ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10094572

EXPLOITING VIRTUAL ARRAY DIVERSITY FOR ACCURATE RADAR DETECTION

Junfeng Guan[⋆†] *Sohrab Madani*[⋆] *Waleed Ahmed*[⋆] *Samah Hussein*[†]
Saurabh Gupta[⋆] *Haitham Hassanieh*[†]

[⋆] University of Illinois Urbana-Champaign [†] École polytechnique fédérale de Lausanne

ABSTRACT

Using millimeter-wave radars as a perception sensor provides self-driving cars with robust sensing capability in adverse weather. However, mmWave radars currently lack sufficient spatial resolution for semantic scene understanding. This paper introduces Radatron++, a system leverages cascaded MIMO (Multiple-Input Multiple-Output) radar to achieve accurate vehicle detection for self-driving cars. We develop a novel hybrid radar processing and deep learning approach to leverage the $10\times$ finer angular resolution while combating unique challenges of cascaded MIMO radars. We train and evaluate Radatron++ with a novel cascaded radar dataset. Radatron++ achieves 93.9% and 58.5% Average Precisions with 0.5 and 0.75 Intersection over Union thresholds respectively in 2D bounding box detection, outperforming prior work using low-resolution radars by 9.3% and 18.1% respectively.

1. INTRODUCTION

Using millimeter-wave (mmWave) radars for perception on self-driving cars has received increased interest in recent years [1, 2, 3, 4, 5, 6, 7, 8]. This is because mmWave radars have unique advantages over the dominant sensory modalities on today's autonomous vehicles, i.e., cameras and LiDARs. Radars can accurately measure distances and even velocities of objects, they are relatively cheap and can operate in adverse weather conditions such as fog, smog, snowstorms, and sandstorms where cameras and LiDARs fail [9, 10].

However, previous studies on mmWave radar-based semantic scene understanding such as object detection were restricted by the low angular resolution of radar, as shown in Fig. 1(b), so they can only coarsely localize objects [1, 11, 12]. Others must fuse radar with LiDARs or cameras to enable object detection [5, 13], or jointly leverage multiple radar frames to correct for inaccurate per-frame detection [14].

In this paper, we develop Radatron++, a mmWave radar-based perception system for self-driving cars that can detect precise bounding boxes of vehicles. Radatron++ leverages the novel cascaded MIMO radar, which combines multiple radar devices to emulate a much larger radar system with more transmitters (TX) and receivers (RX) [15, 16] to improve the angular resolution. However, the high angular resolution of cascaded MIMO radar comes at a price, because

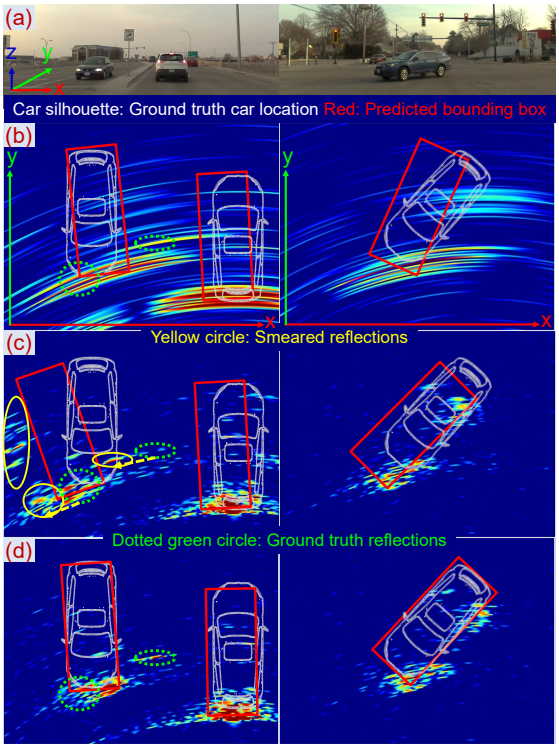

Fig. 1. Bird's-eye view heatmap and detection comparison between (b) low-resolution mmWave radar, (c) high-solution cascaded MIMO radar, (d) Radatron++'s multi-resolution imaging system.

the time-division multiplexing (TDM) of the TX antennas requires combining multiple transmissions to generate a high-resolution heatmap. In highly dynamic scenes like moving cars on the road, the Doppler phase shift across the multiple transmissions can blur and distort the resulting heatmaps [16, 17], as shown in Fig. 1(c). Reflections of objects get smeared and appear in different locations than where they really are, which leads to inaccurate bounding boxes prediction.

To overcome the challenges and fully leverage the high angular resolutions of cascaded MIMO radars, Radatron++ jointly leverages a variety of virtual array topologies that can be emulated from a single radar frame in a hybrid radar signal processing and deep learning solution. On the radar signal processing front, we first exploit a virtual array topology consisting of co-located virtual antennas to disentangle the Doppler phase shift in multiple transmissions and the spatial diversity-induced phase differences. We can hence compensate for the Doppler shift-induced distortion in the high-

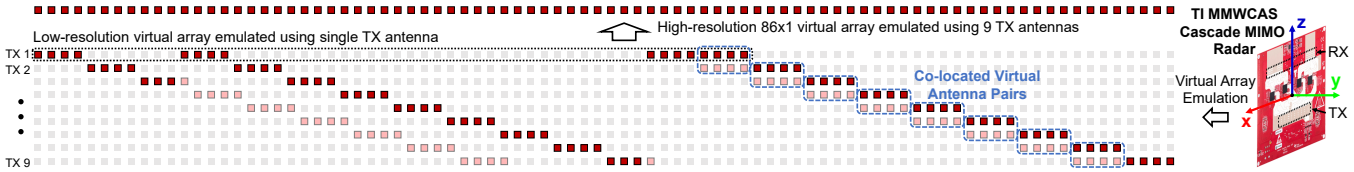

Fig. 2. Varieties of virtual antenna array formations leveraged by Radatron++.

resolution range-azimuth heatmaps. Moreover, Radatron++ manages to extract unambiguous Doppler information using a set of virtual arrays emulated with each TX antenna.

On the deep learning front, we design a novel radar-based object detection network, that incorporates the multi-resolution radar heatmaps from the cascaded MIMO radar. We utilize the richer perceptual information in the high-resolution heatmap for precise bounding box detection. Besides, virtual arrays emulated using a single TX do not suffer from Doppler shift-induced distortions, so even with lower resolution, they are useful in correcting faulty information like smeared or missed cars. Finally, Radatron++ also leverages the additional unambiguous Doppler information to better distinguish cars with different velocities.

Radatron++ is part of a broader radar-assisted perception project. The neural network design details and ablation studies were presented in [18]. This paper focuses on the MIMO radar signal processing algorithm. It also exploits the Doppler information to fully leverage cascaded radar data and to further improve detection performance. We train and extensively evaluate Radatron++ using the novel cascaded MIMO radar dataset we introduced in [18]. Our results show that Radatron++ improves overall detection accuracy by 9.3% for AP_{50} and 18.1% for AP_{75} compared to low-resolution radars used in prior work [1, 4, 12]. Besides, leveraging the additional Doppler information further improves the performance over Radatron [18] by 1.3% for AP_{50} and 2.2% for AP_{75}.

2. CHALLENGES FOR CASCADED MIMO RADAR

The angular resolution of radar is inversely proportional to the antenna aperture size [12], so it only scales linearly with the number of antennas and RF chains. MIMO radars, on the other hand, provide a much more scalable solution to achieve high angular resolution. With each TX-RX antenna pair emulating a *virtual* link, a MIMO radar with N TX and M RX can emulate N×M *virtual* links with only N+M *physical* antennas [15]. This allows the angular resolution to scale bilinearly with the number of antennas. However, MIMO radars require multiple transmissions from the TDM TX antennas to construct a frame, which leads to two unique challenges: (1) Doppler phase shift across transmissions is non-negligible and can severely distort the resulting heatmap [16, 17]. (2) Low frame rate leads to ambiguities in Doppler processing.

Doppler and Spatial Diversity Phase Entanglement: Standard radar processing algorithms estimate the angle-of-arrival (AoA) of reflections (ϕ) and construct range-azimuth (RA) heatmaps using the phase differences across the antenna array: $\frac{2\pi}{\lambda} d sin\phi$, where λ and d represent the wavelength and antenna spacing respectively. For MIMO radars, both TX and RX spatial diversities contribute to the phase differences across the virtual array. Besides, as TDM TXs take turns to transmit with switching delay T, transmissions from the k^{th} and l^{th} TXs received by the i^{th} and j^{th} RXs also experience Doppler phase shift. The resulting phase variation is:

$$\Delta\theta = \frac{2\pi}{\lambda}[\underbrace{sin\phi d(j-i)}_{\textit{RX Spatial Diversity}} + \underbrace{sin\phi d(l-k)}_{\textit{TX Spatial Diversity}} + \underbrace{2vT(l-k)}_{\textit{Doppler Shift}}] \quad (1)$$

In stationary scenes ($v \approx 0$), the Doppler phase shift is negligible. However, with significant relative velocity v, the Doppler phase shift is entangled with phase variations due to antenna spacial diversities and corrupts beamforming. Consequently, the RA heatmap gets distorted as reflections are smeared along the azimuth axis, as shown in Fig. 1(c).

Although Doppler phase compensation algorithms have been proposed for Synthetic Aperture Radar (SAR) and TDM MIMO radars [17], motion compensation for SAR only dissolves a constant Doppler shift throughout the scene caused by the ego motion of SAR. [17] also only considers an overly simple scenario with a single moving target. However, automotive MIMO radars suffer from various Doppler shifts in the scene resulting from relative motions between the moving ego vehicle and different objects in motion. Hence, conventional motion compensation techniques alone are inadequate.

Ambiguity in Doppler Velocity: Another downside caused by the low frame rate of cascaded MIMO radar is the limited maximum unambiguous velocity range that is proportional to the frame rate. Doppler phase shifts over multiple frames can be used to estimate the relative velocity of objects, but there can be ambiguities when the Doppler phase shift exceeds 2π and wraps over adjacent frames. For example, the frame interval in our radar configuration is 547μs, which results in a maximum unambiguous velocity of only ± 6.4 km/h. Therefore, all objects whose velocities (relative to the radar) differ by integer multiples of 12.8 km/h are indistinguishable.

3. PROCESSING DIVERSE VIRTUAL ARRAYS

We leverage a variety of emulatable virtual antenna array topologies to disentangle phase variations due to antenna spatial diversity and Doppler shift and resolve them separately. We design our processing algorithm for TI MMWCAS cascaded MIMO radar [15], a mono-static MIMO radar with 12 TX and 16 RX, but it can generalize to other MIMO radars.

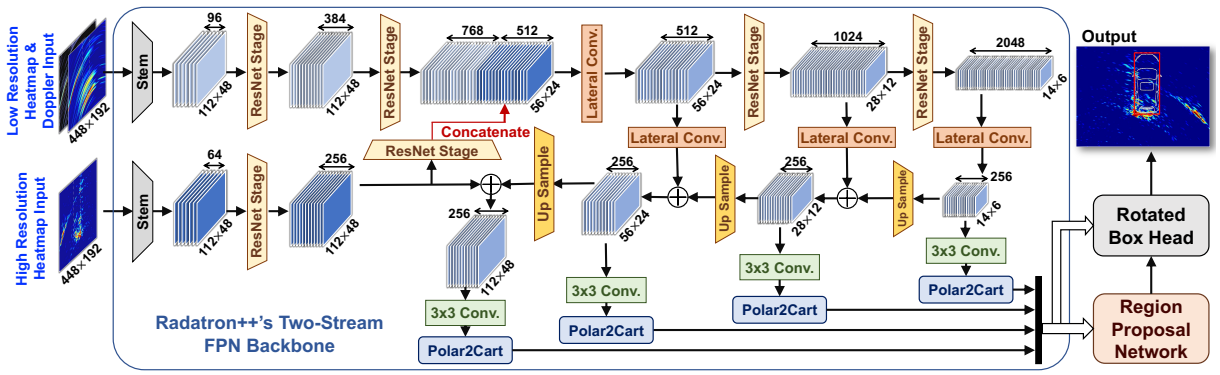

Fig. 3. Radatron++'s network architecture: Multi-resolution radar data are fused at an intermediate layer of the backbone.

Doppler Phase Compensation Leveraging Redundancies: We exploit *redundancies* in the virtual antenna array, i.e., co-located virtual antennas [17], to disentangle the Doppler phase shift from the spatial diversity-induced phase differences and compensate for the Doppler phase shifts. Under the far-field assumption, two TX-RX antenna pairs (i^T, k^R) and (j^T, l^R) emulate co-located virtual antennas if i^T is one unit spacing (d) to the left of j^T, while k^R is d to the right of l^R. The resulting phase centers of the TX-RX links overlap [19], and TX and RX spatial diversities cancel out with each other in Eqn. 1. Therefore, the only phase difference between co-located virtual antennas is the Doppler phase shift, which we can directly measure and then compensate for throughout the remaining virtual antenna array. In the TI MMWCAS MIMO radar topology, there are 32 co-located virtual antenna pairs with unit transmission time interval (T) as shown in Fig. 2. In the radar processing pipeline, after performing the range Fast Fourier transform (FFT), we measure the phase differences between each co-located virtual antenna pair for every range bin and then take the median phase difference of all co-located antenna pairs to reduce noises. We then scale the estimated Doppler phase shift over a unit time interval T by the corresponding transmission delays for all TX antennas. Finally, we compensate for the Doppler phase shift for all virtual antennas by multiplying with unit magnitude phasors with opposite phases. After compensating for the Doppler phase shifts, we can utilize the non-overlapping virtual antennas for beamforming. Figure 1(d) demonstrates an example of effective Doppler-induced distortion compensation.

Although the Doppler phase compensation algorithm works well in general, it has limitations and cannot resolve all the Doppler phase migration in the scene perfectly in the following challenging scenarios. First, the algorithm suffers when there are vehicles moving fast towards the ego-car, as the very high relative velocity results in Doppler ambiguities. Besides, the algorithm assumes that there is only one dominant object and Doppler shift in each range bin. When there are multiple Doppler shifts in a range bin, our algorithm can only correctly estimate and compensate for one of them. Note that, although one can leverage the 4-element arrays of co-located virtual antennas (as shown in Fig. 2) to coarsely resolve the directions of objects, the angular resolution is extremely low. Therefore, we only differentiate different ranges of objects and take the median of all Doppler shift measurements to reduce noise. To ensure accurate detection in these scenarios, Radatron++ further leverages the fact that virtual arrays emulated with a single TX antenna do not suffer from any Doppler phase migration. Therefore, Radatron++ jointly leverages lower-resolution heatmaps generated using a single TX antenna and fuses them with the high-resolution multi-TX heatmaps in the network model to further correct faulty information caused by residual Doppler distortions.

High-Resolution and Low-Ambiguity Doppler Processing: Standard Doppler processing methods estimate the Doppler shifts across the 64 chirp loops, resulting in velocity ambiguities. In contrast, Radatron++ leverages all 12 chips in a chirp loop with 45.6μs Chirp Repetitive Interval (CRI) for Doppler processing. Therefore, the Doppler sampling rate and maximum unambiguous velocity will be increased by 12× to $\frac{c}{4f\times\text{CRI}} \approx \pm 75.4$ km/h. With all $12 \times 64 = 768$ chirps, we can achieve a high-velocity resolution of $\frac{c}{2f\times\text{CRI}\times 768} \approx 0.2$ km/h at the same time.

Unfortunately, the 12 chirps in a chirp loop are transmitted by different TX antennas, so the spatial diversity of the TX antennas also contributes to the phase differences and entangles with the Doppler phase shift. To resolve the Doppler shift across these chirps, we need to disentangle the two sources of phase variations again, but this time, we aim to eliminate the phase differences resulting from spacial diversity instead. To do so, we emulate a 16-element virtual array for every chirp with a single TX antenna (as shown in Fig. 2) and process the corresponding low-resolution range-azimuth (RA) heatmaps. For every RA heatmap and every azimuth angle ϕ, we compute the TX array steering vector, which represents the spatial diversity-resulted phase shift as shown in Eqn. 1. We then eliminate the TX spatial diversity contribution of the phase variation across all RA heatmaps by multiplying it with the complex conjugates of the TX steering vector. Finally, we take a Doppler FFT along the time domain over the 768 RA heatmaps and obtain a 3D range-azimuth-Doppler (RAD) heatmap, where objects with different velocities are grouped into different bins along the Doppler dimension.

Table 1. **Performance against baselines.** Best performing model is boldfaced.

Model / Split	Straight			Oriented			Incoming			Overall		
Eval Metric	AP_{50}	AP_{75}	mAP	AP_{50}	AP_{75}	mAP	AP_{50}	AP_{75}	mAP	AP_{50}	AP_{75}	mAP
A. Radar in Prior work	88.6	45.0	47.3	73.9	24.0	34.4	69.4	24.6	31.2	84.6	40.4	44.2
B. Stand-alone single-TX	92.4	50.2	51.4	77.6	31.6	36.6	74.3	33.6	37.6	88.9	46.4	48.4
C. Stand-alone cascaded	87.7	42.9	45.5	80.9	31.9	38.1	65.9	26.2	30.9	84.6	39.8	43.2
D. Radatron (High-Res Only)	94.7	61.4	56.6	**90.7**	56.3	52.3	73.1	34.6	37.6	92.4	57.1	53.9
E. Radatron [18]	95.6	56.3	53.8	88.7	**57.1**	53.1	79.7	38.2	41.4	92.6	56.3	53.8
Radatron++	**96.2**	**61.8**	**56.8**	89.5	54.8	**53.2**	**79.9**	**40.6**	**41.7**	**93.9**	**58.5**	**54.5**

4. RADATRON++ NETWORK DESIGN

Our network jointly leverages three different versions of radar heatmaps generated using different virtual antenna array topologies. In addition to the Doppler compensated high-resolution heatmaps and the low-resolution but distortion-free single-TX heatmaps that are demonstrated in [18], Radatron++ also leverages the Doppler information extracted by our Doppler processing algorithm. Similar to prior work on radar-based object detection [8], our network design also adopts the Faster R-CNN framework [20], but it incorporates the multi-resolution radar inputs as described below.

Doppler Input: Unlike prior works that directly feed the sparse 3D RAD heatmap to a 3D backbone network [2, 4, 6, 12], we first take an $argmax$ operation along the Doppler dimension to obtain a 2D Doppler feature map, where the pixel value represents the dominant velocity for the range-azimuth pixel. This dimension compression significantly reduces the sparsity of the 3D RAD radar tensor, making it much easier for a smaller network model to learn. Besides, the Doppler feature map has the same dimensions as the single-TX heatmap, so it can be simply concatenated as a second channel to the single-TX input to the backbone network.

Backbone: The backbone network jointly extracts features from the three versions of radar heatmaps. We adopt the feature pyramid network [21] architecture, but we first use two parallel branches at the beginning of the bottom-up network to process the high and low-resolution versions of radar inputs separately and bring them into a common feature space. Each of the two branches first separately goes through a stem layer and two ResNet50 stages [22]. Different from Radatron [18], the number of channels for the low-resolution branch is always $1.5\times$ more than the high-resolution branch (e.g., 96 stem output channels vs 64), in order to encode both single-TX heatmap and Doppler feature map. We then fuse the two branches by concatenating their feature maps of the same dimension across channels and fuse them by applying a lateral 3×3 convolutional layer. We further encode the feature maps by passing them through ResNet stages and then combine them to create feature maps at different scales similar to [20].

RPN and Box Head: We feed RA heatmaps in polar coordinates to Radatron++'s backbone network, then we explicitly map the output features maps from polar coordinates to Cartesian coordinates using bi-linear interpolation, similar to [2]. We can thus use standard Region Proposal Network (RPN) and Box head architectures in Cartesian coordinates to predict 2D bounding boxes with orientations of vehicles.

5. EVALUATION

We train and evaluate Radatron++ on a novel cascaded radar dataset we introduced [18], so the radar configurations follow that in [18] with a maximum detection range of 25 m. We follow the COCO framework [23] to evaluate Radatron++, and our metrics are Average Precisions with 0.5 (AP_{50}) and 0.75 (AP_{75}) Intersection over Union (IoU) thresholds and mean AP (mAP) of IoU values from 0.5 to 0.95 with 0.05 steps. We split vehicle instances in the test set into 3 categories: 2854 straight, 327 oriented, and 512 incoming cars similar to [18].

We compare Radatron++ with the following baselines:

A.Radar in prior work: We implement a virtual array equivalent to the radar used in recent radar datasets [1, 3, 4, 12].

B.Stand-alone single-TX: We pass the single-TX heatmap alone to the stand-alone top stream in Fig. 3.

C.Stand-alone cascaded radar: We process the high-resolution heatmap conventionally without our Doppler compensation algorithm, and feed it to a stand-alone bottom stream in Fig. 3.

D.Radatron [18] (High-res Only): We only feed in the high-resolution processed radar data through the bottom branch.

E.Radatron [18]: Fusion network without Doppler input.

Table 1 summarizes our results. First, Radatron++ outperforms the low-resolution radar baselines (A and B) consistently across all metrics and the performance gaps are even more prominent for higher IoU thresholds and oriented cars. This proves that the higher angular resolution of cascaded MIMO radar is essential for accurate object detection. Next, one can see that using cascaded MIMO radar without any Doppler compensation (baseline C) suffers from distorted heatmaps, especially for highly dynamic objects like incoming cars. The comparison of baselines C and D indicates that our Doppler compensation algorithm can effectively mitigate the distortions and improve accuracies in all aspects. However, as we analyzed in section 3, the compensation algorithm is suboptimal in challenging scenarios, so we need to jointly leverage the distortion-free single-TX heatmaps. The fusion network improves the accuracies in incoming cars by $4 \sim 6\%$ at the cost of slightly lower accuracies for oriented cars.

Finally, with the additional Doppler input, Radatron++ outperforms Radatron [18] in most metrics with overall margins of 1.3% and 2.2% for AP_{50} and AP_{75} respectively. This result confirms our hypothesis that Doppler information can help distinguish car moving in directions with different speeds. We believe the Doppler information can be even more useful when extending bounding boxes detection to also estimate the velocities of vehicles.

6. REFERENCES

[1] Yizhou Wang, Gaoang Wang, Hung-Min Hsu, Hui Liu, and Jenq-Neng Hwang, "Rethinking of radar's role: A camera-radar dataset and systematic annotator via coordinate alignment," in *IEEE/CVF Conf. on Comput. Vis. and Pattern Recog.*, 2021, pp. 2815–2824.

[2] Bence Major, Daniel Fontijne, Amin Ansari, Ravi Teja Sukhavasi, Radhika Gowaikar, Michael Hamilton, Sean Lee, Slawomir Grzechnik, and Sundar Subramanian, "Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors," in *Int. Conf. Comput. Vis. Workshop*, 2019, pp. 924–932.

[3] Marcel Sheeny, Emanuele De Pellegrin, Saptarshi Mukherjee, Alireza Ahrabian, Sen Wang, and Andrew Wallace, "Radiate: A radar dataset for automotive perception in bad weather," in *IEEE Int. Conf. on Robotics and Automation*, 2021, pp. 1–7.

[4] Arthur Ouaknine, Alasdair Newson, Patrick Perez, Florence Tupin, and Julien Rebut, "Multi-view radar semantic segmentation," in *Int. Conf. Comput. Vis.*, Oct. 2021, pp. 15671–15680.

[5] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide, "Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather," in *IEEE/CVF Conf. on Comput. Vis. and Pattern Recog.*, 2020, pp. 11682–11692.

[6] Michael Meyer, Georg Kuschk, and Sven Tomforde, "Graph convolutional networks for 3d object detection on radar data," in *IEEE/CVF Int. Conf. on Comput. Vis.*, 2021, pp. 3060–3069.

[7] Junfeng Guan, Sohrab Madani, Suraj Jog, Saurabh Gupta, and Haitham Hassanieh, "Through fog high-resolution imaging using millimeter wave radar," in *IEEE Conf. on Comput. Vis. and Pattern Recog.*, 2020.

[8] Ruxin Zheng, Shunqiao Sun, David Scharff, and Teresa Wu, "Beyond point clouds: A knowledge-aided high resolution imaging radar deep detector for autonomous driving," *arXiv*, 2021.

[9] Guy Satat, Matthew Tancik, and Ramesh Raskar, "Towards photography through realistic fog," in *IEEE Int. Conf. on Computational Photography*, 2018, pp. 1–10.

[10] Gregory L Charvat, *Small and short-range radar systems*, CRC Press, 2014.

[11] Xu Dong, Pengluo Wang, Pengyue Zhang, and Langechuan Liu, "Probabilistic oriented object detection in automotive radar," in *IEEE Conf. on Comput. Vis. and Pattern Recog. Workshops*, 2020, pp. 102–103.

[12] Xiangyu Gao, Guanbin Xing, Sumit Roy, and Hui Liu, "Ramp-cnn: A novel neural network for enhanced automotive radar object recognition," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 5119–5132, Feb. 2021.

[13] Kun Qian, Shilin Zhu, Xinyu Zhang, and Li Erran Li, "Robust multimodal vehicle detection in foggy weather using complementary lidar and radar signals," in *IEEE/CVF Conf. on Comput. Vis. and Pattern Recog.*, Jun. 2021, pp. 444–453.

[14] Peizhao Li, Pu Wang, Karl Berntorp, and Hongfu Liu, "Exploiting temporal relations on radar perception for autonomous driving," in *IEEE/CVF Conf. on Comput. Vis. and Pattern Recog.*, 2022, pp. 17050–17059.

[15] Texas Instruments Inc., "mmWave cascade imaging radar RF evaluation module," 2022, [Online].

[16] Shunqiao Sun, Athina P. Petropulu, and H. Vincent Poor, "MIMO radar for advanced driver-assistance systems and autonomous driving: Advantages and challenges," *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 98–117, 2020.

[17] Christian M. Schmid, Reinhard Feger, Clemens Pfeffer, and Andreas Stelzer, "Motion compensation and efficient array design for tdma fmcw mimo radar systems," in *Eur. Conf. on Antennas and Propagation*, 2012, pp. 1746–1750.

[18] Sohrab Madani, Junfeng Guan, Waleed Ahmed, Saurabh Gupta, and Haitham Hassanieh, "Radatron: Accurate detection using multi-resolution cascaded mimo radar," in *Eur. Conf. on Comput. Vis.*, 2022.

[19] Jian Li and Petre Stoica, "Mimo radar with colocated antennas," *IEEE Signal Process. Magazine*, vol. 24, no. 5, pp. 106–114, 2007.

[20] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, vol. 28.

[21] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie, "Feature pyramid networks for object detection," in *IEEE/CVF Comput. Vis. and Pattern Recog.*, 2017, pp. 936–944.

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *IEEE/CVF CVPR*, 2016, pp. 770–778.

[23] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, "Microsoft coco: Common objects in context," in *Eur. Conf. on Comput. Vis.*, 2014, pp. 740–755.