Deep-Neural-Network-Enabled Vehicle Detection Using High-Resolution Automotive Radar Imaging

RUXIN ZHENG, Student Member, IEEE
SHUNQIAO SUN, Senior Member, IEEE
HONGSHAN LIU, Student Member, IEEE
The University of Alabama, Tuscaloosa, AL USA

TERESA WU, Senior Member, IEEE
Arizona State University, Tempe, AZ USA

Advanced driver assistance systems (ADASs) and autonomous vehicles rely on different types of sensors, such as camera, radar, ultrasonic, and LiDAR, to sense the surrounding environment. Compared with the other types of sensors, millimeter-wave automotive radar has advantages in terms of cost and reliability under bad weather conditions (e.g., snow, rain, and fog) and does not suffer from light condition variations (e.g., darkness). Typical radar devices used in today's commercial vehicles with ADAS features produce sparse point clouds in low angular resolution with a limited number of antennas. In this article, we present a machine-learning-aided signal processing chain to suppress the radar imaging blur effect introduced by the phase migration in time-division multiplexing multiple-input multiple-output radar, to generate low-level high-resolution radar bird's-eye view (BEV) spectra with rich object's features. Compared with radar point clouds, there is no information loss in radar BEV spectra. We then propose a temporal-fusion distance-tolerant single-stage object detection network, termed as *TDRadarNet*, and an enhanced version, *TDRadarNet+*, to robustly detect vehicles in both long and short ranges on radar BEVs. We introduce a first-of-its-kind multimodel dataset, containing 14 800 frames of high-resolution low-level radar BEV spectra with synchronized stereo camera RGB images and 3-D LiDAR point clouds. Our dataset achieves 0.39-m range resolution and 1.2° degree azimuth angular resolution with 100-m maximum detectable range. Moreover, we create a subdataset, the Doppler Unfolding dataset, containing 244 140 beam vectors extracted from the 3-D radar data cube. With extensive testing and evaluation, we demonstrate that our Doppler unfolding network achieves 93.46% Doppler unfolding accuracy. Compared to YOLOv7, a state-of-the-art image-based object detection network, TDRadarNet, achieves a 70.3% average precision (AP) for vehicle detection, demonstrating a 21.0% improvement; TDRadarNet+ achieves a 73.9% AP, showing a 24.6% improvement in performance.

Manuscript received 18 November 2022; revised 15 March 2023; accepted 7 May 2023. Date of publication 15 May 2023; date of current version 11 October 2023.

DOI. No. 10.1109/TAES.2023.3275887

Refereeing of this contribution was handled by J. Tabrikian.

This work was supported in part by the U. S. National Science Foundation under Grant CCF-2153386 and in part by the Alabama Transportation Institute. An earlier version of this paper was presented in part at the 2022 IEEE Annual Asilomar Conference on Signals, Systems, and Computers [DOI: 10.1109/IEEECONF56349.2022.10051978] and in part at the 2022 IEEE Sensor Array and Multichannel Signal Processing Workshop [DOI: 10.1109/SAM53842.2022.9827798].

Authors' addresses: Ruxin Zheng, Shunqiao Sun, and Hongshan Liu are with the Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL 35233 USA, E-mail: (rzheng9@crimson.ua.edu; hliu75@crimson.ua.edu; shunqiao.sun@ua.edu); Teresa Wu is with the School of Computing and Augmented Intelligence, ASU-Mayo Center for Innovative Imaging, Arizona State University, Tempe, AZ 85287 USA, E-mail: (teresa.wu@asu.edu). *(Corresponding author: Shunqiao Sun.)*

I. INTRODUCTION

Automotive radar sensors are crucial components in advanced driver assistance systems (ADASs) and autonomous vehicles due to their low cost, all-weather sensing capabilities, and immunity to poor visibility conditions [3], [4], [5], [6], [7]. Automotive radar systems typically rely on frequency-modulated continuous-wave (FMCW) signals in the millimeter-wave band, which enable cost-effective and high-resolution sensing for various autonomous driving functions such as automatic emergency braking, blind spot detection, and adaptive cruise control [8], [9], [10].

Object detection and classification are essential for autonomous driving. Humans sense the world through their eyes and ears and constantly use their brains to perform detection and classification tasks. Sensors, akin to human eyes and ears, allow vehicles to perceive their surroundings. Recently, many high-performance object detectors based on camera RGB images and LiDAR point clouds have been proposed [11], [12], [13], [14]. Although cameras allow us to better understand visual scenes, their performance is questionable in poor weather conditions [5]. LiDAR produces 3-D point clouds of the environment with high resolution on a good day by reflecting laser beams off surrounding objects [15], [16]. Its performance, however, degrades significantly in bad weather conditions. The average price of LiDAR products is also high.

Radar, on the other hand, is robust, inexpensive, and reliable even in harsh environments [5], [6], [7]. The wavelength of the millimeter-wave automotive radar operating at 76–81 GHz is in the millimeter range. The high bandwidth of a total of 4 GHz in a carrier frequency of 77–81 GHz is available for short- and medium-range automotive radars to achieve high range resolution. Due to the high carrier frequency, the form factor of automotive radar can be small so that it can be easily incorporated behind vehicle bumpers [5]. Compared with optical sensors, such as LiDAR and cameras, millimeter-wave automotive radar has strong penetration capabilities in fog, rain, snow, smoke, and dust [7]. However, the potential of object detection and classification using automotive radar has not been fully exploited. Today, most radar devices in commercial vehicles

with Level 2 features, such as adaptive cruise control function, have a relatively low angular resolution (around 10°) and low-end embedded computational unit [5], producing sparse point clouds, based on which object tracking is carried out. The Level 4 and Level 5 fully autonomous vehicles would require dense point clouds or radar imaging with a high angular resolution close to LiDAR [7]. Therefore, high-resolution automotive imaging radars [8], [17] are of great interest to support object detection and classification. Several commercial imaging radar products are available with various array configurations. For example, ZF imaging radar [18] has a detection range of up to 350 m and a field of view (FOV) of up to 120°. The ARS540 from Continental [19] has a detection range of up to 300 m and an FOV of up to 120°. However, both only provide radar point clouds as output.

Multiple-input multiple-output (MIMO) radar has been widely used to synthesize a large virtual array aperture for higher angular resolution, while keeping the number of transmit and receive antennas relatively small [5], [20], [21], [22]. MIMO radar relies on waveform orthogonality of the transmitted signals to enable separation at the receiver, which can be achieved using different methods such as time-division multiplexing (TDM), Doppler-division multiplexing (DDM), and frequency-division multiplexing [5]. Although TDM is a simple way to achieve waveform orthogonality, it reduces the maximum unambiguous detectable Doppler or radial velocity by a factor of the number of transmit antennas, an important property for imaging radar synthesized with a large number of transmit antennas [23]. Moreover, the switching delays among the transmit antennas from chirp to chirp introduce phase migration for moving objects, resulting in blurred radar imaging and distorted angular spectrum. To compensate for this, a correct velocity estimation of the moving object is required before carrying out angle estimation. However, moving targets with high speeds are estimated with ambiguity, and therefore, Doppler unfolding/dealiasing is required for accurate velocity estimation of moving objects with high speed under TDM MIMO radar. Various approaches have been proposed to address the challenge of unfolding ambiguous velocities, including methods such as different pulse repetition intervals (PRIs) in conjunction with the Chinese remainder theorem (CRT) [2], overlapped virtual arrays [24], [25], and track-based techniques [26], [27]. These methods often require high operating complexity or additional hardware cost.

High-resolution imaging radar outputs are usually represented as point clouds. Networks originally developed for LiDAR point cloud classification and segmentation, such as PointPillars [28], VoxelNet [29], PointNet [30], and PointNet++ [31], can be directly used or adjusted for radar point clouds [32], [33]. Unfortunately, the thresholding algorithms in generating radar point clouds from radar low-level data often lead to significant information loss. To address this challenge, researchers develop novel deep neural networks [34], [35], [36], [37] on high-resolution radar low-level data representation (e.g., range–azimuth spectra) as it contains rich information of the object (e.g., shape) to support fully autonomous driving.

In this article, we propose a high-resolution radar object detection system that can robustly detect vehicles up to 100 m under various driving scenarios. Our system contains a novel machine-learning-aided TDM MIMO radar signal processing pipeline for generating undistorted high-resolution radar imaging, and a temporal-fusion distance-tolerant radar single-stage object detection network, termed as *TDRadarNet*, for vehicle detection in far and near fields. Our novel signal processing pipeline overcomes the TDM MIMO radar phase migration problem by using a deep-learning-based Doppler unfolding network. After applying the signal processing pipeline, a high-resolution radar range–azimuth spectrum is created in polar coordinates. We transfer the radar range–azimuth spectrum into a radar bird's-eye view (BEV) in Cartesian coordinates using linear interpolation. Considering that the radar BEVs are in general hardly shift invariant over both angle and range, we propose a novel TDRadarNet by separating the radar BEV into far and near fields for object detection. In this article, we focus on recognizing targets not only in the near field with rich shape information but also in the far field with lower resolution. In addition, the proposed network exploits temporal features from historical radar frames for detection. To the best of our knowledge, this research also introduces a first-of-its-kind multimodel dataset, which contains high-resolution radar spectra with synchronized stereo camera RGB images and 3-D LiDAR point clouds under different driving scenarios. Unlike existing datasets with low-resolution or short maximum detectable range, our radar dataset achieves 0.39-m range resolution and 1.2° azimuth angular resolution with 100-m maximum detectable range. The dataset includes 14 800 radar BEV frames representing 30-min driving in different driving scenarios, such as on campuses, highways, and urban streets. We also create a subdataset for TDM MIMO radar Doppler unfolding, which consists of 244 140 beam vectors with varying signal-to-noise ratios (SNRs) and target directions of arrival (DOAs). To validate our proposed system, we train and test the Doppler unfolding network of the signal process pipeline using a subdataset. The Doppler unfolding network achieves 93.46% accuracy. We further train and evaluate our TDRadarNet using our self-collected dataset to demonstrate the outperformance of our proposed system over state-of-the-art image-based object detection networks with 10.6%, 17.1%, and 14.1% improvements in precision, recall, and F1-score, respectively. Overall, our contributions are as follows.

1) A novel signal processing pipeline with a deep learning network for TDM MIMO radar to overcome phase migration problems is developed.
2) With field experiments, a multimodel sensor dataset that contains high-resolution radar BEVs with synchronized stereo cameras RGB images and 3-D LiDAR point clouds is generated. In addition, a subdataset containing beam vectors with different SNRs

and target DOAs for Doppler unfolding network training is created.
3) A novel deep neural network, TDRadarNet, is proposed. TDRadarNet leverages historical radar frames to exploit temporal features and separates far and near fields to address inconsistent resolution in radar BEV frames, demonstrating the promise of robust radar object detection in various driving scenarios.
4) An enhanced version of the neural network, TDRadarNet+, is proposed to utilize Doppler information for object detection. TDRadarNet+ demonstrates the potential of robust radar object detection by effectively utilizing Doppler information.

Existing research has explored the application of machine learning and deep learning on Doppler radar data. For example, Zheng et al. [1] presented machine-learning-aided Doppler unfolding with preliminary simulation results. Object detection based on high-resolution radar range–azimuth spectra obtained at a few locations using a vanilla-image-based neural network was presented in [2]. We argue that our contributions go beyond [1] and [2]. Here, we propose a novel temporal-fusion and distance-resilient neural network for vehicle detection. We develop a high-resolution low-level automotive radar imaging dataset with a maximum detection range of 100 m under different driving scenarios. In addition, comprehensive numerical studies and verification have been conducted based on field experiment results.

The rest of this article is organized as follows. In Section II, we review existing approaches to solve the motion-induced phase errors in TDM MIMO radars, the publicly available radar datasets for autonomous driving, and different deep learning approaches for radar object detection using different radar data representations. The system model of automotive FMCW MIMO radar and its challenges for TDM MIMO radar are addressed in Section III. In Section IV, we propose a deep-learning-aided signal processing pipeline to generate high-resolution radar imaging by unfolding the ambiguous Doppler estimations. The details of the proposed TDRadarNet are presented in Section V. Our own dataset is introduced in Section VI, and the neural network experimental performance on our dataset is shown in Section VII. Finally, Section VIII concludes this article.

II. RELATED WORKS

In this section, we briefly discuss the existing Doppler unfolding approaches in automotive TDM MIMO radars, automotive radar data representation and datasets for autonomous driving, and relevant radar machine learning work.

A. Doppler Unfolding Approaches

Under TDM MIMO radar, only one transmitter is turned ON at each time slot. The switching delay between transmitters introduces a phase migration between chirp and chirp. As a result, it would corrupt the angular phase information of moving targets leading to blurred radar imaging. Unlike cameras overcoming the blurred image issue by increasing frames per second (FPS), TDM MIMO radars require special antenna geometries or adaptive signal processing techniques to tackle the phase migration.

1) *Overlapping Virtual Array Elements:* Phase error is estimated using overlapping elements in the virtual aperture [24]. If two virtual elements corresponding to different transmitters share the same virtual position, the only phase difference between them in an ideal situation is the motion-induced phase error. The issue of this approach is that it requires redundancy of virtual elements and, thus, increases hardware costs. In addition, this approach suffers from low-SNR situations.

2) *Adaptive Discrete Fourier Transform (DFT):* Bechter et al. [38] concluded that overlapping element is unnecessary and the motion compensation can be resolved via applying adaptive DFT on signals corresponding to different transmitters. However, such an approach may cause ambiguities in the Doppler domain, and a special waveform design is required to carry out Doppler dealiasing.

3) *Staggered TDM With CRT:* Several proposed studies use multiple pulse repetition frequencies (PRFs) techniques combined with the CRT to address the Doppler ambiguities [2], [27]. The general approach involves searching for coincidences between the unfolded estimates obtained for each PRF. However, such an approach is sensitive to noise, and the pairing process has a high computational cost.

4) *Track-Based Approach:* Examples of track-based approaches are multiple hypothesis tracking (MHT) [26] and the track-before-resolve (TBR) [27] method. MHT uses multiple PRFs and initiates multiple track hypotheses to represent potential target positions, while TBR resolves the ambiguity using the tracking filters of multiple models. Both approaches have shown great promise in resolving Doppler ambiguities at low SNRs.

B. Radar Datasets for Autonomous Driving

Environmental perception plays a key role in autonomous driving. The popular datasets in autonomous vehicle perception, such as KITTI [39] and Waymo Open Dataset [12], only contain cameras and LiDAR recordings. As radar gains more attention in autonomous driving, several datasets containing radar data have been released recently, such as nuScenes [40], Oxford Radar RobotCar [41], Astyx [42], RADIATE [43], CRUW [34], Zendar [44], CARRADA [45], RadarScenes [46], RADIal [36], View-of-Delft [47], and Radatron [37]. The publicly available datasets are summarized and compared in Table I. CARRADA and CRUW datasets use single-chip Texas Instruments (TI) radar with a low angular resolution, i.e., larger than 10°. RADIATE and Oxford Radar RobotCar use mechanical scanning radars, CTS350-X, providing high-resolution radar images with 4 FPS. Yet, the Doppler information of targets is missing, and such a low frame rate introduces uncorrectable motion blur to the radar imaging. Synthetic aperture radar (SAR) technology for static targets

TABLE I
Overview of Publicly Available Radar Datasets

Dataset	# of Frames	Data Type	LiDAR	Camera	Radar		Radar/Technology
					Range	Resolution	
nuScenes [40]	40,000	Sparse PC	√	√	LR	Low	Continental ARS408
Oxford Radar [41]	240,000	RA	√	√	MR	High	Navtech Spinning Radar
Astyx [42]	500	PC	√	√	MR	High	Astyx 6455
RADIATE [43]	44,000	RA	√	√	MR	High	Navtech Spinning Radar
CRUW [34]	396,241	RA	×	√	USR	Low	TI AWR1843
Zendar [44]	94,460	ADC,RD,PC	√	√	MR	High	SAR
CARRADA [45]	12,666	RA,RD,RAD	×	√	SR	Low	TI AWR1843
RadarScenes [46]	975	Dense PC	×	√	MR	High	77GHz Middle-Range Radar
RADIal [36]	25,000	ADC,RD,PC	√	√	MR	High	Valeo Middel Range DDM
View-of-Delft [47]	8,693	PC+Doppler	√	√	LR	High	ZF FRGen21 Radar
Radatron [37]	152,000	RA	×	√	USR	High	TI Cascade Imaging Radar
BAMA (Ours)	14,800	RA+Doppler	√	√	MR	High	TI Cascade Imaging Radar

Data type: Raw ADC data (ADC), range–Doppler map (RD), range–azimuth map (RA), and point clouds (PC). Range: LR, MR, SR, and USR stands for long range (> 200 m), medium range (< 200 m), short range (< 50 m), and ultrashort range (< 25 m) [48].

is adopted in the Zendar dataset with multiple measurements from different vehicle locations. The Astyx dataset contains only 500 frames with sparse radar point clouds. High-definition imaging radar with DDM MIMO configuration is adopted in RADIal, resulting in interleaved range–Doppler maps. The View-of-Delft dataset utilizes the ZF FRGen21 radar [18], a long-range and high-resolution imaging radar. It provides only point cloud data, and object annotations are available for objects within a short range of 50 m. Radadtron uses TI cascade imaging radar, similar to ours. It only contains targets in near fields, i.e., within 25 m.

C. Radar-Based Deep Learning

Deep learning has found wide application in radar systems [49], [50]. For example, low-cost radar, such as Soli radar [51], is used to capture hand gesture for human–computer interaction. Short-range radar is also proposed in the medical field to remotely monitor human vital signs [52]. Radar has a long application history in commercial automobiles [6] since the 1990s, spanning from ADAS to the recently emerging autonomous driving techniques [5]. Different automotive radar data representations have been exploited, which can be, in general, divided into three categories.

1) *Radar Point Clouds:* Radar data can be represented as point clouds by applying filtering and thresholding algorithms, such as constant false alarm rate, on the radar range–azimuth map. In this way, radar produces sparse point clouds, and it can be viewed as a low-quality LiDAR. Point-cloud-based networks, such as PointPillars [28], VoxelNet [29], and PointNets [30], can be directly used or adjusted [32] for radar point clouds. Moreover, Scheiner et al. [53] present a comparison between five real-time capable object detector architectures on radar point clouds. Such thresholding algorithms in generating radar point clouds may lead to significant information loss of objects.

2) *Radar Data Tensor:* To avoid loss of information, radar data can be processed in 3-D tensors, i.e., range–Doppler–azimuth for 1-D antenna array, or 4-D tensors, i.e., range–Doppler–azimuth–elevation for 2-D antenna array. Deep-learning-based radar detector [50] directly learns from 4-D complex radar tensors for object detection and localization. It is also possible to project the 3-D radar tensors along different views to extract 2-D features for semantic segmentation [54] and object recognition [55], [56], [57].

3) *Radar BEVs:* Radar BEVs were generated from a radar range–azimuth map through coordinate transformation. Radar BEVs obtained from high-resolution radar contain targets' geometric information. Object detection based on radar BEV was proposed in [2], [37], and [58], achieving relatively accurate object detection. However, only highway scene is considered in [58], which are considered as the clean and easy scenario in autonomous driving. In [2], the radar is placed at intersections and only moving targets in the near fields are of interest. Similarly, objects within ultra short range are considered in [37].

In radar machine learning, taking advantage of temporal and spatial information can effectively improve detector performance [48]. Extensive studies have been conducted on the combination of different radar frames, such as summation among neighboring frames [54], concatenation in frame level [59], and stacking in feature level [34], [55]. In [55], a convolutional long short-term memory layer is adopted after the encoder network to extract temporal features from a sequence of feature maps. In [34], frame-level feature maps are concatenated and temporal features are extracted by a 3-D convolutional neural network (CNN) layer. Other than using a CNN-based network, an isotropic graph convolution network that leverages spatial information from neighboring nodes is proposed in [60] to boost radar detection performance.

III. SYSTEM MODEL OF AUTOMOTIVE MIMO RADAR

FMCW radar is a widely used technology in automotive applications due to its ability to provide high-range resolution at a low cost. In this section, we will explain the principles of FMCW radar and how it is used in automotive

MIMO radar systems to synthesize a large virtual array followed by the discussion of challenges that arise with MIMO radar technology.

A. FMCW Radar

The FMCW radar signal is generated by transmitting a continuous-wave signal that linearly increases in frequency over time, with a carrier frequency f_c, bandwidth B, and chirp duration T. The transmitted signal's phase $\varphi_T(t)$ is obtained by integrating the transmitted frequency $f_T(t)$ over time, yielding $\varphi_T(t) = 2\pi \int_{-T/2}^{t} f_T(t)dt$. When the transmitted signal encounters a target at a range R with radial velocity v, the signal is delayed by a round-trip time of $\tau = 2(R + vt)/c$, where c is the speed of light. The received signal is mixed with the transmitted signal, resulting in a beat signal with a phase $\varphi_B(t)$ that can be approximated as

$$\varphi_B(t) = 2\pi \left[\frac{2f_c R}{c} + \left(\frac{2f_c v}{c} + \frac{2BR}{Tc} \right) t \right] \tag{1}$$

where the beat frequency is given by $f_b = f_R + f_D$, with $f_R = 2BR/Tc$ representing the range frequency and $f_D = 2f_c v/c$ representing the Doppler frequency. To improve the radar's dynamic range, the beat signal typically goes through a bandpass filter that compensates for the gain in targets at different distances, followed by an analog-to-digital converter (ADC) that samples the signal at a rate greater than twice the maximum beat frequency $f_b^{\max}$. Range and Doppler information of the target can then be obtained by applying fast Fourier transforms (FFTs) along the fast-time and slow-time dimensions.

B. Automotive MIMO Radar and Waveform Orthogonality

MIMO radar has been increasingly used in automotive radar design due to its ability to synthesize a large virtual array for angle estimation using a small number of transmit and receive antennas [5], [20]. An example of a MIMO radar with $N_t = 2$ transmit and $N_r = 4$ receive antennas is shown in Fig. 1(a), where a virtual uniform linear array with eight elements is synthesized. By applying digital beamforming [61] or super-resolution algorithms, such as MUSIC [62], ESPRIT [63], and compressive sensing [64], the DOA is estimated.

In automotive MIMO radar, it is crucial to transmit orthogonal FMCW sequences so that at the receiving side, the contribution of each transmit antenna can be extracted from the receive signals. There are different ways to achieve waveform orthogonality in MIMO radar, such as DDM and TDM [5].

In the DDM scheme, waveform orthogonality is achieved in the slow-time domain by multiplying a phase code on each transmitted FMCW chirp. At the receive side, the contribution of each transmitter can be shifted to a higher Doppler frequency or treated as random noise by applying slow-time Doppler demodulation after range FFT.

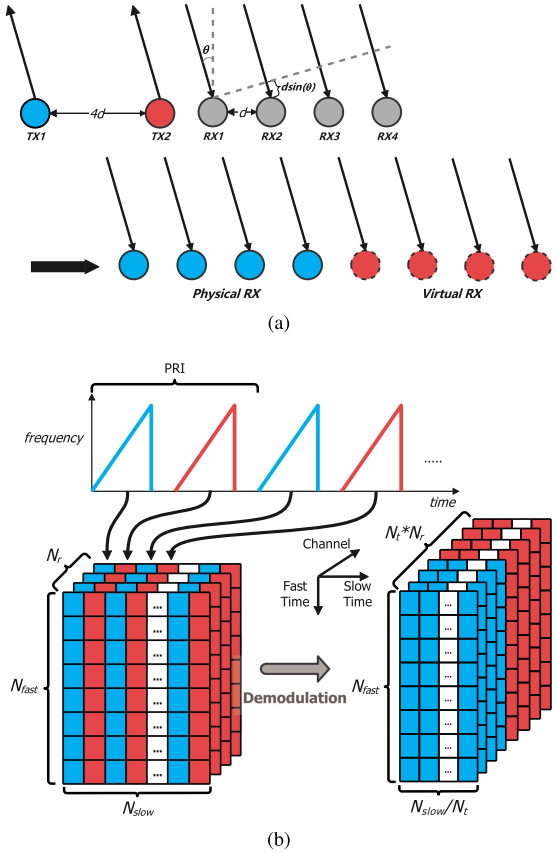


Fig. 1. Example of MIMO radar with eight virtual array elements, synthesized by two transmit and four receive antenna elements via TDM. (a) Physical and virtual arrays. (b) Illustration of waveform orthogonality through TDM.

DDM allows all transmit antennas to transmit simultaneously. However, it either reduces the maximum unambiguous detectable Doppler or masks objects with low radar cross section by the waveform residual from other transmit antennas [5].

On the other hand, the TDM scheme selects only one transmit antenna to transmit at each time. A signal processing example of a TDM MIMO radar with $N_t = 2$ transmit antennas and $N_r = 4$ receive antennas is shown in Fig. 1(b). Assume that N_{slow} chirps are transmitted in one coherent processing interval and the number of ADC samples in one chirp is N_{fast}. All odd chirps (blue) are transmitted by TX1, and all even chirps (red) are transmitted by TX2. The radar data matrix can be assembled at each receive antenna into two matrices corresponding to odd and even chirp sequences, respectively. Therefore, a radar data cube with a dimension of $(N_{\text{slow}}/N_t) \times N_{\text{fast}} \times (N_t N_r)$ can be obtained from the original $N_{\text{slow}} \times N_{\text{fast}} \times N_r$ data cube. Due to its simplicity in implementation, we adopt TDM to achieve waveform orthogonality in this article.

C. Challenges of TDM MIMO Radar

1) *Doppler Aliasing:* In automotive radar, the maximum unambiguous detectable velocity is given by

$$v_{\max} = c/(4f_c T_{\text{PRI}}) \tag{2}$$

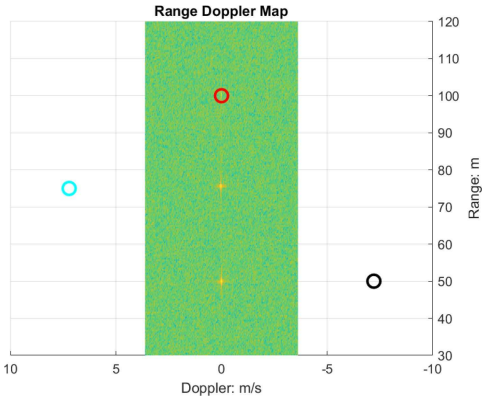


Fig. 2. Illustration of Doppler folding/aliasing. The FMCW radar simulator is configurated to have a maximum unambiguous detectable radial speed, $v_{\max} = 3.6$ m/s. Range–Doppler map with three targets at ranges of 100, 75, and 50 m and corresponding velocities of 0, $2v_{\max}$, and $-2v_{\max}$ m/s. Circle markers show the targets' ground truth parameters.

where c is the speed of light, f_c is the carrier frequency, and T_{PRI} is the PRI. However, in TDM MIMO radar, where only one transmit antenna is scheduled to transmit at each time slot, T_{PRI} of each transmit antenna is enlarged by the number of transmit antennas N_t. This increase in T_{PRI} reduces $v_{\max}$ by N_t times, meaning that the TDM MIMO radar can only estimate velocities without ambiguity within a smaller range compared to a traditional radar with the same PRI but only one transmit antenna.

This reduction in $v_{\max}$ can exacerbate the problem of Doppler aliasing, which occurs when targets move with velocities beyond the maximum unambiguous detectable velocity. In Doppler aliasing, targets appear on the range–Doppler map with incorrect Doppler estimates, and this will fold back targets at incorrect locations within the range–Doppler map. We illustrate this problem in Fig. 2, where an FMCW radar simulator is used to generate radar data for three targets with different range and Doppler information. Although those three targets have different velocities, they share the same Doppler/velocity estimation on the range–Doppler map due to Doppler folding. In this case, the velocity information of the object cannot be estimated accurately, and an object moving with an aliasing velocity, such as $2v_{\max}$ or $-2v_{\max}$, could be detected as stationary.

2) *Phase Migration:* The scheduling delay, Δt, between different transmit antennas can result in phase migration for moving targets between different chirps, defined as follows:

$$\phi = (4\pi/\lambda) v \Delta t \tag{3}$$

where $\lambda = c/f_c$ is the wavelength. For the mth switched transmit antenna, the phase migration is written as [65]

$$\phi_m = (m-1)\pi \frac{v}{v_{\max}}. \tag{4}$$

This phenomenon will cause a distortion in the virtual array beampattern, leading to inaccurate angle finding. For a MIMO radar with N_t transmit and N_r receive antennas, the transmit and receive steering vectors are

$$\mathbf{a}_t(\theta_t) = \left[e^{j\frac{2\pi}{\lambda}d_t^1 \sin(\theta_t)}, \ldots, e^{j\frac{2\pi}{\lambda}d_t^{N_t} \sin(\theta_t)}\right]^T \tag{5}$$

$$\mathbf{a}_r(\theta_r) = \left[e^{j\frac{2\pi}{\lambda}d_r^1 \sin(\theta_r)}, \ldots, e^{j\frac{2\pi}{\lambda}d_r^{N_r} \sin(\theta_r)}\right]^T \tag{6}$$

where d_t^m denotes the distance of the mth transmit antenna to the reference transmit antenna; similarly, d_r^n denotes the nth receive antenna to the reference receive antenna. For targets in far field, it holds that $\theta_t = \theta_r \triangleq \theta$. The virtual array steering vector is

$$\mathbf{a}(\theta) = \mathbf{a}_t(\theta) \otimes \mathbf{a}_r(\theta) \tag{7}$$

where $\otimes$ is the Kronecker product. The element of the virtual array steering vector corresponding to the mth transmit and nth receive antenna is denoted as $e^{j\varphi_{mn}}$, where

$$\varphi_{mn} = \frac{2\pi}{\lambda}\left(d_t^m + d_r^n\right)\sin(\theta). \tag{8}$$

As a result, the element of the virtual array steering vector corresponding to the mth transmit and nth receive antenna is $e^{j(\varphi_{mn}+\phi_m)}$. The term ϕ_m needs to be compensated to avoid beampattern distortion.

To illustrate this effect, we performed simulations using an array configuration shown in Fig. 8, which is similar to the one used in TI imaging radar [25]. Assume that there is a moving target at 10° with a velocity of 20 m/s. Fig. 3(a) shows the angle spectrum obtained from the virtual array without velocity compensation. To remove phase migration, a compensation value of $e^{-j\phi_m}$ needs to be multiplied along the virtual array before angle finding for every moving target. The velocity of the target is estimated from the 2-D FFT result of a single RX antenna data matrix. Fig. 3(b) shows the correct angle spectra after compensation. In addition, Fig. 3(d) and (e) demonstrates the radar BEVs of highway guardrails before and after motion compensation. Without compensation, the radar BEV is severely blurred due to the phase errors induced by the moving host vehicle. The compensated radar BEV matches well with the camera image and LiDAR BEV of the guardrail shown in Fig. 3(c) and (f), respectively, which serves as the ground truth.

IV. DEEP-LEARNING-AIDED SIGNAL PROCESSING PIPELINE OF HIGH-RESOLUTION IMAGING RADAR

In this section, we introduce a novel signal processing pipeline that leverages the power of deep learning to synthesize a virtual uniform linear antenna array with 86 elements for TDM MIMO radar. As a result, high-resolution radar imaging with an azimuth resolution of 1.2° is achieved by cascading four AWR2243 chipsets [25].

The proposed deep-learning-aided TDM MIMO radar signal processing pipeline is illustrated in Fig. 4. To synthesize MIMO virtual array, a waveform decoding process is applied to the 3-D radar data cube at the radar receive side. This will separate the transmit signals from different transmit antennas and rearranges the data cube to have a large virtual array.

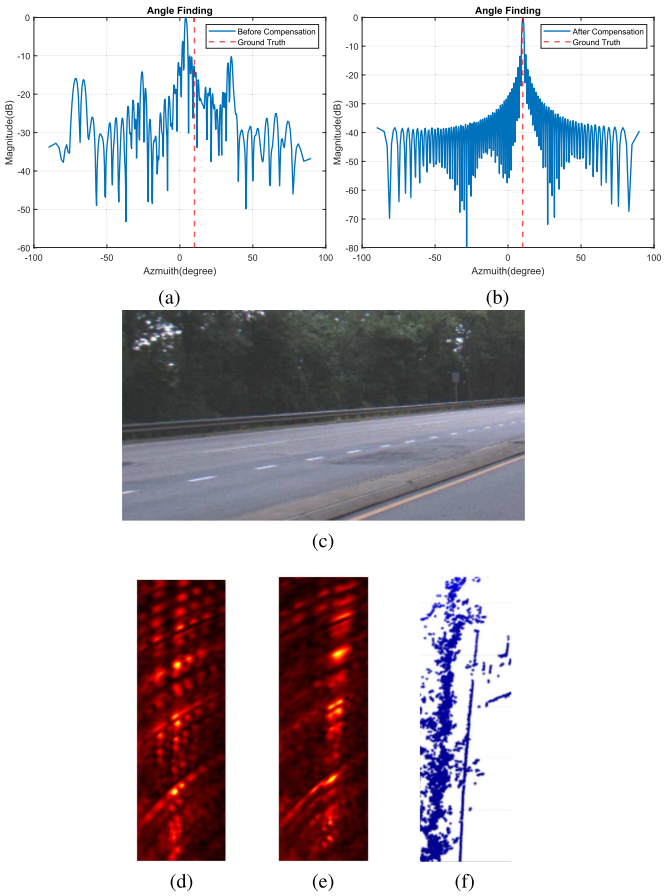


Fig. 3. Angle spectra of a moving target with a velocity of 20 m/s and azimuth angle of 10°: (a) before and (b) after phase compensation. The radar is configured to select 9 TX and 16 RX antennas with a chirp duration of 50 μs. Guardrail from (c) camera image and (d) radar BEV before, (e) after phase compensation, and (f) LiDAR BEV.

In practice, there may exist frequency, phase, and amplitude mismatches across the four radar chipsets. These mismatches can be caused by various factors, such as path length mismatch, chip-to-chip variation, and antenna coupling. To address this issue, a corner reflector at the boresight direction is utilized to generate a frequency and amplitude calibration matrix to compensate for the different transmit–receive antenna path delays.

To estimate the target range and Doppler information, the dc offset is first removed and an N_{fast}-point Hanning window is applied to the 3-D radar data cube along the fast time axis before performing an FFT to estimate the targets' range.

Next, another N_{slow}/N_t-point Hanning window is applied to the data cube along the slow time axis, followed by a second FFT to estimate the targets' Doppler. The Hanning window reduces the effects of sidelobes in the frequency domain, which can distort the range or Doppler estimation. Overall, the combination of dc offset removal and Hanning windowing helps to improve the accuracy and precision of the range and Doppler estimation in the signal processing chain.

To account for the motion-induced phase error, a Doppler unfolding and phase compensation step is necessary. In practical applications such as autonomous driving, the range of velocities is typically limited. For instance, the velocity range is typically within $[-120, 120]$ mi/h. This implies that the number of possible velocity values is finite. Thus, when the maximum velocity, $v_{\max}$, is sufficiently large, it is possible to unfold the velocity a limited number of times to estimate the velocity correctly. In our case, only nine possible unfolded velocity candidates are considered, i.e.,

$$\mathcal{S} = \{v - 4 \times (2v_{\max}), \ldots, v, \ldots, v + 4 \times (2v_{\max})\}. \quad (9)$$

As shown in Fig. 4, we use a 1-D CNN, comprising three CNN layers and one fully connected layer, to estimate Doppler in automotive TDM MIMO radar. The CNN is pretrained with simulated data and fine-tuned on real data with the same antenna configuration as the TI imaging radar. The simulated data consist of beam vectors of virtual arrays obtained after range and Doppler FFTs. The dataset includes nine velocity candidates, with labels 0–8. The beam vector of the virtual array is selected from the 3-D radar data cube along the channel dimension, and a signal preprocessing step is applied to remove the angle-dependent phase in the beam vector to ensure that the input data for the Doppler unfolding network only contain motion-induced phase error, mutual coupling error, and noise. This preprocessing step significantly reduces the amount of data required for training. Next, phase features in the beam vector are extracted using a 1-D CNN. The network output is the estimated velocity used to compensate for the phase migration in the beam vector of the virtual array.

It is important to note that mutual coupling and fabrication imperfections may lead to degraded antenna performance, highlighting the need for antenna array calibration in producing high-quality radar imaging [66], [67]. While the TI cascaded imaging radar employs a simple one-time boresight calibration method for improved angle performance, this approach suffers from reduced performance when targets are off-boresight. To address this, we adopt an angle-dependent calibration method that uses radar measurements of a corner reflector at various angles to calibrate the already compensated virtual array beam vector.

Next, an angle FFT is conducted along the already compensated and calibrated virtual array with a Chebyshev window with 50-dB sidelobe attenuation, generating a 3-D radar spectra that contains range, Doppler, and azimuth information. A range–azimuth heatmap is obtained by collapsing the 3-D data spectra along the Doppler dimension. Finally, the polar coordinates (r, θ) of the range–azimuth map are converted to Cartesian coordinates (X, Y), generating a radar BEV. This representation provides a 2-D view of the surrounding environment, with the X-axis representing the horizontal distance and the Y-axis representing the vertical distance from the radar sensor.

An example of the radar range–azimuth spectra of a street intersection and its corresponding BEV generated through polar-to-Cartesian transformation are presented in

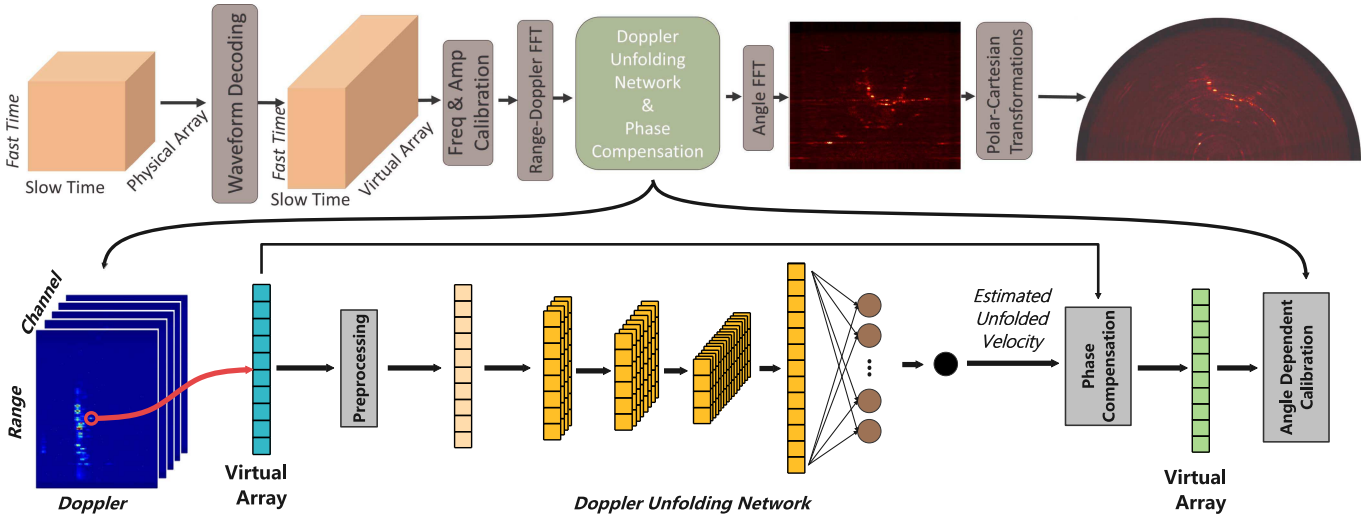


Fig. 4. Illustration of deep-learning-aided TDM MIMO radar signal processing pipeline.

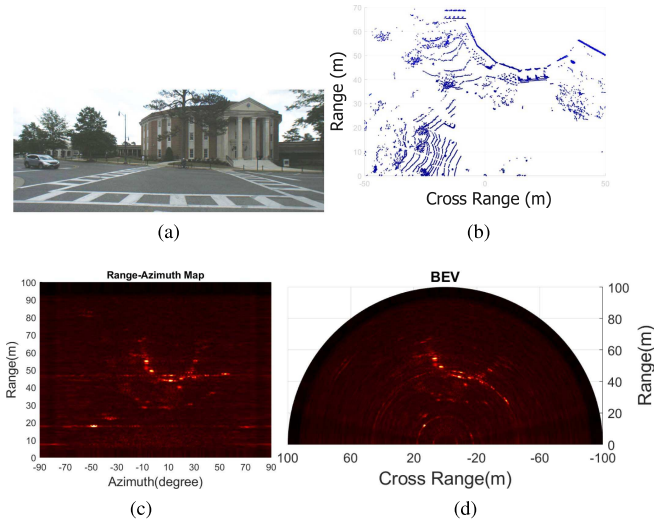


Fig. 5. (a) Camera. (b) LiDAR. (c) Polar coordinate and its (d) Cartesian coordinate transformation.

Fig. 5. The high-resolution radar BEVs provide rich shape information of various objects such as buildings, cars, trees, and light poles, as observed in Fig. 5, as well as in Figs. 13 and 15. The shape information enables accurate target detection using the proposed TDRadarNet (see Section V).

V. TDRADARNET FOR RADAR OBJECT DETECTION

Radar BEVs are single-channel grayscale images with varying resolutions and SNR. Considering these unique characteristics, we propose a new object detection network called temporal-fusion distance-tolerant radar object detection network (TDRadarNet). TDRadarNet utilizes temporal features and handles objects in far and near fields differently to improve detection performance on radar BEVs.

A. Not Every Pixel Is Created Equally

To enable object detection via deep learning, high-resolution radar BEVs in Cartesian coordinate are utilized, as they contain rich shape information of objects. However, it is important to note that not every pixel in radar BEV is generated with the same accuracy. The accuracy of the pixels is generally sensitive to both range and angle. This will have implications on the performance of any detection algorithm utilizing radar BEV, as small measurement errors can result in large variances in the Cartesian coordinate for targets in the far field. This issue will be discussed in more detail in the following sections.

1) *Effective Antenna Aperture Relies on Angle:* In a uniform linear array, the half power beamwidth [68] is given by $\theta_B \approx \frac{0.886\lambda}{Nd\cos\theta}$, where θ is the angle of view, N is the number of antenna elements, d is the antenna spacing, and λ is the wavelength. The maximum effective antenna aperture, and hence the best angular resolution, is achieved along the boresight direction. The effective aperture of the array decreases as the view angle increases, resulting in poorer angular resolution.

2) *SNR Drops as Range Increases:* According to the radar range equation, the received power decreases as the range increases. The radar receive power of a target of range r with radar cross section of σ is [69]

$$P_r = \frac{P_t C \sigma}{(4\pi)^3 r^4} \tag{10}$$

where C can be considered as a constant number for the same radar, which includes antenna gain, effective antenna area, and efficiency, and P_t is the transmit power. Therefore, $P_r \propto 1/r^4$. Typically, targets at a far distance have lower SNR, as a result of which the angle estimation error is relatively large for targets with long ranges.

3) *Information Loss in Coordinate Systems Transform:* The radar range–azimuth spectrum is typically obtained in polar coordinates, which are then transformed or interpolated into a Cartesian coordinate system before fed into

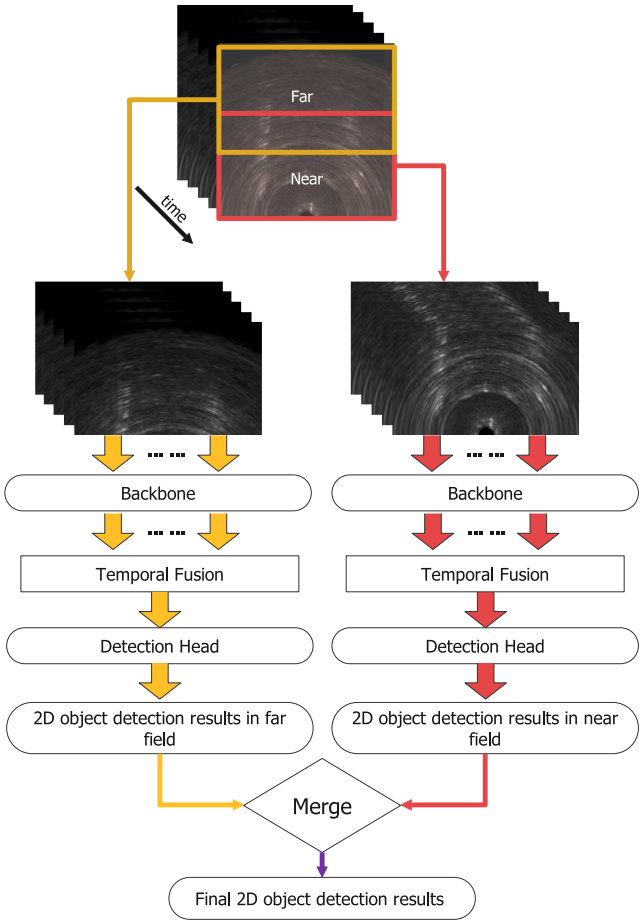


Fig. 6. TDRadarNet. Far-near fields are divided and used to train two sets of learnable parameters of the model. Temporal fusion works by extracting temporal features from a sequence of frames.

deep neural networks. As the range increases, the distance between adjacent bins becomes larger, resulting in a larger variance in the distance between adjacent pixels in Cartesian coordinate systems for targets at longer ranges. A small error in angle estimation will cause a large variance in the radar BEV pixels in far fields. And the SNR decreases at longer ranges, leading to relatively large DOA estimation errors, which further amplify the variance in these pixels. This issue is particularly significant for radar BEVs, where inconsistencies in pixel resolution and SNR are already present.

To summarize, the radar BEV image obtained using MIMO radar with fixed antenna arrays exhibits shift variance over both angles and ranges, making it inherently difficult to be shift invariant.

B. Network Architecture of TDRadarNet

Given the unique characteristics of radar BEVs, we introduce TDRadarNet (see Fig. 6). This network comprises two identical subnetworks trained for object detection in the far and near fields, respectively. The input radar frame sequences are partitioned into overlapping sequences of far and near fields. A backbone network extracts features from each frame, followed by a temporal fusion stage to explore

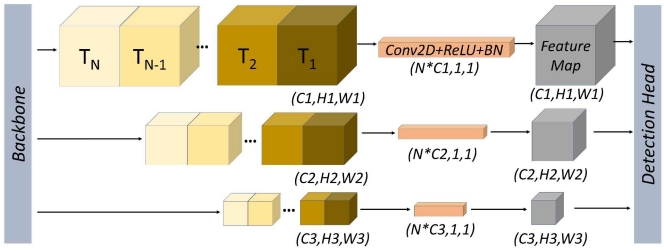


Fig. 7. Temporal fusion module. Feature maps extracted from N consecutive frames are concatenated in three scales separately. The temporal features are then extracted by applying a convolution kernel followed by batch normalization and ReLU activation.

the historical relationship between the features across consecutive frames. The detection head produces predictions, which are combined for the far and near fields. Inspired by You Only Look Once (YOLO) v7 [70], TDRadarNet is optimized for radar BEVs by learning distinct features for far and near fields and integrating temporal information from historical radar frames.

The details of TDRadarNet are outlined as follows.

1) *Far and Near Fields:* Detecting objects in radar BEVs can be challenging due to varying reflection intensity, shape, and contrast caused by differences in resolution. This dissimilarity is particularly notable when attempting to detect the same type of object, such as cars, in both near and far fields. To address this issue, we divided the radar BEV frame into two regions: far field and near field, as shown in Fig. 6. These regions overlap to ensure that no information is lost at the boundaries. A deep learning model on each region is trained to learn two sets of parameters optimized for detecting objects in far and near fields. We decide to adopt a separate processing approach for far and near areas for multiple reasons. First, translating polar to Cartesian before feeding it into the object detection network allows us to leverage LiDAR and camera images for extrinsic calibration. Second, separately processing the far and near areas helps to handle error amplification in coordinate translation as well as SNR deviations due to path loss. Third, it enables the network to learn field-related specific features to improve the vehicle detection performance in far field that experiences lower resolution and SNR drop.

2) *Temporal Fusion:* While object detection can be performed with a single radar frame, we contend incorporating temporal information from multiple consecutive frames can improve detection accuracy. As shown in Fig. 7, a sequence of N frames is processed by the backbone to extract feature maps of three scales for each frame, resulting in a total of $3 \times N$ feature maps. To incorporate temporal information, the feature maps that correspond to the same scale are concatenated and passed through a convolutional layer, followed by batch normalization and rectified linear unit (ReLU) activation, to extract temporal features.

3) *Merging:* Once the TDRadarNet generates object detection results for the far and near fields, the results are merged into a single frame. In cases where there are overlapping detections in the common region of both fields, we use nonmaximum suppression to filter out duplicate

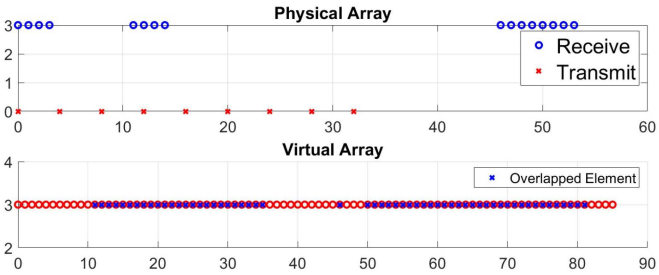


Fig. 8. TI imaging radar. Four AWR2243 radar transceivers are cascaded together, providing 9 transmit and 16 receive antennas in horizontal direction, enabling the synthesis of 86 unique virtual array elements with half-wavelength spacing. Note 58 virtual array elements are overlapped.

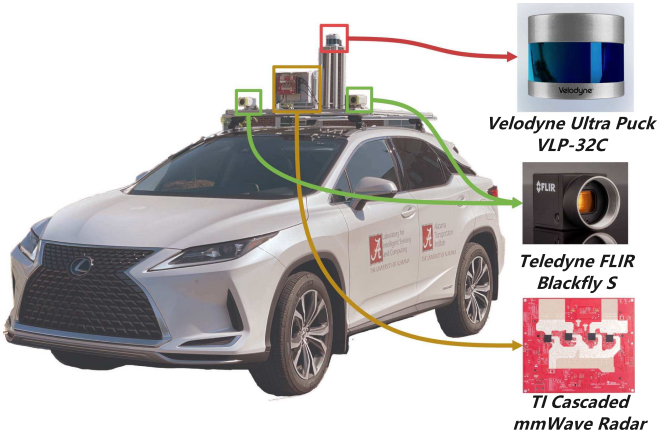


Fig. 9. Data acquisition vehicle platform of Lexus RX450h with high-resolution imaging radar, LiDAR, and stereo cameras is used to carry out field experiments at the University of Alabama.

detections. Specifically, we set an intersection over union (IoU) threshold to determine the extent of the overlap. If multiple detections share the same intersection, the detection with the highest confidence score is kept.

C. TDRadarNet+ Exploiting Doppler Information

Radar BEVs are grayscale images that represent reflection intensity with valuable Doppler information such as the velocity of objects in the scene relative to the radar sensor. To make use of the Doppler information, we create a radar Doppler BEV. This involves creating a Doppler feature map by replacing the pixel values of the range–azimuth heatmap with corresponding Doppler information obtained through our Doppler unfolding network. A polar–Cartesian transformation to both the range–azimuth and Doppler feature maps is employed to ensure consistency. As a result, the radar Doppler BEV has the same dimensions as the radar BEV, with the only difference being that the pixel values represent Doppler information rather than reflection intensity. We propose an improved version called TDRadarNet+ that leverages the Doppler information provided by the radar sensor to improve performance. TDRadarNet+ is specifically designed to incorporate a radar Doppler BEV as part of its input. Unlike TDRadarNet, which takes radar frame sequences directly as input, TDRadarNet+ takes one less radar frame and includes a single radar Doppler BEV.

VI. BAMA HIGH-RESOLUTION AUTOMOTIVE RADAR IMAGING DATASET

We implemented the proposed deep-learning-aided imaging radar signal processing chain shown in Fig. 4 on TI cascaded imaging radar, which is a chirp configurable MIMO radar with 12 TX and 16 RX antennas, cascaded by four radar transceivers of AWR2243 [25]. The azimuth 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 58 virtual array elements are overlapped (see Fig. 8). The 3-dB beamwidth of the imaging radar in azimuth is $\Delta\theta_{\mathrm{AZ}} = 2\arcsin(\frac{1.4\lambda}{\pi D_x}) \approx 1.2^\circ$, where $D_x = 42.5\lambda$ is the virtual array aperture in the horizontal direction. Antenna calibration is required to reduce the frequency, phase, and amplitude mismatches across those four radar transceivers. Instead of using a one-time boresight calibration method, we adopt angle-dependent calibration for better performance. The processed radar image is presented as a single-channel grayscale image instead of an RGB image because each pixel only presents the target's reflection intensity.

TABLE II
Multimodal Sensors

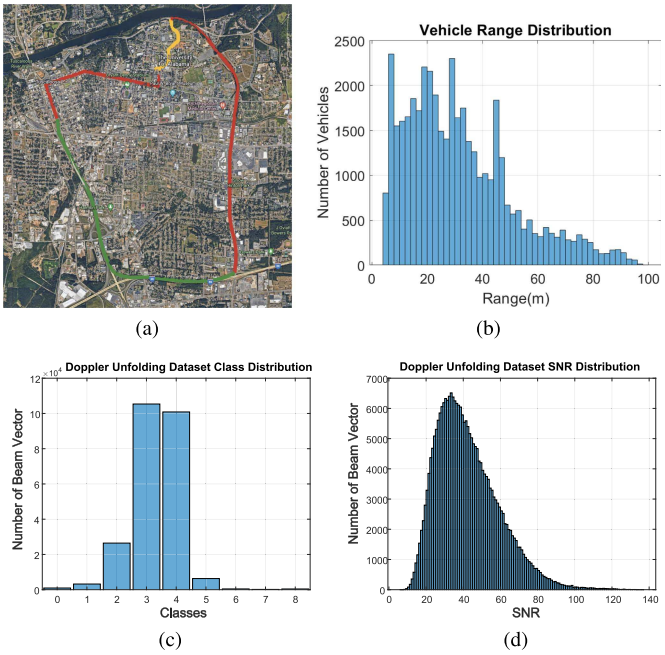
Sensors	Model
Radar	TI Imaging Radar, • Bandwidth: 384MHz • Range Resolution: 0.39m • Doppler Resolution: 0.23m/s • Azimuth Resolution: 1.2° • Azimuth FOV: 70°
LiDAR	Velodyne Ultra Puck VLP-32C • Azimuth Resolution: $0.1^\circ - 0.4^\circ$ • Vertical FOV: 40° • Maximum Range: 200 m
Camera	Teledyne FLIR Blackfly S, Stereo • Image Resolution: 2048×1536

Note our system, as per TI user manual, can detect objects up to 150 m in the MIMO mode and 350 m in the beamforming mode. In our study, the maximum detectable range is set to be 100 m because only one transmitter is selected to transmit at each time slot under TDM, resulting in lower transmit power, and labeling objects beyond 100 m is challenging, as obtaining ground truth from LiDAR and cameras on the same host vehicle can be difficult due to occlusion.

A. Data Acquisition and Distribution

Our field experiments included three multimodal 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. 9. The measurements of cameras and LiDAR are used as ground truth for labeling the radar data. The sensor features are summarized in Table II.

We drove over 30 min to collect data around the city of Tuscaloosa, AL, USA. As shown in Fig. 10(a), the experimental driving route consists of three types of driving scenarios, i.e., campus road, urban street, and highway.



(a) (b) (c) (d)

Fig. 10. (a) Data collection route in the city of Tuscaloosa, AL, USA. The lines with different colors denote the different driving scenarios. Yellow: campus road; Red: urban street; Green: highway. (b) Vehicle range distribution. (c) Doppler Unfolding Dataset class distribution. (d) SNR distribution of beam vector.

The BAMA dataset contains 14 800 radar BEV frames with synchronized stereo camera images and LiDAR 3-D point clouds. There are different types of objects of interest, including pedestrian, car, truck, and bus. Thanks to the high-resolution azimuth capability, the collected low-level radar BEV imaging contains rich object geometry features, as shown in Fig. 13. Specifically, the zoom-in radar BEV views of typical objects, such as bike, car, bus, truck, and pedestrian are plotted in Fig. 13(b). The high-resolution radar BEVs representing the objects' shape provide more information than radar point clouds [53] for object detection using the proposed TDRadarNet. For demonstration, in this article, we focus on vehicle detection only [35]. Using camera images and LiDAR 3-D point clouds as ground truth, a total number of 42 390 vehicles at various ranges are labeled with 2-D bounding boxes that encompass their physical dimensions (see examples in Fig. 15). 2-D bounding box labeling is a conventional label method for object detection and has been shown to be effective in many applications [70]. Vehicle range distribution is shown in Fig. 10(b). Examples under various driving scenarios are shown in Fig. 13 (also in Fig. 15).

B. Doppler Unfolding Dataset

We created a separate subdataset for our Doppler unfolding network using 1700 3-D radar data cube frames, from which we extracted a total of 244 140 beam vectors. The SNR of each beam vector was estimated using $\text{SNR} = 10\log_{10}(\mathbf{S}^2/\mathbf{N}^2)$, where $\mathbf{S}$ represents the signal amplitude and $\mathbf{N}$ represents the noise floor from the range–Doppler spectrum of each channel. In our processing, the range–Doppler 2-D FFT provides a signal processing gain of $10\log_{10}(N_{\text{fast}}N_{\text{slow}}) = 42.14$ dB, where $N_{\text{fast}} = 256$ and $N_{\text{slow}} = 64$ are the number of samples in the fast-time and slow-time dimensions, respectively. The class distribution and SNR distribution of the dataset are depicted in Fig. 10(c) and (d), respectively.

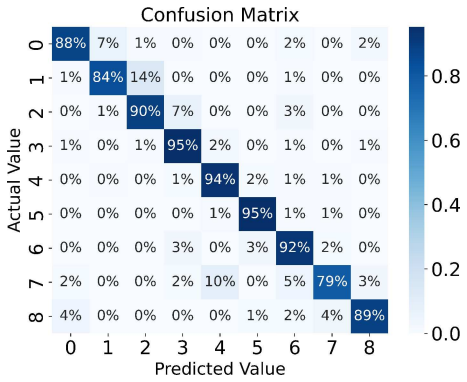
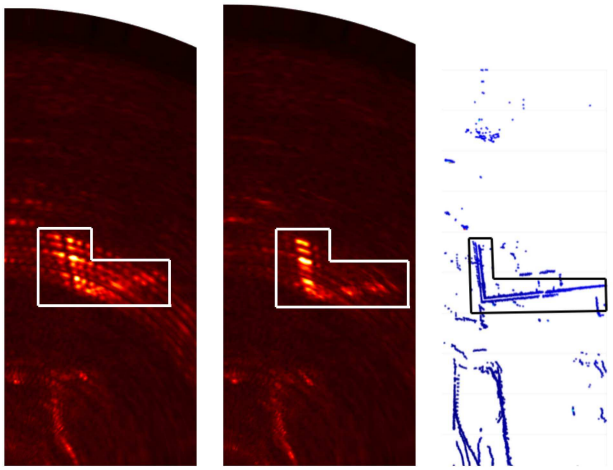


Fig. 11. Confusion matrix of Doppler Unfolding Dataset.

(a)

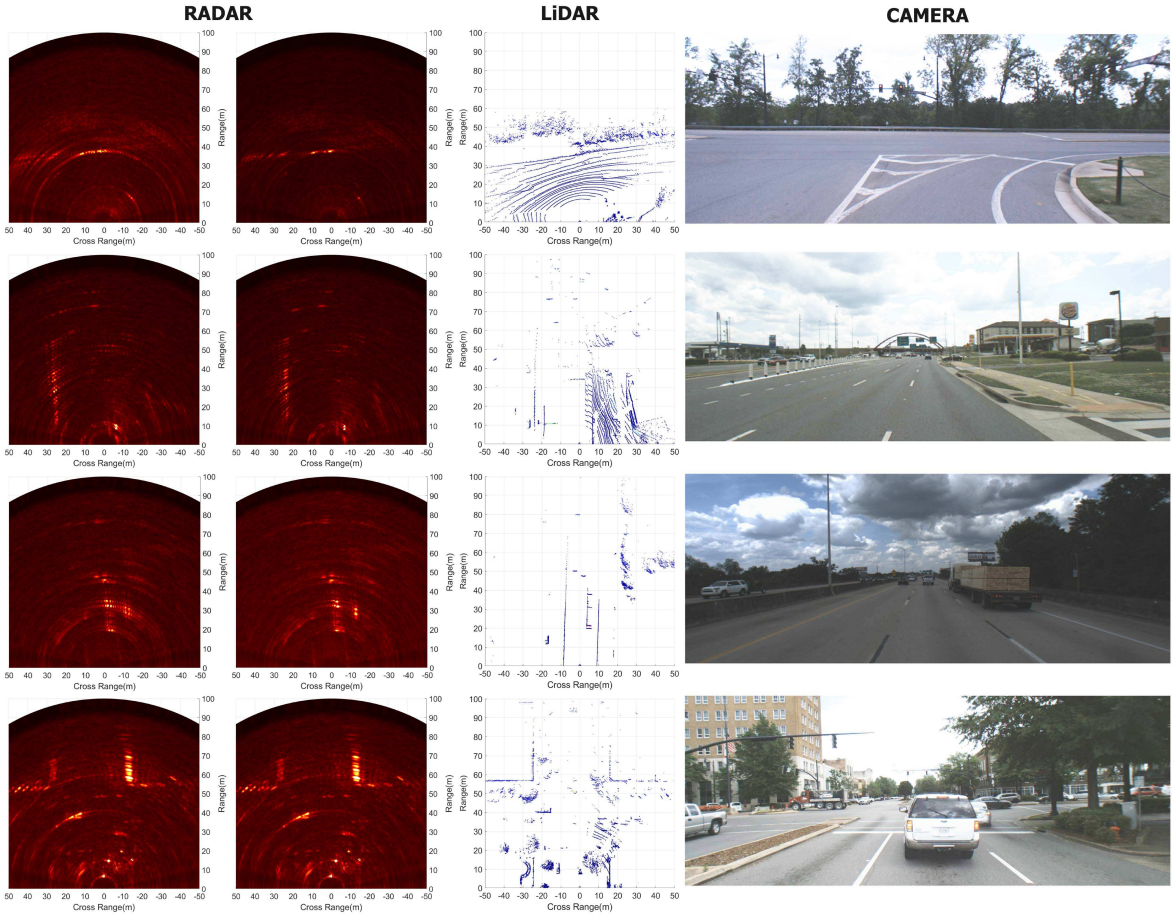


(b) (c) (d)

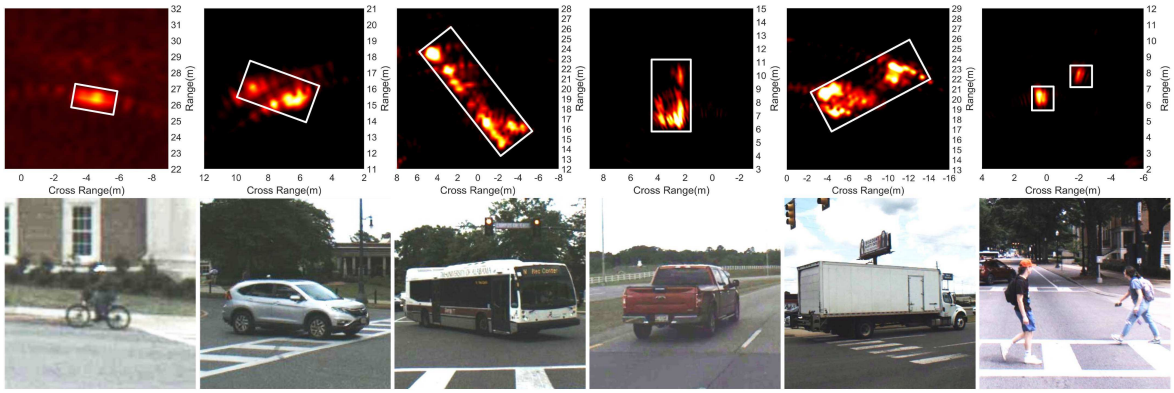
Fig. 12. Motion-induced error removed by Doppler Unfolding Network. (a) Camera image. (b) Blurred radar BEV. (c) Undistorted radar BEV. (d) LiDAR BEV.

VII. EXPERIMENTS

In this section, we evaluate the Doppler Unfolding Network and TDRadarNet with the BAMA dataset quantitatively and qualitatively. We demonstrate that our Doppler Unfolding Network performs accurate Doppler unfolding results on signals with various SNRs. We further analyze the TDRadarNet by comparing it with a baseline model and performing ablation experiments.



(a)



(b)

Fig. 13. (a) Examples from the BAMA dataset showing the effect of velocity compensation on radar BEV frames. The first column shows BEVs before motion compensation, and the second column shows BEVs after motion compensation via the proposed Doppler Unfolding Network. The third column shows LiDAR point clouds in BEV format, and the fourth column shows an image from the left camera. (b) Zoom-in radar BEV views of typical objects, such as bike, car, bus, truck, and pedestrian.

A. Doppler Unfolding Network

To train the Doppler Unfolding Network, we use a simulated dataset that contains 46 000 beam vectors with different velocities and DOAs, evenly distributed across nine classes. To further enhance the network's performance on real-world data, we use a small training set from the BAMA Doppler Unfolding dataset that contains 900 samples, evenly distributed across the same classes. The proposed network achieves an accuracy of **93.46%**. The confusion matrix for the BAMA dataset is presented in Fig. 11. Examples of radar BEV frames captured under various driving scenarios, both with and without compensation, are shown in Figs. 12 and 13. When driving at high speeds on a highway, motion-induced phase errors severely affect the radar BEVs, resulting in a blur effect that can cause false targets due to high sidelobes. After

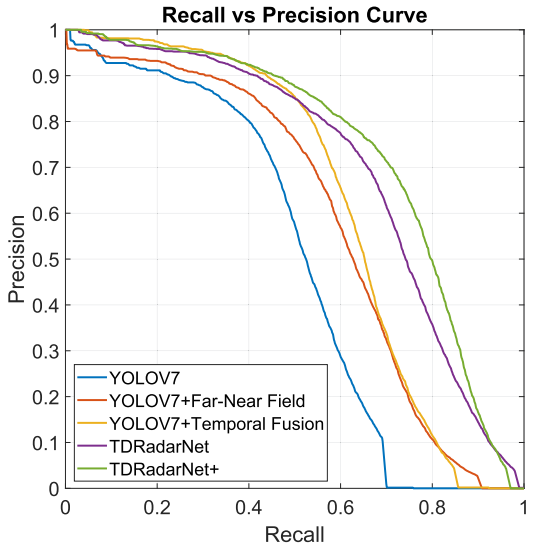


Fig. 14. Recall–precision curve.

compensation with the correct velocity predicted by the proposed Doppler Unfolding Network, the blur effect is significantly suppressed. As a result, false targets are successfully mitigated, demonstrating the effectiveness of the deep-learning-aided signal processing chain. In addition, we include examples from intersections where most targets are stationary or have low velocities, and compensation is not required. Overall, our proposed approach successfully suppresses the "motion-blur" effect of TDM MIMO radar for various driving scenarios.

B. TDRadarNet

We used a total of 14 800 high-resolution radar BEV frames, out of which 11 500 frames were selected for training, and the remaining 3300 frames were used for testing. It is important to note that our testing data were obtained from an independent sequence and were not obtained by randomly shuffling frames. This ensures that our testing data are representative of real-world scenarios and are not biased toward the training data. The YOLOv7 network trained with a single-frame dataset is used for the baseline. For the proposed TDRadarNet, we feed the network with a sequence of radar frames, and the annotations in the last frame of the input frame sequence are used as reference. The sequences of radar frames are overlapping in time order so that each frame is guaranteed to be trained and evaluated. The experiment was built in Python 3.8, PyTorch 1.10, CUDA 11.1 on four Nvidia RTX A6000 GPUs. The baseline model YOLOv7 using a single-frame dataset was trained 200 epochs with a batch size of 8. For the TDRadarNet, the model was trained 200 epochs to ensure convergence, with a batch size of 8, and a linear decaying learning rate initialized as 0.001. Besides, we performed two ablation experiments to evaluate the far-near field and temporal fusion design, with all the hyperparameters the same as the implementation of TDRadarNet.

TABLE III
Object Detection Results of Deep Learning Models in Precision, Recall, F-1 Score, and AP

Networks	Precision	Recall	F1-score	AP
YOLOv7 (baseline)	58.3%	49.7%	53.7%	49.3%
YOLOv7 + Far-Near Field	62.8%	57.9%	60.2%	58.9%
YOLOv7 + Temporal Fusion	67.9%	59.0%	63.2%	62.9%
TDRadarNet	68.9%	66.8%	67.8%	70.3%
TDRadarNet+	74.9%	67.4%	70.9%	73.9%

To assess the detection performance, we employ several standard metrics including precision, recall, F1-score, and average precision (AP), using an IoU threshold of 0.5. In addition, we present the recall versus precision curves for all the models, as illustrated in Fig. 14. Notably, YOLOv7 reports its precision and recall based on the optimal F1-score, which ensures a desirable tradeoff between precision and recall. To maintain consistency, we follow the same approach.

Table III presents the quantitative evaluation results. Our proposed TDRadarNet outperforms the baseline model, achieving a 10.6% improvement in precision, a 17.1% improvement in recall, a 14.1% improvement in F1-score, and a 21.0% improvement in AP. By leveraging Doppler information, TDRadarNet+ achieves even better performance, with a 6.0% improvement in precision, a 0.6% improvement in recall, a 3.1% improvement in F1-score, and a 3.6% improvement in AP compared to TDRadarNet. Regarding the far–near-field design and temporal fusion design, our ablated models generally outperform the baseline method.

To assess the complexity of our proposed TDRadarNet, we evaluated the number of parameters and FPS. We compared it with the baseline YOLOv7 network, which has 36.5M parameters and achieves 100 FPS. Since TDRadarNet and TDRadarNet+ share the same architecture, they have the same number of parameters and FPS. Therefore, we focused on evaluating the complexity of TDRadarNet. The results showed that TDRadarNet has 80.2M parameters and achieves 34 FPS. While the number of parameters is higher than the baseline, the achieved FPS is still suitable for real-world applications.

The representative detection results over three scenarios, the corresponding original radar frames, and ground truth annotations are shown in Fig. 15. In the second column of Fig. 15, the implementation of far–near field not only helps to detect vehicles that are far in distance and have a lower resolution but also improves the detection in the near field, by learning field-related specific features, respectively. In addition, compared to the baseline model, the temporal fusion also shows better performance, especially in the example of the highway scenario, in providing accurate and complete predictions. The observations overall agree with the quantitative evaluation. The proposed TDRadarNet has shown superior capability in detecting objects in both far and near fields and performs well in the campus and highway scenarios where fewer data ($n = 1700$ and $n = 2300$, respectively) are available compared to the urban scenario ($n = 7500$).

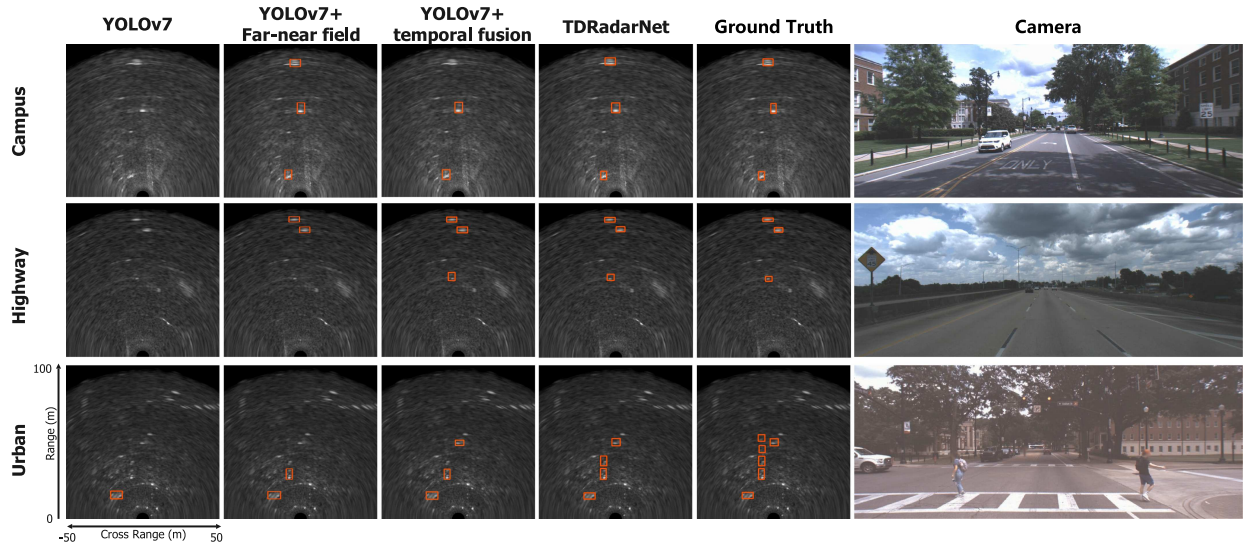


Fig. 15. Examples from our test set. Detection results are marked in red. (Note: The pixels represent locations beyond 100 m from ego vehicle located at original (0, 0) are set as zeros, corresponding to dark pixels.)

VIII. CONCLUSION

In this article, we reported a novel deep-learning-aided TDM MIMO radar signal pipeline that contains a network for solving the Doppler unfolding problem. After extensive tests, we showed that our signal process pipeline is able to generate undistorted radar BEVs containing rich object features under various driving scenarios.

In an innovative way, the proposed TDRadarNet combined temporal fusion and far-near field designs to perform vehicle detection tasks on high-resolution radar BEVs, which do not suffer from information loss compared to radar point clouds. Furthermore, the enhanced TDRadarNet+ maximized the utilization of Doppler information and achieved better vehicle detection performance. We investigated and compared the detection performance of TDRadarNet and TDRadarNet+ on our own dataset containing 14 800 frames of high-resolution radar BEVs, together with synchronized stereo camera images and LiDAR 3-D point clouds, collected in diverse driving scenarios. The results showed that the proposed model is superior to the image-based neural network baseline model in producing accurate vehicle detection results. The ablation experiments demonstrated the effectiveness of our designs.

The potential limitation of TDRadarNet could be the extra computational cost introduced by the temporal fusion over a sequence of frames. An investigation of the tradeoff between the number of historical frames fed into neural network and detection accuracy is desired. The imbalance of radar BEVs in different ranges and driving scenarios may be an issue that would impact the TDRadarNet training and detection performance. To overcome the limitations, in future work, we plan to investigate the lightweight deep learning detection model, conduct additional experiments to optimize the number of historical frames in a sequence, and implement data augmentation for limited available driving scenarios to further enhance the TDRadarNet detection accuracy and efficiency.

REFERENCES

[1] R. Zheng, H. Liu, and S. Sun, "A deep learning approach for Doppler unfolding in automotive TDM MIMO radar," in *Proc. IEEE 56th Annu. Asilomar Conf. Signals, Syst., Comput.*, 2022, pp. 1121–1125.

[2] R. Zheng, S. Sun, D. Scharff, and T. Wu, "SpectraNet: A high resolution imaging radar deep neural network for autonomous vehicles," in *Proc. IEEE Sens. Array Multichannel Signal Process. Workshop*, 2022, pp. 301–305.

[3] S. Patole, M. Torlak, D. Wang, and M. Ali, "Automotive radars: A review of signal processing techniques," *IEEE Signal Process. Mag.*, vol. 34, no. 2, pp. 22–35, Mar. 2017.

[4] F. Engels, P. Heidenreich, A. M. Zoubir, F. K. Jondral, and M. Wintermantel, "Advances in automotive radar: A framework on computationally efficient high-resolution frequency estimation," *IEEE Signal Process. Mag.*, vol. 34, no. 2, pp. 36–46, Mar. 2017.

[5] 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, Jul. 2020.

[6] C. Waldschmidt, J. Hasch, and W. Menzel, "Automotive radar—From first efforts to future systems," *IEEE J. Microw.*, vol. 1, no. 1, pp. 135–148, Jan. 2021.

[7] M. Markel, *Radar for Fully Autonomous Driving*. Boston, MA, USA: Artech House, 2022.

[8] 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, Jun. 2021.

[9] G. Duggal, S. Vishwakarma, K. V. Mishra, and S. S. Ram, "Doppler-resilient 802.11ad-based ultrashort range automotive joint radar-communications system," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 5, pp. 4035–4048, Oct. 2020.

[10] F. Engels, P. Heidenreich, M. Wintermantel, L. Stäcker, M. Al Kadi, and A. M. Zoubir, "Automotive radar signal processing: Research directions and practical challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 4, pp. 865–878, Jun. 2021.

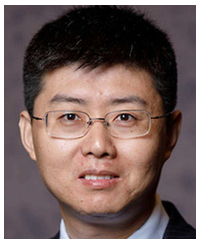
[11] D. Feng et al., "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1341–1360, Mar. 2021.

[12] P. Sun et al., "Scalability in perception for autonomous driving: Waymo open dataset," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2446–2454.
[13] Y. Xiao, F. Codevilla, A. Gurram, O. Urfalioglu, and A. M. López, "Multimodal end-to-end autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 537–547, Jan. 2022.
[14] Y. Cui et al., "Deep learning for image and point cloud fusion in autonomous driving: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 722–739, Feb. 2022.
[15] Y. Li and J. Ibanez-Guzman, "Lidar for autonomous driving: The principles, challenges, and trends for automotive Lidar and perception systems," *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 50–61, Jul. 2020.
[16] S. Chen, B. Liu, C. Feng, C. Vallespi-Gonzalez, and C. Wellington, "3D point cloud processing and learning for autonomous driving: Impacting map creation, localization, and perception," *IEEE Signal Process. Mag.*, vol. 38, no. 1, pp. 68–86, Jan. 2021.
[17] S. Sun and Y. D. Zhang, "Four-dimensional high-resolution automotive radar imaging exploiting joint sparse-frequency and sparse-array design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 8413–8417.
[18] Accessed: May, 2023. [Online]. Available: https://www.zf.com/products/en/cars/products_64255.html
[19] Accessed: May, 2023. [Online]. Available: https://www.continental-automotive.com/en-gl/Passenger-Cars/Autonomous-Mobility/Enablers/Radars/Long-Range-Radar/ARS540
[20] J. Li and P. Stoica, "MIMO radar with colocated antennas," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 106–114, Sep. 2007.
[21] J. Li and P. Stoica, *MIMO Radar Signal Processing*. Hoboken, NJ, USA:Wiley, 2009.
[22] J. Bergin and J. R. Guerci, *MIMO Radar: Theory and Application*. Boston, MA, USA: Artech House, 2018.
[23] S. L. Cassidy, S. Pooni, M. Cherniakov, E. G. Hoare, and M. S. Gashinova, "High resolution automotive imaging using MIMO radar and Doppler beam sharpening," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 2, pp. 1495–1505, Apr. 2023.
[24] C. M. Schmid, R. Feger, C. Pfeffer, and A. Stelzer, "Motion compensation and efficient array design for TDMA FMCW MIMO radar systems," in *Proc. 6th Eur. Conf. Antennas Propag.*, 2012, pp. 1746–1750.
[25] "Design Guide: TIDEP-01012 Imaging Radar Using Cascaded MMWave Sensor Reference Design, 2020, Rev. A," Texas Instrum., Dallas, TX, USA, 2020. [Online]. Available: https://www.ti.com/lit/ug/tiduen5a/tiduen5a.pdf
[26] K. Christopher, S. S. Blackman, R. A. Rosen, and R. J. Dempster, "Methods for resolving radar ambiguities using multiple hypothesis tracking," U.S. Patent US2 013 009 361 7A1, Apr. 18, 2013.
[27] S.-J. Shin and E. H. Kim, "Track-before-resolving: Doppler ambiguity resolution at low SNR," *IET Radar, Sonar Navigat.*, vol. 17, pp. 117–127, 2023.
[28] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast encoders for object detection from point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12697–12705.
[29] Y. Zhou and O. Tuzel, "VoxelNet: End-to-end learning for point cloud based 3D object detection," in *Proc. Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4490–4499.
[30] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 77–85.
[31] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, 2017, pp. 5105–5114.
[32] A. Danzer, T. Griebel, M. Bach, and K. Dietmayer, "2D car detection in radar data with PointNets," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2019, pp. 61–66.
[33] B. Yang, R. Guo, M. Liang, S. Casas, and R. Urtasun, "RadarNet: Exploiting radar for robust perception of dynamic objects," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 496–512.
[34] 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 Conf. Appl. Comput. Vis.*, 2021, pp. 504–513.
[35] K. Qian, S. Zhu, X. Zhang, and L. E. Li, "Robust multimodal vehicle detection in foggy weather using complementary Lidar and radar signals," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 444–453.
[36] J. Rebut, A. Ouaknine, W. Malik, and P. Pérez, "Raw high-definition radar for multi-task learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 17021–170 30.
[37] S. Madani, J. Guan, W. Ahmed, S. Gupta, and H. Hassanieh, "Radatron: Accurate detection using multi-resolution cascaded MIMO radar," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 160–178.
[38] J. Bechter, F. Roos, and C. Waldschmidt, "Compensation of motion-induced phase errors in TDM MIMO radars," *IEEE Microw. Wireless Compon. Lett.*, vol. 27, no. 12, pp. 1164–1166, Dec. 2017.
[39] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 3354–3361.
[40] H. Caesar et al., "nuScenes: A multimodal dataset for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11618–11628.
[41] 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 Int. Conf. Robot. Autom.*, 2020, pp. 6433–6438.
[42] M. Meyer and G. Kuschk, "Automotive radar dataset for deep learning based 3D object detection," in *Proc. 16th Eur. Radar Conf.*, 2019, pp. 129–132.
[43] M. Sheeny, E. D. Pellegrin, M. Saptarshi, A. Ahrabian, S. Wang, and A. Wallace, "RADIATE: A radar dataset for automotive perception," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2021, pp. 1–7.
[44] 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 Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 450–457.
[45] A. Ouaknine, A. Newson, J. Rebut, F. Tupin, and P. Pérez, "CARRADA dataset: Camera and automotive radar with range-angle-Doppler annotations," in *Proc. 25th Int. Conf. Pattern Recognit.*, 2021, pp. 5068–5075.
[46] O. Schumann et al., "RadarScenes: A real-world radar point cloud data set for automotive applications," in *Proc. IEEE 24th Int. Conf. Inf. Fusion*, 2021, pp. 1–8.
[47] A. Palffy, E. Pool, S. Baratam, J. Kooij, and D. Gavrila, "Multi-class road user detection with 3+1D radar in the view-of-Delft dataset," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 4961–4968, Apr. 2022.
[48] Y. Zhou, L. Liu, H. Zhao, M. López-Benítez, L. Yu, and Y. Yue, "Towards deep radar perception for autonomous driving: Datasets, methods, and challenges," *Sensors*, vol. 22, no. 11, 2022, Art. no. 4208.
[49] E. Mason, B. Yonel, and B. Yazici, "Deep learning for radar," in *Proc. IEEE Radar Conf.*, 2017, pp. 1703–1708.
[50] D. Brodeski, I. Bilik, and R. Giryes, "Deep radar detector," in *Proc. IEEE Radar Conf.*, 2019, pp. 1–6.
[51] J. Lien et al., "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–19, 2016.
[52] C. Li, J. Cummings, J. Lam, E. Graves, and W. Wu, "Radar remote monitoring of vital signs," *IEEE Microw. Mag.*, vol. 10, no. 1, pp. 47–56, Feb. 2009.
[53] N. Scheiner, F. Kraus, N. Appenrodt, J. Dickmann, and B. Sick, "Object detection for automotive radar point clouds—A comparison," *AI Perspectives*, vol. 3, no. 1, pp. 1–23, 2021.
[54] A. Ouaknine, A. Newson, P. Pérez, F. Tupin, and J. Rebut, "Multi-view radar semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 15671–15680.
[55] B. Major et al., "Vehicle detection with automotive radar using deep learning on range-azimuth-Doppler tensors," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2019, pp. 924–932.

[56] X. Gao, G. Xing, S. Roy, and H. Liu, "Ramp-CNN: A novel neural network for enhanced automotive radar object recognition," *IEEE Sens. J.*, vol. 21, no. 4, pp. 5119–5132, Feb. 2021.
[57] A. Zhang, F. E. Nowruzi, and R. Laganiere, "RADDet: Range-azimuth-Doppler based radar object detection for dynamic road users," in *Proc. 18th Conf. Robots Vis.*, 2021, pp. 95–102.
[58] X. Dong, P. Wang, P. Zhang, and L. Liu, "Probabilistic oriented object detection in automotive radar," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 458–467.
[59] J. Peršić, L. Petrović, I. Marković, and I. Petrović, "Spatio-temporal multisensor calibration based on Gaussian processes moving object tracking," *IEEE Trans. Robot.*, vol. 37, no. 5, pp. 1401–1415, Oct. 2020.
[60] M. Meyer, G. Kuschk, and S. Tomforde, "Graph convolutional networks for 3D object detection on radar data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop*, 2021, pp. 3060–3069.
[61] S. Tokoro, K. Kuroda, A. Kawakubo, K. Fujita, and H. Fujinami, "Electronically scanned millimiter-wave radar for precrash safety and adaptive cruise control system," in *Proc. IEEE Intell. Veh. Symp.*, 2003, pp. 304–309.
[62] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.
[63] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotation invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 17, no. 7, pp. 984–995, Jul. 1989.
[64] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, 2007.
[65] D. Zoeke and A. Ziroff, "Phase migration effects in moving target localization using switched MIMO arrays," in *Proc. Eur. Radar Conf.*, 2015, pp. 85–88.
[66] A. J. Weiss and B. Friedlander, "Array shape calibration using sources in unknown locations—A maximum likelihood approach," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1958–1966, Dec. 1989.
[67] C. M. Schmid, S. Schuster, R. Feger, and A. Stelzer, "On the effects of calibration errors and mutual coupling on the beam pattern of an antenna array," *IEEE Trans. Antennas Propag.*, vol. 61, no. 8, pp. 4063–4072, Aug. 2013.
[68] C. Balanis, *Antenna Theory: Analysis and Design*, 4th ed. Hoboken, NJ, USA: Wiley, 2016.
[69] M. A. Richards, *Fundamentals of Radar Signal Processing*, 3rd ed. New York, NY, USA: McGraw-Hill, 2022.
[70] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.

Ruxin Zheng (Student Member, IEEE) received the B.S. degree from the University of Pittsburgh, Pittsburgh, PA, USA, in 2018, and the M.S. degree from the University of Michigan, Ann Arbor, MI, USA, in 2020, both in electrical engineering. He is currently working toward the Ph.D. degree in electrical engineering with the University of Alabama, Tuscaloosa, AL, USA.

His research interests include automotive radar, radar signal processing, multiple-input multiple-output radar with sparse sensing, and machine learning.



Shunqiao Sun (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Rutgers, The State University of New Jersey, New Brunswick, NJ, USA, in 2016.

From 2016 to 2019, he was with the Radar Core Team, Aptiv, Malibu, CA, USA, where he has worked on advanced radar signal processing and machine learning algorithms for self-driving vehicles and lead the development of direction-of-arrival estimation techniques for next-generation short-range radar sensor that has been used in over 120 million automotive radar units. In 2019, he joined the Department of Electrical and Computer Engineering, The University of Alabama, Tuscaloosa, AL, USA, as a tenure-track Assistant Professor. His research interests include the interface of statistical and sparse signal processing with mathematical optimizations, automotive radar, multiple-input multiple-output radar, machine learning, and smart sensing for autonomous vehicles.

Dr. Sun received the 2016 IEEE Aerospace and Electronic Systems Society Robert T. Hill Best Dissertation Award for his thesis "MIMO radar with Sparse Sensing" and the Best Student Paper Award at 2020 IEEE Sensor Array and Multichannel Signal Processing Workshop. He is the Vice Chair of the IEEE Signal Processing Society's Autonomous Systems Initiative Steering Committee (for period 2023–2024). He is an Associate Editor for IEEE OPEN JOURNAL OF SIGNAL PROCESSING.

Hongshan Liu (Student Member, IEEE) received the B.S. degree in physics from Zhejiang University, Hangzhou, China, in 2018, and the M.S. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2020. She is currently working toward the Ph.D. degree in electrical engineering with the University of Alabama, Tuscaloosa, AL, USA.

Her research interests include deep-learning-based image processing.

Teresa Wu (Senior Member, IEEE) received the Ph.D. degree in industrial engineering from the University of Iowa, Iowa City, IA, USA, in 2001.

She is currently a Professor of Industrial Engineering Program with the School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, USA. She has authored or coauthored more than 140 journal articles in journals such as IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and *Information Science*. Her current research interests include deep learning on heterogenous data, health informatics, and distributed decision support.