Supplementary Materials for 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.

Table of Contents

We created a demo video to show Radatron's performance over a larger number of continuous frames Demo Video. Please check it out.

Appendix A: Consistency of Evaluation

As mentioned in sec. 6 of the paper, 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 3. For the results shown in sec. 6.1-6.2, we chose *Day 2* 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 1 and Table 2 show the results when we use *Day 3* and *Day 4* respectively as the test set, while using all other days for training. For both cases, following the trends reported in sec. 6.1, all three implementations of Radatron consistently outperform the three baselines. Next, we discuss the results for each day in detail.

Eval Metric		AP 50				AP 75				mAP			
Model	Split	str.	ori.	inc.	overall	str.	ori.	inc.	overall	str.	ori.	inc.	overall
Radar used 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 (No fusion)		**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		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 1: Quantitative results on Day #3. 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 used 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 (No fusion)		**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		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 2: Quantitative results on Day #4. Best performing model is boldfaced.

Day 3: 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 3* has significantly more oriented cars as shown in Table 3. 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.

Day 4: The trends seen for *Day 2* and *Day 3* more or less follow in case of *Day 4* 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.

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

Table 3: Distribution of different categories across all days

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 4: Ablation results. Best performing model is boldfaced.

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 3* and *Day 4*, the absolute AP values for oriented and incoming cars are lower than those reported for *Day 2* in sec. 6.1. This is because *Day 3* and *Day 4* 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.

Appendix B: Ablation Studies

Impact of data augmentation: To study the impact of the two forms of data augmentations applied (discussed in sec. 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 4 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. 4.2), while the 27.5% AP_{75} improvement for oriented cars shows affirms that angular shift can help with oriented vehicle predictions.

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 4 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. **??** 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.

Fusion at Different Stages: In sec. 4.2, 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 5 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.

Doppler: As we mentioned in sec. 7 in the paper, 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, 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

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

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 tasks such as speed estimation for future work.

Appendix C: 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 car is partially blocked by another car. In these scenarios, Radatron can either miss the occluded car, e.g. Fig. 1(1), or predict misplaced bounding boxes, e.g. Fig. 1(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. 1(3,4), or inaccurate orientation, e.g. Fig. 1(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 [6], as a result, even if the car is not occluded, reflections from some parts of the car cannot propagate back to the 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. 1(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

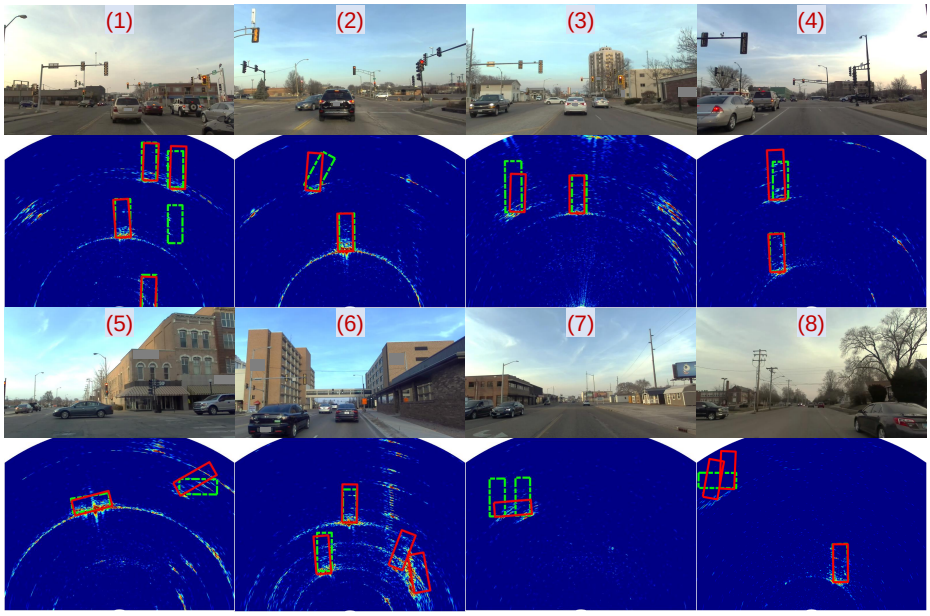

Fig. 1: Typical prediction errors in our test set. Ground truth is marked in green and predictions are marked in red. Top row of each example shows the original scene and the bottom row shows Radatron's predictions and ground truth bounding boxes overlaid on the input radar heatmaps.

confuses background reflections for cars. For example, in Fig. 1(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. 1(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. 1(8). Fortunately, as we discussed in appendix B, 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. 1(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

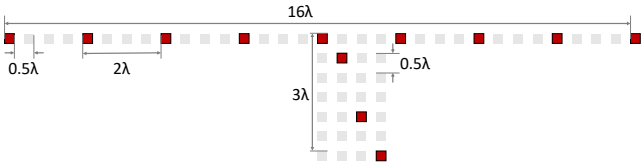

Fig. 2: Physical TX antenna array of Radatron's cascaded radar.

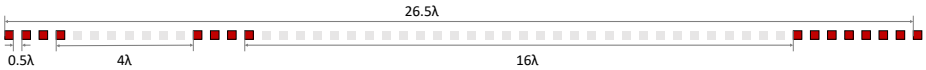

Fig. 3: Physical RX antenna array of Radatron's cascaded radar.

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.

Appendix D: Cascaded MIMO Radar System

We collect our own mmWave radar data featuring high angular resolution using TI MMWCAS mmWave cascaded MIMO radar [4]. 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.

Virtual Antenna Array Emulation: Fig. 2 shows the physical positions of the 12 TX antennas, while Fig 3 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. 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 paper. 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 paper.

Radar Configurations: We report our cascaded radar parameters as well as its configuration in our data collection experiments in Tab. 7.

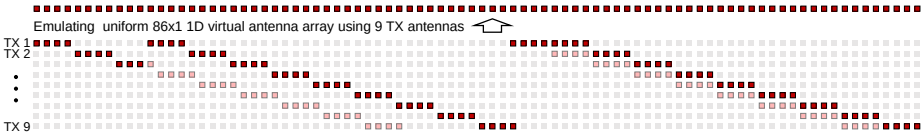

Fig. 4: Emulating large 1D virtual antenna array using Radatron's cascaded radar. Virtual antenna elements used to emulate large 1D array are marked in red, whereas unused virtual antenna elements are marked in pink.

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	$\sim 1.2^{\circ}$	Elevation Resolution	$\sim 18^{\circ}$

Table 7: Parameters and experimental configurations of Radatron's mmWave cascaded MIMO radar.

Appendix E: Mathematical Formulation of the Motion-Induced Distortion Problem

As we described in sec. 3 of the paper, temporal shifts in RF-signals translate to phase shifts of the electromagnetic waves. In other words, a difference in the Time-of-Flight (ToF) of two copies of the same signal that reaches two different antenna elements will translate into phase shifts.

There are two mechanisms that contribute to a ToF disparity between two differnet antenna elements. The first mechanism stems from the fact that displacement of antenna element i and j causes a slight difference in the path that the RF signal has to traverse before it reaches the two antennas. As this displacement occurs for all pairs of antennas, this ToF disparity effect holds for all TX and RX antennas. Here we denote The ToF delay between TX antenna i and RX antenna j with $\Delta\tau_{ij}$.

The second mechanism has to do with the movement of the objects in the environment. As the TX antennas take turns to send the same copy of the RF signal, the objects in the environment move ever so slightly between consecutive transmissions. Naturally, the movement of each object during this time increases depending on its speed v. More concretely, if TX antenna i and antenna j transmit their corresponding signals with δt_{ij} delay, then a ToF difference between

the two signals received at the same RX antennas will be

$$\delta t_{ij} \frac{2v}{c}, \tag{1}$$

where c is the speed of light. We now proceed to explain the challenge of motion-induced distortion and our solution in a more rigorous way.

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 is represented with the amplitude (A), frequency (f), and initial phase (θ_0) of a sinusoidal function. Therefore, for

$$\text{Signal} = A\sin(2\pi ft + \theta_0) \tag{2}$$

the corresponding phasor will be

$$\text{Phasor} = Ae^{j(2\pi ft+\theta_0)}. \tag{3}$$

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:

$$\text{FMCW Radar Waveform} = Ae^{j[2\pi(f_0 t+\frac{\alpha}{2}t^2)+\theta_0]}, \tag{4}$$

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:

$$\text{Reflected Signal} = e^{-j\{2\pi[f_0(t-\tau)+\frac{\alpha}{2}(t-\tau)^2]+\theta_0\}}. \tag{5}$$

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:

$$\text{Beat Frequency Signal} \approx e^{j2\pi(\alpha t\tau+f_0\tau)}, \tag{6}$$

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 $2\pi f_0\tau$.

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

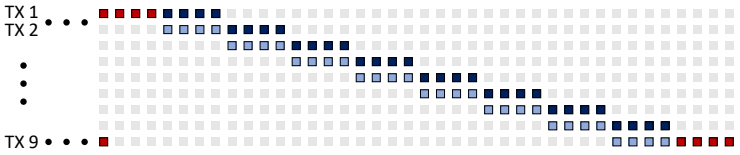

Fig. 5: Emulated co-located virtual antennas used for motion-induced phase variance estimation. Co-located virtual antenna pairs that are emulated using adjacent TX antennas (time gap equals single chirp interval) are marked in navy and light blue.

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

$$\Delta\theta_{ij} = 2\pi f_0 \Delta\tau_{ij} = 2\pi f_0(\tau_j - \tau_i) = 2\pi \frac{l\sin(\phi)}{\lambda}(j-i), \tag{7}$$

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. 7 becomes:

$$\Delta\theta'_{ij} = 2\pi \frac{lsin(\phi)}{\lambda}(j-i) + 2\pi f_0 \delta t_{ij} \frac{2v}{c} \tag{8}$$

As we has discussed in sec. 3 of the paper, 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.

Appendix F: 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 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. 8 becomes

$$\Delta\theta^{\dagger}_{ii'} = 2\pi f_0 \delta t_{ii'} \frac{2v}{c}. \tag{9}$$

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. 5. 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 the paper, 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 [3,1], 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.

Appendix G: Doppler Pre-Processing Algorithm

As we mentioned in sec. 7 in the paper, our cascaded radar also provides Doppler information, which we also try to leverage. Here, we describe our Doppler pre-processing algorithm.

As we have described in appendix C, we combine 9 TX chirps to create a range-azimuth (RA) radar heatmap. However, a radar frame further include 64 such chirp loops, which we leverage to extract Doppler information. Similar to how the we estimate the motion-induced phase variances, we can calculate the phase differences of the same virtual antenna over time ($\Delta\theta^{\dagger}$) to estimate the velocity-induce Doppler shift, and hence the velocity:

$$v = \frac{c}{4\pi f_0 \cdot 9T}\Delta\theta^{\dagger} = \frac{\lambda}{36\pi T}\Delta\theta^{\dagger} \tag{10}$$

where T is the time interval between consecutive chirps.

A standard algorithm applies another fast Fourier transform along the 64 chirp loops, that outputs a 3D range-azimuth-Doppler (RAD) radar tensors. Objects with different velocities are grouped into different bins along the Doppler dimension in the 3D radar tensors. Prior work [3,10,7,8,12] take this 3D radar tensor and collapse it into three different 2D radar feature maps for processing, and then recombined the encoded latent vectors. Considering the sparse 3D RAD

radar tensor, this multi-view network design also reduces the sparsity in each 2D feature maps, making it easier to learn.

However, simply applying the 3D radar tensor processing using Doppler FFT to our cascaded radar is also problematic. This is because the much long time gap between two chirps used for Doppler processing leads to aliasing in the Doppler/velocity domain. For example, in our experimental radar configuration, the time gap between when the same TX antenna transmits in adjacent chirp loops is $45.62\mu s \times 12 = 547\mu s$. This results in a the maximum unambiguous velocity of only ± 1.73 m/s. As a result, all objects whose velocities differ by $n \times 3.47$ m/s will end up in the same velocity/Doppler bin.

To resolve the velocity/Doppler ambiguity, we leverage the fact that the minimum time gap between chirps transmitted by our cascaded MIMO radar is only one chirp interval (T). This very short time gap can be leveraged to resolve a lot of aliasing. Therefore, we try to combine the 12 TX chirps in a chirp loop and the 64 chirp loops to achieve high-resolution and less aliased Doppler estimation. Unfortunately, chirps transmitted by adjacent TX antennas are not co-located, so that in addition to the phase variance introduced by motion, they also experience AoA dependent phase differences. Earlier, when we tried to accurately estimate AoA, we tried to disentangle these two sources of phase variances by compensating for the motion-induced phase. Here, in order to accurately estimate Doppler/velocity, we need to compensate for the AoA dependent phase differences instead.

To do so, we processed low-resolution range-azimuth (RA) heatmaps with every single TX antennas in every chirp loop separately, which provides us with $12 * 64 = 768$ 2D RA heatmaps. Every RA heatmaps is created using only one TX chirp with one chirp interval time gap in between. For each azimuth angle in these heatmaps, we compensate for the AoA dependent phase differences by multiplying with the complex conjugate of our TX antenna array steering vector. Then we take a fast Fourier transform along the 768 RA heatmaps, which outputs a 3D range-azimuth-Doppler (RAD) radar heatmap, whose azimuth resolution is the same as the low-resolution input RA heatmap to Radatron's network. This 3D RAD radar tensor has very high velocity resolution of 0.05 m/s, and a maximum unambiguous velocity of ± 20.85 m/s.

Although there are still residual aliasing along the Doppler dimension due to imperfect AoA phase compensation, the dominant velocity of each object always correspond to the highest power bin the the Doppler dimension. Therefore, we further take a $argmax$ operation along the Doppler dimension to extra the dominant velocity for each range-azimuth bin. In this way, the aliases in Doppler are neglected due to their lower power, and we can obtain a 2D Range-Azimuth Doppler index feature map, whose pixel values represent the dominant velocity of the corresponding range-azimuth bin. Moreover, the sparsity of the 3D RAD radar tensor also significantly reduced, making it much easier for a relatively smaller neural network model to learn. We concatenate this 2D RA Doppler index feature map as a second channel to the single-TX input of our network.

Appendix H: Additional Qualitative Results

We show additional *randomly sampled* qualitative results samples from our test set in Fig. 6. 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 [2,11,9,5].

Appendix I: Training Details

Here we provide the training details of our network.

• *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 $[0°, 180°]$, with 0.94° resolution. The output after conversion to Cartesian (sec. 4 of the paper) 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.

• *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 $-90°, \pm 45°$, and $0°$.

• *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.

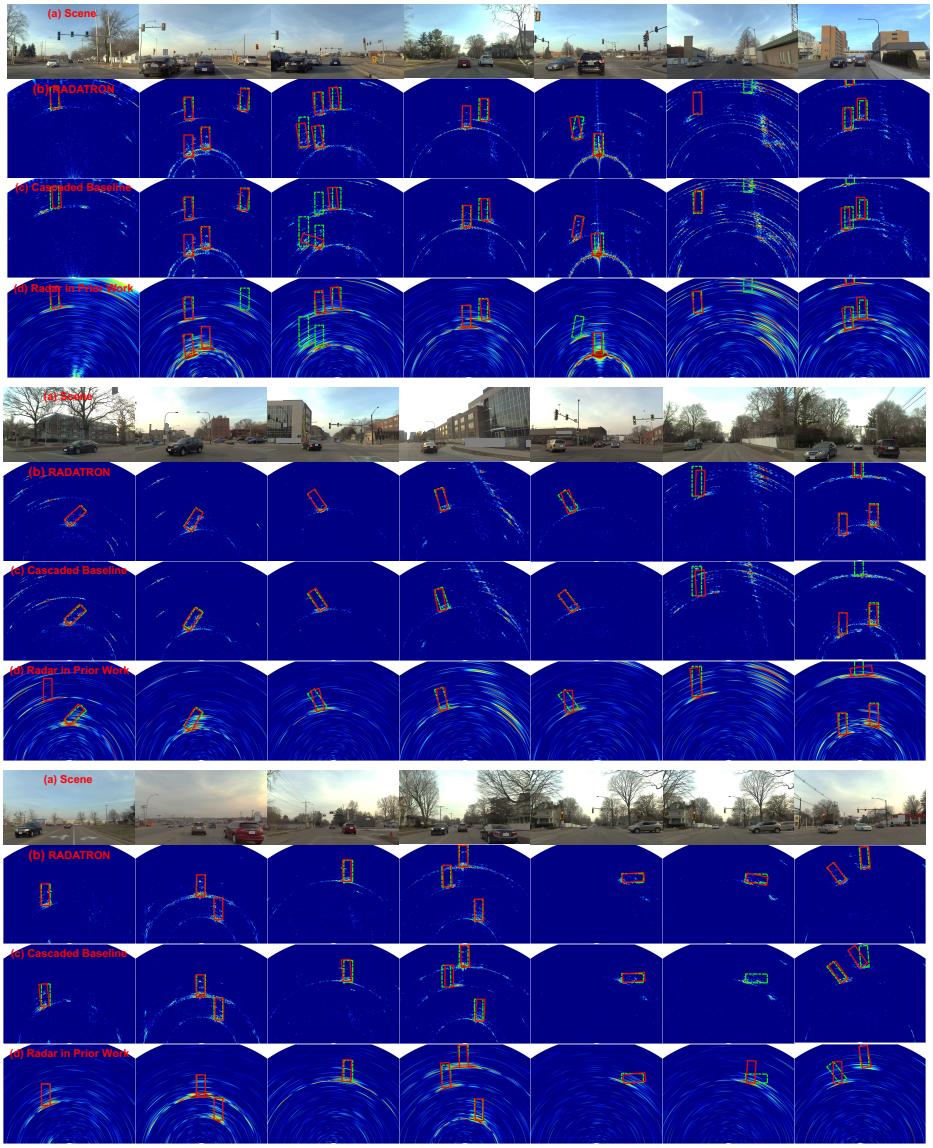

Fig. 6: Randomly sampled examples from our test set. Ground truth is marked in green and predictions in red. Row (a) shows the original scene. Row (b) shows Radatron's performance overlaid on distortion compensated radar heatmaps. Row (c) and (d) show the performances of our baselines with stand-alone cascaded radar and the radar used in prior work along with their input radar heatmaps respectively.

References

1. Bechter, J., Roos, F., Waldschmidt, C.: Compensation of motion-induced phase errors in tdm mimo radars. IEEE Microwave and Wireless Components Letters

27(12), 1164–1166 (2017). https://doi.org/10.1109/LMWC.2017.2751301 11
2. 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) 13
3. 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) 11
4. Inc., T.I.: mmWave cascade imaging radar RF evaluation module. https://www.ti.com/tool/MMWCAS-RF-EVM (2021), [Online; accessed oct-27-2021] 7
5. 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) 13
6. Lu, J.S., Cabrol, P., Steinbach, D., Pragada, R.V.: Measurement and characterization of various outdoor 60 ghz diffracted and scattered paths. In: 2013 IEEE Military Communications Conference. pp. 1238–1243 (Nov 2013). https://doi.org/10.1109/MILCOM.2013.212 5
7. 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). https://doi.org/10.1109/ICCVW.2019.00121 11
8. 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) 11
9. 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) 13
10. 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) 11
11. 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) 13
12. 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) 11