K., Zhu, S., Bharadia, D.: Pointillism: Accurate 3d bounding box estimation with multi-radars. In: Proceedings of the 18th Conference on Embedded Networked Sensor Systems. p. 340–353. SenSys '20 (2020)
2. Barnes, D., Gadd, M., Murcutt, P., Newman, P., Posner, I.: The oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset. In: 2020 IEEE International Conference on Robotics and Automation (ICRA). pp. 6433–6438. IEEE (2020)
3. Bijelic, M., Gruber, T., Mannan, F., Kraus, F., Ritter, W., Dietmayer, K., Heide, F.: 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. pp. 11682–11692 (2020)
4. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuscenes: A multimodal dataset for autonomous driving. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11621–11631 (2020)
5. Chadwick, S., Maddern, W., Newman, P.: Distant vehicle detection using radar and vision (2019)
6. Cho, S., Lee, S.: Fast motion deblurring. In: ACM SIGGRAPH Asia 2009 papers, pp. 1–8 (2009)
7. Danzer, A., Griebel, T., Bach, M., Dietmayer, K.: 2d car detection in radar data with pointnets. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). pp. 61–66 (2019)
8. Dong, X., Wang, P., Zhang, P., Liu, L.: Probabilistic oriented object detection in automotive radar. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 102–103 (2020)
9. Feng, D., Haase-Schütz, C., Rosenbaum, L., Hertlein, H., Glaeser, C., Timm, F., Wiesbeck, W., Dietmayer, K.: Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges. IEEE Transactions on Intelligent Transportation Systems **22**(3), 1341–1360 (2020)
10. Gao, X., Xing, G., Roy, S., Liu, H.: Experiments with mmwave automotive radar test-bed. In: 2019 53rd Asilomar Conference on Signals, Systems, and Computers. pp. 1–6. IEEE (2019)
11. Gao, X., Xing, G., Roy, S., Liu, H.: Ramp-cnn: A novel neural network for enhanced automotive radar object recognition. IEEE Sensors Journal **21**(4), 5119–5132 (Feb 2021)
12. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE conference on computer vision and pattern recognition. pp. 3354–3361. IEEE (2012)
13. Golovachev, Y., Etinger, A., Pinhasi, G., Pinhasi, Y.: Propagation properties of sub-millimeter waves in foggy conditions. Journal of Applied Physics **125**(15), 151612 (2019)
14. Guan, J., Madani, S., Jog, S., Gupta, S., Hassanieh, H.: 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)
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)

16. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. pp. 448–456. PMLR (2015)
17. Iovescu, C., Rao, S.: The fundamentals of millimeter wave sensors. Texas Instruments pp. 1–8 (2017)
18. Kim, J., Kim, Y., Kum, D.: 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)
19. Kim, Y., Choi, J.W., Kum, D.: 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). pp. 10857–10864. IEEE (2020)
20. Li, T., Fan, L., Zhao, M., Liu, Y., Katabi, D.: Making the invisible visible: Action recognition through walls and occlusions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 872–881 (2019)
21. Lim, T.Y., Ansari, A., Major, B., Fontijne, D., Hamilton, M., Gowaikar, R., Subramanian, S.: Radar and camera early fusion for vehicle detection in advanced driver assistance systems. NeurIPS Machine Learning for Autonomous Driving Workshop (2019)
22. Lim, T.Y., Markowitz, S.A., Do, M.N.: 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 **15**(4), 941–953 (2021)
23. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
24. Long, Y., Morris, D., Liu, X., Castro, M., Chakravarty, P., Narayanan, P.: Radar-camera pixel depth association for depth completion (2021)
25. Lu, C.X., Rosa, S., Zhao, P., Wang, B., Chen, C., Stankovic, J.A., Trigoni, N., Markham, A.: See through smoke: Robust indoor mapping with low-cost mmwave radar. In: Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services. p. 14–27. MobiSys '20, Association for Computing Machinery, New York, NY, USA (2020)
26. Major, B., Fontijne, D., Ansari, A., Sukhavasi, R.T., Gowaikar, R., Hamilton, M., Lee, S., Grzechnik, S., Subramanian, S.: Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). pp. 924–932 (2019)
27. Manikas, A.: Beamforming: Sensor Signal Processing for Defence Applications, vol. 5. World Scientific (2015)
28. Meyer, M., Kuschk, G.: Automotive radar dataset for deep learning based 3d object detection. In: 2019 16th European Radar Conference (EuRAD). pp. 129–132. IEEE (2019)
29. Meyer, M., Kuschk, G., Tomforde, S.: Graph convolutional networks for 3d object detection on radar data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3060–3069 (2021)
30. Mostajabi, M., Wang, C.M., Ranjan, D., Hsyu, G.: 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. pp. 100–101 (2020)

31. Nabati, R., Qi, H.: 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). pp. 1527–1536 (January 2021)
32. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Icml (2010)
33. Nowruzi, F.E., Kolhatkar, D., Kapoor, P., Al Hassanat, F., Heravi, E.J., Laganiere, R., Rebut, J., Malik, W.: Deep open space segmentation using automotive radar. In: 2020 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM). pp. 1–4. IEEE (2020)
34. Ouaknine, A., Newson, A., Perez, P., Tupin, F., Rebut, J.: Multi-view radar semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 15671–15680 (October 2021)
35. Ouaknine, A., Newson, A., Rebut, J., Tupin, F., Pérez, P.: Carrada dataset: camera and automotive radar with range-angle-doppler annotations. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 5068–5075. IEEE (2021)
36. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 652–660 (2017)
37. Qian, K., Zhu, S., Zhang, X., Li, L.E.: 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). pp. 444–453 (June 2021)
38. Satat, G., Tancik, M., Raskar, R.: Towards photography through realistic fog. In: 2018 IEEE International Conference on Computational Photography (ICCP). pp. 1–10. IEEE (2018)
39. Schumann, O., Hahn, M., Dickmann, J., Wöhler, C.: Semantic segmentation on radar point clouds. In: 2018 21st International Conference on Information Fusion (FUSION). pp. 2179–2186 (2018)
40. Schumann, O., Wöhler, C., Hahn, M., Dickmann, J.: 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). pp. 1–6 (2017). https://doi.org/10.1109/SDF.2017.8126350
41. Shah, M., Huang, Z., Laddha, A., Langford, M., Barber, B., Zhang, S., Vallespi-Gonzalez, C., Urtasun, R.: Liranet: End-to-end trajectory prediction using spatio-temporal radar fusion (2020)
42. Shan, Q., Jia, J., Agarwala, A.: High-quality motion deblurring from a single image. Acm transactions on graphics (tog) **27**(3), 1–10 (2008)
43. Sheeny, M., De Pellegrin, E., Mukherjee, S., Ahrabian, A., Wang, S., Wallace, A.: Radiate: A radar dataset for automotive perception. arXiv preprint arXiv:2010.09076 **3**(4), 7 (2020)
44. Stereolabs Inc.: Zed Stereo Camera. `https://www.stereolabs.com/zed/` (2022), [Online; accessed mar-7-2022]
45. Texas Instruments Inc.: mmWave cascade imaging radar RF evaluation module. `https://www.ti.com/tool/MMWCAS-RF-EVM` (2022), [Online; accessed mar-7-2022]
46. Times, N.Y.: 5 things that give self-driving cars headaches. `https://www.nytimes.com/interactive/2016/06/06/automobiles/autonomous-cars-problems.html` (2016)
47. Uhnder Inc.: Uhnder - Digital Automotive Radar. `https://www.uhnder.com/` (2022), [Online; accessed mar-7-2022]

48. Wang, Y., Jiang, Z., Gao, X., Hwang, J.N., Xing, G., Liu, H.: Rodnet: Radar object detection using cross-modal supervision. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 504–513 (January 2021)
49. Wang, Y., Wang, G., Hsu, H.M., Liu, H., Hwang, J.N.: 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. pp. 2815–2824 (2021)
50. Waymo: A fog blog. `https://blog.waymo.com/2021/11/a-fog-blog.html` (2021)
51. Wu, Y., Kirillov, A., Massa, F., Lo, W.Y., Girshick, R.: Detectron2. `https://github.com/facebookresearch/detectron2` (2019)
52. Y. Golovachev, et. al.: Millimeter wave high resolution radar accuracy in fog conditions-theory and experimental verification. Sensors **18**(7), 2148 (2018)
53. Yang, B., Guo, R., Liang, M., Casas, S., Urtasun, R.: Radarnet: Exploiting radar for robust perception of dynamic objects. In: European Conference on Computer Vision. pp. 496–512. Springer (2020)
54. Zhang, A., Nowruzi, F.E., Laganiere, R.: Raddet: Range-azimuth-doppler based radar object detection for dynamic road users. arXiv preprint arXiv:2105.00363 (2021)
55. Zhang, Z., Tian, Z., Zhou, M.: Latern: Dynamic continuous hand gesture recognition using fmcw radar sensor. IEEE Sensors Journal **18**(8), 3278–3289 (2018). https://doi.org/10.1109/JSEN.2018.2808688
56. Zhao, M., Li, T., Alsheikh, M.A., Tian, Y., Zhao, H., Torralba, A., Katabi, D.: Through-wall human pose estimation using radio signals. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7356–7365 (2018)
57. Zhao, M., Liu, Y., Raghu, A., Zhao, H., Li, T., Torralba, A., Katabi, D.: Through-wall human mesh recovery using radio signals. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 10112–10121 (2019)
58. Zhao, M., Tian, Y., Zhao, H., Alsheikh, M.A., Li, T., Hristov, R., Kabelac, Z., Katabi, D., Torralba, A.: Rf-based 3d skeletons. In: Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication. p. 267–281. SIGCOMM '18, Association for Computing Machinery, New York, NY, USA (2018)