SPECTRANET: A HIGH RESOLUTION IMAGING RADAR DEEP NEURAL NETWORK FOR AUTONOMOUS VEHICLES

Ruxin Zheng[1], *Shunqiao Sun*[1], *David Scharff*[1] *and Teresa Wu*[2,3]

[1]Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL 35487
[2]School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, 85281
[3]ASU-Mayo Center for Innovative Imaging, Arizona State University, Tempe, AZ, 85281

2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM) | 978-1-6654-0633-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/SAM53842.2022.9827798

ABSTRACT

The potentials of automotive radar for autonomous driving have not been fully exploited due to the difficulty of extracting targets' information from the radar signals and the lack of radar datasets. In this paper, a novel signal processing pipeline is proposed to address the max ambiguous velocity reduction issue introduced by staggered time division multiplexing (TDM) scheme of high resolution imaging radar system with a large number of transmit antennas. A dataset of $1,410$ synchronized frames (stereo cameras, LiDAR, radar) with three classes, i.e., bus, car, and people, is constructed from field experiments. Next, we implement a vanilla SpectraNet and show its promising performance on moving object detection and classification with a mean average precision (mAP) of 81.9% at an intersection over union (IoU) of 0.5.

***Index Terms*—Automotive radar, machine learning, deep neural network, autonomous vehicles**

I. INTRODUCTION

Automotive radar operating at millimeter-wave frequency, i.e., 76-81 GHz, plays an important role in autonomous driving systems due to its robustness in environment perception under all weather conditions [1], [2]. Existing automotive radar transceivers, such as NXP Semiconductors MR3003 and Texas Instruments AWR1243 [3], support up to 3 transmit and 4 receive antennas, yielding an angular resolution of around 10°, which is not capable for Level 4 and Level 5 autonomous driving where a vehicle drives itself in all conditions without any human interaction. High resolution imaging radar systems are highly desired for Level 4 and Level 5 autonomous driving to provide *point clouds* of the surrounding environment [1], [4]–[6]. Cascaded radar chips rendering 12 transmit and 16 receive antennas are being developed [7]–[9] to synthesize a large virtual array using multi-input multi-output (MIMO) radar technology at a low cost [1], [2], [10]. Several products are available with different array configurations, such as forward-looking full-range radar ARS540 by Continental [11]. With imaging radar, it is of great interest to investigate environment perception using deep neural networks.

The popular data sets in autonomous vehicle perception, such as KITTI [12] and Waymo Open Dataset [13] only contain camera and LiDAR recordings. Recently, datasets containing automotive radar have been published, such as nuScenes [14], Oxford Radar RobotCar [15], Astyx [16], RADIATE [17], CRUW [18], Zendar [19], CARRADA [20] and RadarScenes [21]. However, CARRADA and CRUW datasets are small sizes and the angular resolution of automotive radar is low, i.e., larger than 10°. Some data sets, such as RADIATE, Oxford Radar RobotCar, used a specific radar, such as mechanical scanning radar, which provided denser radar image. Yet, the Doppler information of targets is missing. Synthetic aperture radar (SAR) technology which is for static targets was adopted in Zendar dataset with multiple measurements from different vehicle locations. The Astyx dataset only contains 500 frames with sparse radar point clouds.

Radar datasets with point clouds suffer from information loss due to the thresholding operation in radar signal processing. To our best knowledge, there is no open dataset containing high resolution radar spectra obtained from a large virtual antenna array, stereo cameras images and LiDAR 3D point clouds. In this paper, we construct radar data representation without losing information for moving objects detection and classification. Our contributions are:

- We design a novel transmit and receive signal processing pipeline of imaging radar systems to generate high resolution radar spectra from raw data.
- With field experiments, we construct a high resolution radar range-azimuth spectra dataset, which encodes the critical details of object's shape.
- Based on our dataset, we propose a vanilla SpectraNet to detect moving targets of interests.

II. SYSTEM MODEL

In this section, we briefly present a novel radar signal processing chain to generate high resolution radar spectra.

II-A. FMCW Radar Principles

A FMCW radar transmits a *chirp*, which is a complex sinusoid signal whose frequency changes linearly with time. The transmit frequency, $f_T(t)$, for on chirp with bandwidth B and chirp duration T, can be expressed as

$$f_T(t) = f_c + \frac{B}{T}t, \tag{1}$$

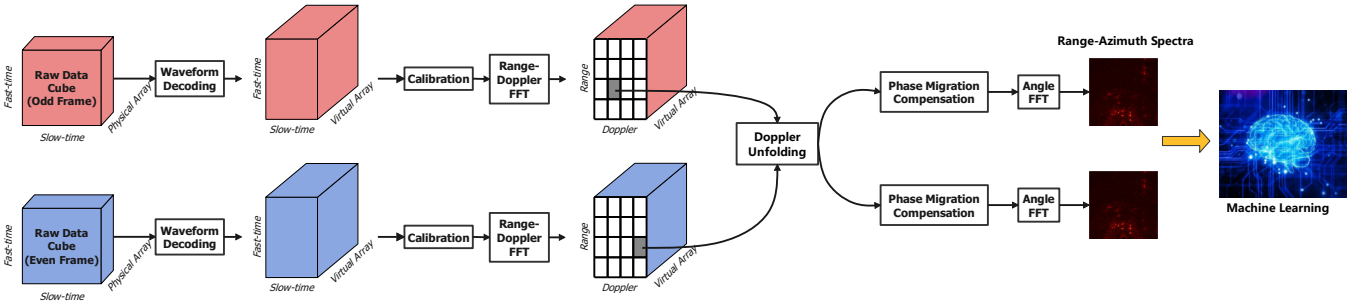


Fig. 1: Imaging radar signal processing pipeline to generate high resolution radar spectra for object detection with machine learning.

where f_c is carry frequency. The phase $\varphi_T(t)$ of the transmitted signal could be obtained after integration as $\varphi_T(t) = 2\pi \int_{-T/2}^{t} f_T(t)\, dt$. The noiseless received signal is a delayed version of transmit signal. For a target at range of R with radial velocity of v, the round-trip delay can be expressed as $\tau = 2(R+vt)/c$. The received signal is mixed with the transmit signal, and the output of the mixer is called *beat signal*, whose phase could be approximated as

$$\varphi_B(t) = 2\pi \left[\frac{2f_c R}{c} + \left(\frac{2f_c v}{c} + \frac{2BR}{Tc} \right) t \right], \qquad (2)$$

where the beat frequency is $f_b = f_R + f_D$ with $f_R = \frac{2BR}{Tc}$ being the range frequency and $f_D = \frac{2f_c v}{c}$ being the Doppler frequency. The beat signal typically goes through a band pass filter (BPF) to compensate the gain for targets in distance to improve the radar dynamic range, followed by an analog-to-digital converter (ADC), whose sampling rate is greater than twice of maximum beat frequency $f_b^{\max}$. Range and Doppler information of the target could be obtained by applying fast Fourier transforms (FFTs) along fast time and slow time.

II-B. High Resolution Imaging Radar Signal Processing

MIMO radar can synthesize a large virtual array for angle estimation using multiple transmit and multiple receive antennas [1], [10]. In MIMO radar, at transmitting side, multiple transmit antennas transmit orthogonal FMCW sequences; at receiving side, due to the waveform orthogonality, the contribution of each transmit antenna can be extracted from the received signal at each receive antenna. There are different ways to achieve waveform orthogonality in MIMO radar, such as Doppler-division multiplexing (DDM) and time division multiplexing (TDM) [1].

TDM was adopted in this paper to achieve waveform orthogonality due to its simplicity. However, under TDM, the maximum unambiguous detectable velocity is reduced to $v_{\max}/N_{\mathrm{TX}}$ with N_{TX} being the number of TX antennas [1]. As N_{TX} increases, the maximum unambiguous detectable velocity becomes small, and moving targets with relatively high speed will be aliased. We propose a staggered TDM scheme to resolve the Doppler ambiguity using the Chinese remainder theorem (CRT) and overlapped array elements. To improve the robustness of velocity unfolding, in this paper, we utilize both CRT and overlapped array. The narrow velocity candidate list obtained from CRT is used for compensating the phase migration among the overlapped array elements, and the candidate which gives the smallest phase difference will be chosen as the actual target velocity.

Under TDM scheme, only one TX antenna is selected to transmit at each time. The scheduling delay, Δt, between different transmit antennas would causes phase migration for moving targets between different chirps, i.e.,

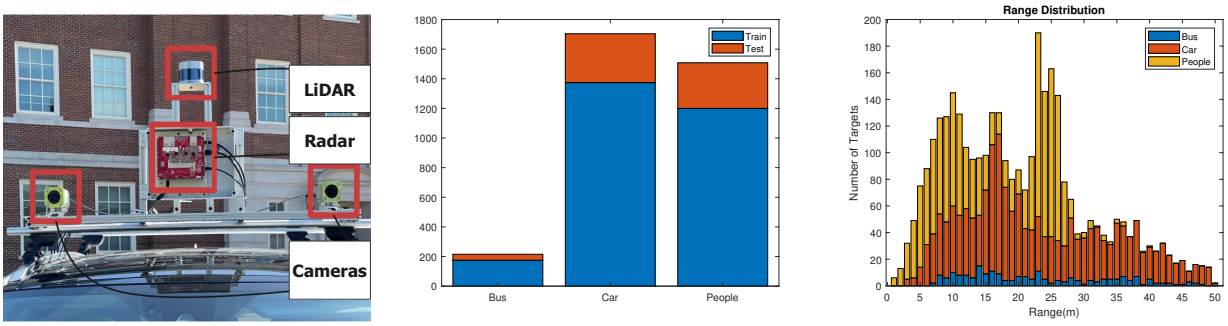
$$\phi = (4\pi/\lambda) v \Delta t. \qquad (3)$$

That phase migration introduces a distortion in the virtual array pattern, which results in inaccurate angle finding. In order to remove phase migration, for every moving target, a compensation value $e^{-j\phi}$ needs to be multiplied along with the virtual array before angle finding. The imaging radar signal processing pipeline is shown in Fig. 1. More details can be found in [22].

III. RADAR MACHINE LEARNING FRAMEWORK

We implemented the proposed imaging radar signal processing chain shown in Fig. 1 on TI imaging radar, which is a chirp configurable MIMO radar with 12 TX and 16 RX antennas, cascaded by 4 radar transceivers. The azimuth field of view (FOV) is 70°. A virtual uniform linear array with 86 elements and half wavelength spacing can be synthesized with 9 TX and 16 RX antennas, of which 32 virtual array elements are overlapped. The 3 dB beam width of the imaging radar in azimuth is $\Delta\theta_{\mathrm{AZ}} = 2\arcsin\left(\frac{1.4\lambda}{\pi D_x}\right) \approx 1.2^\circ$, where $D_x = 42.5\lambda$ is the virtual array aperture in horizontal direction. The parameters of consecutive frames are shown in Table I. Antenna calibration is required to reduce the frequency, phase, and amplitude mismatches across those 4 radar transceivers. A one-time boresight calibration method is used as our calibration method.

Our field experiments included three multi-modal sensors, i.e., a TI imaging radar, stereo cameras of Teledyne FLIR Blackfly S, and Velodyne Ultra Puck VLP-32C LiDAR, as shown in Fig. 2 (a). The measurements of cameras and



(a) (b) (c)

Fig. 2: Field experiments: (a) Multi-modal sensors; (b) Target classes distributions; (c) Target range distributions.

Parameters	**Frame** n	**Frame** $n+1$
Max Range	50 m	50 m
Max Velocity	7.2 m/s	5.5 m/s
Range Resolution	0.4 m	0.4 m
Angle Resolution	1.2°	1.2°

Table I: Radar Parameters

Fig. 3: Radar field experiments locations, where star ⋆ denotes the sensors' location.

LiDAR are used as ground truth for labeling the radar data. The sensor features are summarized in Table II.

Sensors	**Model**
Radar	TI Imaging Radar, Azimuth Resolution: 1.2°, Azimuth FOV: 70°
LiDAR	Velodyne Ultra Puck VLP-32C, Azimuth Resolution: 0.1° - 0.4° Vertical FOV: 40°, Maximum Range: 200 m
Camera	Teledyne FLIR Blackfly S, Stereo, Image Resolution: 2048×1536

Table II: Multi-Modal Sensors

III-A. High Resolution Imaging Radar Dataset

Reliable detection of moving objects, such as buses, cars, and people using high resolution imaging radar, is of great interest to autonomous vehicles. The radar field experiments were carried out at four locations at The University of Alabama campus, shown in Fig. 3. To avoid using consecutive radar frames for machine learning, frames were selected sparsely at random. After removing corrupted frames, the dataset contains a total of $1,410$ radar frames that are labeled using LiDAR 3D point clouds and camera images as ground truth. The target class distribution and range distribution are shown in Fig. 2 (b) and (c), respectively.

III-B. Target Detection with SpectraNet

Incorporating structured information into a perception algorithm using machine learning would avoid information loss through the constant false alarm rate (CFAR) detector and beamforming of angle finding. With the raw radar data, two range-azimuth spectra are generated corresponding to odd and even frames, using the signal processing pipeline shown in Fig. 1. Those spectra are then translated into Cartesian coordinates. The radar spectra generated by FFTs in range, Doppler, and spatial domains contain all the information about targets that is available in radar time series. Representative examples of high resolution bird's-eye-view (BEV) radar range-azimuth spectra are shown in Fig. 4, indicating that the high resolution range-azimuth spectra could represent the target shape and radar's performance is comparable to the LiDAR system.

We present a vanilla SpectraNet for fast object detection using high resolution radar spectra. The SpectraNet contains four convolutional blocks (Conv-4) with each consisting of a convolutional layer, a batch normalization (BN) layer, a ReLU layer, and max pooling layer, as the feature extractor backbone, followed by a YOLO detection head [23] due to its advantage in detection speed and accuracy. To avoid information loss, the radar range-azimuth spectra with the size of 512×512, are used as the input of vanilla SpectraNet, whose architecture is plotted in Fig. 5.

III-C. Performance Evaluation

Class-wise random sampling method was used to divide the dataset into training and testing sets with a ratio of $8 : 2$, ensuring that training and testing sets follow the overall dataset distribution. Additionally, a 5-fold cross-validation method was employed during the training process to find the optimal model by reducing data bias due to the lack of bus data in our dataset. The performance of SpectraNet was evaluated with various backbones, such as VGG16, VGG19 [24], and ResNet18 [25]. All networks were trained on a Dell workstation with dual Intel Xeon Gold 6230 CPU and an Nvidia Quadro RTX 4000 GPU. Radar

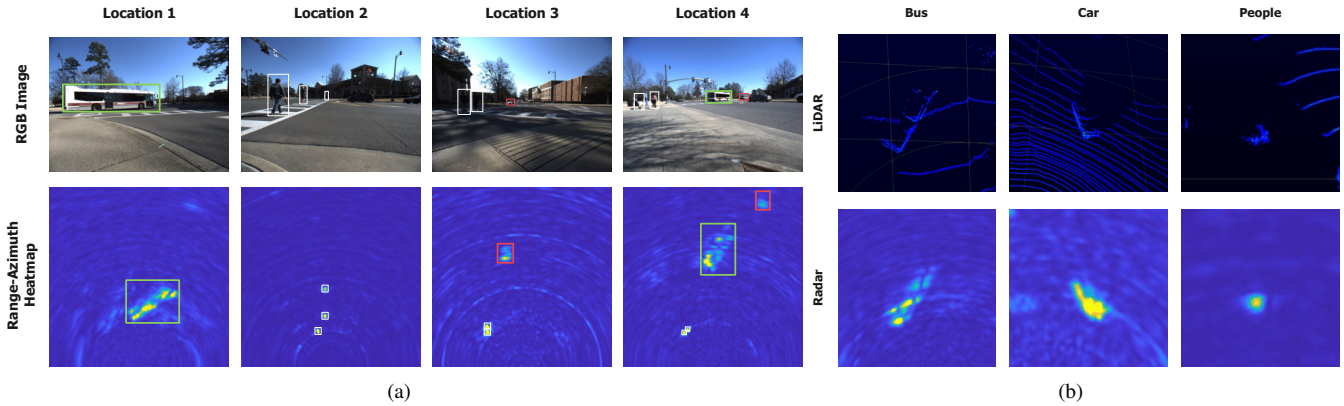


Fig. 4: Examples of high resolution radar spectra representing target's shape. (a) Field experiments at different locations, with the first row as the RGB images, and the second row as the corresponding range-azimuth heatmap in Cartesian coordinate. Buses, cars, and people are labeled with bounding boxes in green, red, and white colors, respectively. (b) Radar's performance is comparable to LiDAR, with 3D LiDAR point clouds BEV at the first row, and range-azimuth spectra at the second row.

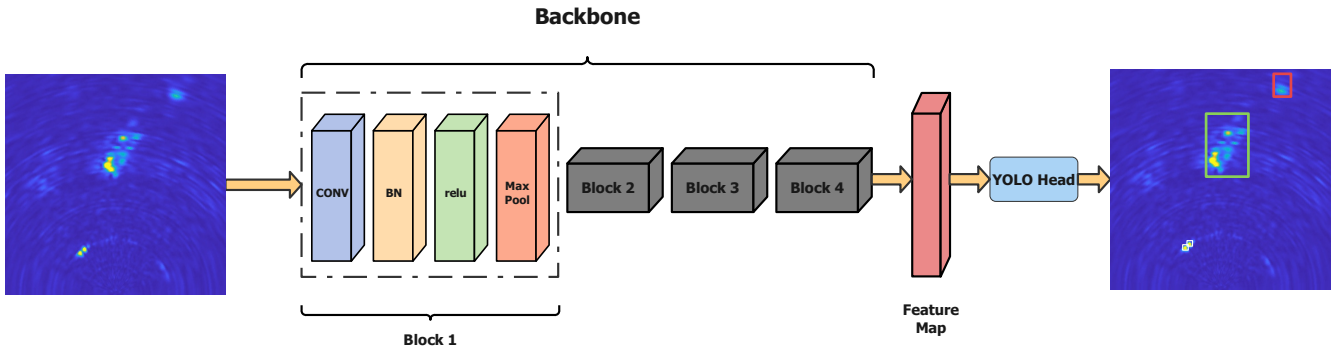


Fig. 5: Vanilla-SpectraNet has four convolutional blocks with each consisting of a convolutional layer, a batch normalization layer, a ReLU layer, and max pooling layer, as the feature extractor backbone, followed by a YOLO detection head.

Backbone	mAP	AP@0.5			AR@0.5		
		Bus	Car	People	Bus	Car	People
Conv-4	66.3%	56.4%	77.6%	64.9%	70.0%	86.7%	**80.8%**
VGG16	54.5%	36.7%	77.3%	49.6%	42.5%	83.0%	61.0%
VGG19	67.7%	63.0%	82.0%	58.1%	67.5%	87.3%	73.1%
ResNet18	**81.9%**	**87.4%**	**89.2%**	**69.0%**	**87.5%**	**90.3%**	72.1%
ResNet50	59.7%	52.2%	78.8%	48.1%	55.0%	82.1%	54.9%

Table III: Evaluation metrics of the SpectraNet.

machine learning performance is summarized in Table III. It is interesting to observe that deeper network (e.g., ResNet50) does not guarantee a better performance. We contend this may be because unlike RGB images which contain rich object information, radar spectra hold the reflection intensity of objects. It can be found that at 0.5 IoU, SpectraNet with ResNet18 as backbone yields the best overall performance in detecting bus, car, and people with average precision (AP) of 87.4%, 89.2%, 69.0%, and average recall (AR) of 87.5%, 90.3%, 72.1%, respectively. The corresponding confusion matrix is shown in Fig. 6.

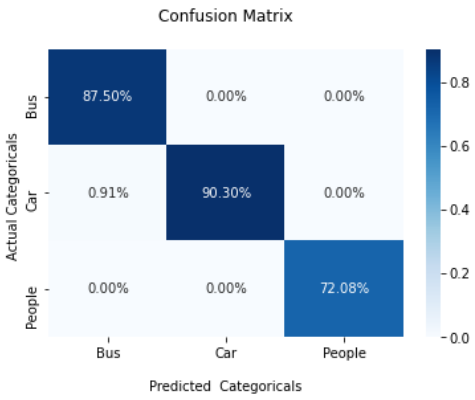


Fig. 6: Confusion matrix.

IV. CONCLUSIONS

In this paper, we presented a novel imaging radar signal processing chain to generate high resolution radar range-azimuth spectra without information loss and a vanilla SpectraNet for moving object detection. Through field experiments, we demonstrated the promising performance of moving object detection and classification using SpectraNet on high resolution radar range-azimuth spectra, that unlocks the radar potentials for Level 4/5 autonomous driving.

V. REFERENCES

[1] S. Sun, A. P. Petropulu, and H. V. 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.

[2] S. Sun and Y. D. Zhang, "4D automotive radar sensing for autonomous vehicles: A sparsity-oriented approach," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 4, pp. 879–891, 2021.

[3] Texas Instruments, "AWR1243 single-chip 77- and 79-GHz FMCW transceiver," datasheet, 2017.

[4] I. Bilik and *et al.*, "Automotive MIMO radar for urban environments," in *Proc. IEEE Radar Conference*, Philadelphia, PA, May 2016.

[5] F. Meinl, M. Stolz, M. Kunert, and H. Blume, "An experimental high performance radar system for highly automated driving," in *Proc. International Conference on Microwaves for Intelligent Mobility (ICMIM)*, Nagoya, Japan, Mar. 2017.

[6] S. Alland and *et al.*, "Virtual radar configuration for 2D array," U.S. Patent 9 869 762, Jan. 16, 2018.

[7] I. Bilik and *et al.*, "Automotive multi-mode cascaded radar data processing embedded system," in *Proc. IEEE Radar Conference*, Oklahoma City, OK, April 2018.

[8] V. Giannini and *et al.*, "A 192-virtual-receiver 77/79GHz GMSK code-domain MIMO radar system-on-chip," in *Proc. IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, CA, Feb. 17-21, 2019.

[9] Texas Instruments Inc., "Design guide: TIDEP-01012 imaging radar using cascaded mmWave sensor reference design (REV. A)," [Available Online] https://www.ti.com/lit/ug/tiduen5a/tiduen5a.pdf, Mar., 2020.

[10] J. Li and P. Stoica, "MIMO radar with colocated antennas," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 106–114, 2007.

[11] https://www.continental-automotive.com, May 2021.

[12] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, June 16-21, 2012.

[13] P. Sun and *et. al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, June 14-19, 2020.

[14] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A multimodal dataset for autonomous driving," *arXiv preprint arXiv:1903.11027*, 2019.

[15] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The Oxford radar robotcar dataset: A radar extension to the oxford robotcar dataset," in *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, Paris, France, May 31-Oct. 31, 2020.

[16] M. Meyer and G. Kuschk, "Automotive radar dataset for deep learning based 3D object detection," in *Proc. 16th European Radar Conference (EuRAD)*, Paris, France, Oct. 2-4, 2019.

[17] M. Sheeny, E. D. Pellegrin, M. Saptarshi, A. Ahrabian, S. Wang, and A. Wallace, "RADIATE: A radar dataset for automotive perception," *arXiv preprint arXiv:2010.09076*, 2020.

[18] Y. Wang, Z. Jiang, X. Gao, J.-N. Hwang, G. Xing, and H. Liu, "RODNet: Radar object detection using cross-modal supervision," in *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, Jan. 5-9, 2021.

[19] M. Mostajabi, C. M. Wang, D. Ranjan, and G. Hsyu, "High resolution radar dataset for semi-supervised learning of dynamic objects," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, June 14-19, 2020.

[20] A. Ouaknine, A. Newson, J. Rebut, F. Tupin, and P. Perez, "CARRADA dataset: Camera and automotive radar with range-angle-Doppler annotations," in *Proc. 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy, Jan. 10-15, 2021.

[21] O. Schumann, M. Hahn, N. Scheiner, F. Weishaupt, J. F. Tilly, J. Dickmann, and C. Wohler, "RadarScenes: A real-world radar point cloud data set for automotive applications," *arXiv preprint arXiv:2104.02493*, 2021.

[22] R. Zheng, S. Sun, D. Scharff, and T. Wu, "Beyond point clouds: A knowledge-aided high resolution imaging radar deep detector for autonomous driving," *arXiv preprint arXiv:2111.01246*, 2021.

[23] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, July 22-25, 2017, pp. 7263–7271.

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. International Conference on Learning Representations (ICLR)*, San Diego, CA, May 7-9, 2015.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, June 26-July 1, 2016, pp. 770–778.