A Deep Automotive Radar Detector Using the *RaDelft* Dataset

Ignacio Roldan, *Graduate Student Member, IEEE*, Andras Palffy, *Member, IEEE*, Julian F. P. Kooij, *Member, IEEE*, Dariu M. Gavrila, *Member, IEEE*, Francesco Fioranelli, *Senior Member, IEEE*, and Alexander Yarovoy, *Fellow, IEEE*

***Abstract*— The detection of multiple extended targets in complex environments using high-resolution automotive radar is considered. A data-driven approach is proposed where unlabeled synchronized lidar data are used as ground truth to train a neural network (NN) with only radar data as input. To this end, the novel, large-scale, real-life, and multisensor *RaDelft* dataset has been recorded using a demonstrator vehicle in different locations in the city of Delft, The Netherlands. The dataset, as well as the documentation and example code, is publicly available for those researchers in the field of automotive radar or machine perception. The proposed data-driven detector can generate lidar-like point clouds (PCs) using only radar data from a high-resolution system, which preserves the shape and size of extended targets. The results are compared against conventional constant false alarm rate (CFAR) detectors as well as variations of the method to emulate the available approaches in the literature, using the probability of detection, the probability of false alarm, and the Chamfer distance (CD) as performance metrics. Moreover, an ablation study was carried out to assess the impact of Doppler and temporal information on detection performance. The proposed method outperforms different baselines in terms of CD, achieving a reduction of 77% against conventional CFAR detectors and 28% against the modified state-of-the-art deep learning (DL)-based approaches.**

***Index Terms*— Automotive radar, deep learning (DL), point cloud (PC) generation, radar dataset, radar target detection.**

I. Introduction

IN THE domain of environmental sensing technology, radar sensors can provide unique advantages over other sensors. While lidar offers high-resolution imaging capabilities, making it excellent for detailed environmental mapping, radar provides superior performance in adverse weather conditions, such as fog or rain, or in the case of low-light conditions [1]. Furthermore, radar can accurately and directly measure objects' velocity via the Doppler effect. All this makes radar a crucial sensor for vehicular autonomy [2].

A notable trend in automotive radar is the shift toward imaging radar, which achieves high angular resolution in both azimuth and elevation by leveraging a larger number of antennas and thus a larger aperture [3]. Furthermore, neural networks (NNs) and deep learning (DL) techniques are increasingly being applied to signal and data processing [4]. These algorithms can excel in multiple steps of the radar signal processing pipeline, such as detection [5], [6], [6], [7], [8], [9], [10], [11], classification [12], [13], [14], and signal enhancement [15], offering a richer interpretation of radar data. However, their effectiveness relies on extensive and high-quality datasets for training, to accurately identify and react to diverse driving scenarios. To the best of authors' knowledge, there is a lack of suitable public datasets for radar practitioners where analog-to-digital converter (ADC)-level data from large-aperture radars are collected using real vehicles. Therefore, the *first contribution* of this article is the introduction of *RaDelft*, a large-scale, real-world multisensory dataset recorded in various driving scenarios in the city of Delft, The Netherlands, which is publicly shared.

In terms of signal processing, challenges remain for the integration of radar technology into automotive systems. A primary hurdle in this context is the use of the well-known constant false alarm rate (CFAR) detectors for generating radar point clouds (PCs) from the dense radar data cube. While CFAR detectors have proven optimal in other environments [16], their application in the dynamic and unpredictable conditions of road traffic scenarios suffers from poor performance [7], [17]. Namely, they are designed to maintain a constant rate of false alarms amidst varying clutter, but they struggle to adapt to the rapidly changing environments typical of roadways. Complications such as nonuniform clutter (or the lack of reliable clutter models for this task), target masking, and shadowing can significantly reduce the effectiveness of CFAR detectors in automotive radar settings. Additionally, CFAR detectors are constrained by a fundamental limitation: they typically assume a fixed, expected target size based on predefined guard and training cell hyperparameters. However, in an automotive context, this assumption is problematic as the size of potential targets can widely vary, ranging from medium-sized objects such as

Received 30 May 2024; revised 11 September 2024; accepted 17 October 2024. Date of publication 23 October 2024; date of current version 1 November 2024. *(Corresponding author: Ignacio Roldan.)*

Ignacio Roldan, Francesco Fioranelli, and Alexander Yarovoy are with the Microwave Sensing, Signals and Systems (MS3) Group, Department of Microelectronics, Delft University of Technology (TU Delft), 2628 CD Delft, The Netherlands (e-mail: i.roldanmontero@tudelft.nl; f.fioranelli@tudelft.nl; a.yarovoy@tudelft.nl).

Andras Palffy is with the Microwave Sensing, Signals and Systems (MS3) Group, Department of Microelectronics, Delft University of Technology (TU Delft), 2628 CD Delft, The Netherlands, and also with Perciv AI, 2628 BC Delft, The Netherlands (e-mail: andras.palffy@perciv.ai).

Julian F. P. Kooij and Dariu M. Gavrila are with the Intelligent Vehicles (IV) Group, Department of Cognitive Robotics, Delft University of Technology (TU Delft), 2628 CD Delft, The Netherlands (e-mail: j.f.p.kooij@tudelft.nl; d.m.gavrila@tudelft.nl).

Digital Object Identifier 10.1109/TRS.2024.3485578

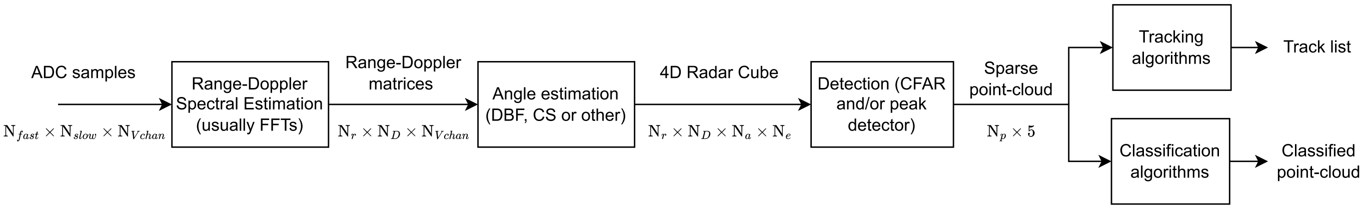
ADC samples
$N_{fast} \times N_{slow} \times N_{Vchan}$
Range-Doppler Spectral Estimation (usually FFTs)
Range-Doppler matrices
$N_r \times N_D \times N_{Vchan}$
Angle estimation (DBF, CS or other)
4D Radar Cube
$N_r \times N_D \times N_a \times N_e$
Detection (CFAR and/or peak detector)
Sparse point-cloud
$N_p \times 5$
Tracking algorithms
Track list
Classification algorithms
Classified point-cloud

Fig. 1. Typical radar processing pipeline, from the raw ADC samples to the output of classification and tracking steps. N_{fast} and N_{slow} are the number of samples in a chirp and in a CPI, respectively. N_{Vchan} are the number of virtual channels, in an MIMO system the product of the number of Tx and Rx channels. N_r, N_D, N_a, and N_e are the number of range, Doppler, azimuth, and elevation cells, respectively. Finally, N_p is the number of points after the detector, with the three spatial coordinates plus Doppler and power.

pedestrians to large vehicles such as trucks or buses. Moreover, the perceived size of these targets in the radar's angular dimension changes with distance. Large objects occupying multiple cells at close range can appear as simpler point-like targets at further distances. This relationship between angular target size and distance adds another layer of complexity to using CFAR detectors in automotive radar, necessitating alternative solutions to accurately detect and classify objects under varying road conditions.

To address these limitations, the *second contribution* of this work is to present a new data-driven radar target detector using a unique cross-sensor supervision pipeline. The proposed data-driven detector is initially trained with synchronized radar and lidar data together, and can subsequently generate denser point clouds using only raw data from a high-resolution automotive radar. The proposed approach is extensively validated using the aforementioned *RaDelft* dataset.

Compared to the initial results presented in our conference submission [18], two additional contributions are presented in this work. First, the proposed data-driven radar detector is expanded to include temporal information across frames, and a more rigorous analysis of the impact of each processing block is included. Second, the multisensor dataset used for validation, *RaDelft*, is presented and shared with the broader research community, including example code for easier utilization [19].

The rest of this article is organized as follows. As automotive radar is part of a wider multidisciplinary field on autonomous vehicles, clarifying the terminology used in this work is important to prevent confusion. This is done in Section II, which also briefly reviews the conventional radar processing pipeline. Section III reviews the available automotive radar datasets and summarizes the state of the art of automotive radar detectors. Section IV introduces our new publicly available dataset *RaDelft*, detailing its characteristics for data-driven approaches. Our proposed data-driven detector is presented in Section V. Section VI shows the results of the proposed method and compares them with those of conventional CFAR detectors. Finally, Section VII concludes this article.

II. Terminology and Radar Processing Review

In this section, the terminology used in this work is first clarified, followed by a brief review of the conventional radar processing pipeline and its steps.

A. Terminology

In recent years, automotive radar has become part of a wider multidisciplinary field in autonomous vehicles where scientists from different backgrounds are cooperating. As different research communities might use different terms [12], [20], a list of definitions used in this work is provided here.

1) *Raw radar data* or *ADC data* refer to the complex baseband samples the ADC provides at each receiver channel.
2) *Virtual channel* or *channel* refers to one of the multiple unique combinations of Tx–Rx antenna in a multi-in multi-out (MIMO) radar, meaning the signal transmitted from a Tx is received, downconverted, and sampled at the Rx.
3) *Radar frame* refers to the set of ADC samples from a coherent processing interval (CPI) of each virtual channel. It has dimensions of $N_{\text{fast}} \times N_{\text{slow}} \times N_{\text{Vchan}}$, where these are the number of samples in fast time, the number of samples in slow time, and the number of virtual channels, respectively.
4) *Radar cube* refers to the spherical coordinate, discretized representation of the radar data, meaning the range, azimuth, elevation, and Doppler estimation have already been performed. Each cell in the *radar cube* contains a scalar value indicating the reflected power in that cell. The size of each cell is related to the characteristics of the radar, such as the transmitted bandwidth or the antenna array topology. In general, the cells do not have the same size over the whole grid.
5) An *extended target* is a target occupying multiple cells in one or several dimensions, in contrast to a *point target*, which occupies a single cell. Point targets present a clear peak in the estimation space (range–Doppler–angle), while extended targets do not.
6) *Detection* is the binary decision problem determining whether a *radar cube* cell contains only noise or noise plus target. On the other hand, *classification* aims to associate a class to each detected cell, such as "pedestrian," "vehicle," or "light pole." In general, these two tasks are treated as two blocks in a conventional radar processing pipeline.
7) *3-D occupancy grid* refers to a binary cube, also in spherical coordinates, which contains ones in voxels that are occupied by detected targets, and zeros otherwise. Such a *3-D occupancy grid* could be generated directly from a lidar point cloud, but also from a *radar cube* through a detector as this work aims to. In the latter

case, the resulting *3-D occupancy grid* and the *radar cube* share the same grid.

8) *Point cloud* refers to a set of N_p points, each containing L features that result from selecting only those cells containing ones in a *3-D occupancy grid* and converting them to Cartesian coordinates. For radar point clouds, it is typically assumed that $L = 5$, adding Doppler and power information to the three spatial dimensions, while for lidar point clouds, $L = 4$ since Doppler is not provided.

B. Radar Processing Pipeline Review

The conventional radar processing pipeline is illustrated in Fig. 1. The steps are as follows.

1) Range and Doppler spectral estimation is performed from the baseband or ADC samples organized in fast-time, slow-time, and channel dimensions. Usually, this is achieved by applying a window with the fast Fourier transform (FFT) algorithm independently in fast time and slow time. However, this step may be enhanced by compensating the range/Doppler migration due to ego-vehicle and target motion [21].
2) Once a range–Doppler matrix is computed per channel, the angle estimation is performed (1-D in azimuth or 2-D in both azimuth and elevation, depending on the antenna array topology). Direction of arrival (DoA) estimation is a current area of widespread interest, with much active research. Usually, digital beam-forming (DBF) is used for simplicity by means of FFT-based implementation, but many research works explore alternatives such as compressive sensing approaches [22], [23], Doppler beam sharpening [24], [25], or machine learning [26]. Sometimes, especially in real-time embedded systems, the detection stage is performed before the angle estimation to reduce the computational load [27], sacrificing the increase in signal-to-noise ratio (SNR) due to spatial coherent integration prior to detection. This process outputs a 4-D radar cube.
3) The detection stage then identifies the cells that contain the targets. Usually, a combination of a CFAR detector in some dimensions and peak finding in the rest is used though some works have also explored using machine learning algorithms [5], [6], [6], [7], [8], [9], [10]. In this stage, the data are often sparse since most of the space in the field of view (FoV) does not reflect sufficient power or is simply empty. The detector outputs a 3-D occupancy grid, but a conversion to point cloud is usually performed since it is a convenient format for visualizations or for dataset storage.
4) After the detection process and the generation of a point cloud, additional steps can be implemented to extract more task-relevant information. For instance, in the automotive context, it is critical to know the nature of each of the detected points to make the appropriate decisions, meaning if this originated from a pedestrian, a vehicle, or some road infrastructure, among others. Therefore, it is common to apply a classifier on the point cloud, usually based on DL techniques [12], [13], [14]. If needed for the application, tracking algorithms can also be applied on the point cloud by using past information to reduce the estimation noise, eliminate false detections, and predict future target positions based on the trajectory. In the automotive radar domain, tracking algorithms have to deal with the problem of the extended nature of targets over the angular domain [28], [29].

III. Related Work

This work introduces two contributions: the recording and sharing of the *RaDelft* dataset and the proposed data-driven detector. Therefore, two related work subsections are included to review the state of the art and highlight the need for new radar datasets and new detection algorithms.

A. Radar Datasets

Several automotive radar datasets have recently been published for different tasks, covering many of the processing steps listed in Section II-B. However, most of them are unsuitable or, at the very least, limited for radar practitioners since the data are already processed, often to the point cloud level. Thus, it is impossible to apply signal processing algorithms that operate on lower level data. Some datasets also provide the radar cube data, but few give the raw ADC data needed to test advanced signal processing methods. Essentially, each already-performed processing step limits the scope of the research that can be performed with that data. On the other hand, this simplifies the steps needed to make it suitable for other subsequent tasks.

A recent summary of the available automotive radar datasets can be found in [30]. Nevertheless, in this article, only those datasets providing data before the point cloud level of processing are considered since they are the most useful for radar practitioners. Table I summarizes such datasets. As can be seen, most of these available datasets are recorded with automotive radars with linear antenna arrays, meaning that there is only azimuth resolution and no information about the elevation of targets. While useful for some tasks, this type of data is not representative of the data of the next-generation 4-D radars that are becoming the standard in the automotive field. On the other hand, some datasets already include a 4-D imaging radar [6], [31], [32]. The RADial [31] dataset provides ADC data level suitable for radar practitioners, but the array topology used is not public, and thus advanced array processing methods cannot be applied. The ColorRadar [32] dataset uses a commercially available radar; therefore, its datasheet is public. However, most of the scenes are recorded indoors and without a vehicle. Moreover, camera information is not provided. Finally, the K-Radar [6] dataset is the most complete, providing range–azimuth–elevation–Doppler cubes, many auxiliary sensors, and useful code to parse the data. However, no ADC-level data are provided, which may limit the potential research scope of the dataset.

Considering the limitations of the aforementioned public datasets, this work presents a new dataset, *RaDelft*, aiming to close the gaps in the existing available datasets collected

TABLE I

Available Public Datasets Providing Either ADC or Predetection Data. In the *Data Type* Column, R, D, A, E, and C Stand for Range, Doppler, Azimuth, Elevation, and Channel, Respectively, While PC Means Point Cloud. In the *Array Type* Column, "dense" Means That All the Half-Wavelength Spacing Is Filled With Virtual Elements. In the *Other Sensors* Column, C, L, and O Stand for the Camera, Lidar, and Odometry, Respectively

Name	**Data Type**	**Array Type**	**Virtual Aperture (x×z $\frac{\lambda}{2}$ spacing)**	**Other Sensors**	**Record Time (Radar Frames)**	**Potential Gaps**
Zendar [33]	RDC / PC	Dense ULA	4x1	CLO	478s (4780)	No elevation, no ADC data, small aperture.
Radiate [34]	RA	No array mechanical scanning	N/A	CLO	5h (44000)	No elevation, no Doppler, no ADC data
CARRADA [35]	RA / RD	Dense ULA	8x1	C	12.1m (12666)	No elevation, no ADC data, small aperture
RADet [36]	RAD	Dense ULA	8x1	C	1015s (10158)	No elevation, no ADC data, small aperture
CRUW [14]	RA	Dense ULA	8x1	C	3.5h (400000)	No elevation, no Doppler, no ADC data, small aperture
Radical [37]	ADC	Dense ULA	8x1	C	104m (189000)	No elevation, small aperture
SCORP [9]	ADC / RAD	Dense ULA	8x1	C	N/A (3913)	No elevation, small aperture
ColoRadar [32]	ADC / RAE / PC	Sparse URA	86x7	LO	145m (43000)	No camera, mostly indoor
RADial [31]	ADC / RAD / PC	N/A	N/A	CLO	2h (25000)	Unknow array topology
K-Radar [10]	RAED	NUA	N/A	CLO	N/A (35000)	No ADC data
RaDelft (Ours)	ADC / RAED / PC	Sparse URA	86x7	CLO	35m (16975)	

with a commercially available radar. Our dataset contains three different levels of data processing, namely, ADC-level, radar cubes, and point clouds as defined in Section II-A, such that it can serve different future research directions. Additionally, the dataset contains synchronized data from camera, lidar, and odometry, recorded in real-world driving scenarios in the city of Delft. Additional details are provided in Section IV, specifically the sensors used and the developed radar signal processing pipeline.

B. Radar Detectors

The radar detection problem can be formulated as a binary decision task for each radar cube cell, whose objective is to determine whether there is a target or only noise in that specific cell. As mentioned in Section I, the automotive radar field has particular challenges when tackling the detection problem. First, the definition of clutter is not univocal in this application since targets of very different natures should be detected, including pedestrians, vehicles, bridges, potholes, road debris, and buildings, among others. Second, since modern automotive radars have high resolution in range, Doppler, and to some extent angle, targets occupy more than a single cell, behaving as extended targets. Finally, not only do the sizes of targets to be detected have a large variance, but also, for the same target, its perceived size can change over time. This is due to two different physical phenomena: the dependence of the angle estimation with its cosine with respect to the radar line of sight, and the relationship of the Cartesian size of the cell with the range due to the angle. Due to all these reasons, conventional CFAR detectors are expected to perform poorly in automotive radar data [7], [17].

In the past years, several works have been published on detecting extended targets in radar data. Image-based detector techniques have been explored in [38], but usually rely on high-contrast data where sharp transitions occur between noise and target. However, due to the finite length nature of signals, spectral leakage in the Fourier processing makes, in general, these transitions soft. Moreover, subspace detectors for extended targets in range and Doppler have been developed [39], [40], but still, an expected spread size of the target energy is needed, in addition to a high computational cost, making them unsuitable for real-time imaging automotive radars.

Also, DL techniques have been applied to the radar detection problem [5], [6], [6], [7], [8], [9], [11]. In [5], a DL detector is proposed, outperforming several 2-D cell-averaging (CA)-CFAR detectors, but only tested in simulated data. Lin et al. [8] and Gao et al. [41] propose a similar network structure using three autoencoders in three 2-D projections (range–angle, range–Doppler, and angle–Doppler) using the annotated dataset in [14] by a camera, avoiding full 3-D detection. However, using camera detections as ground truth may be limited due to the 2-D nature of camera images. Also, Zheng et al. [11] propose a DL-based detector using bird-eye view radar data, but focusing only on vehicle detection. On the other hand, Cheng et al. [7] and Paek et al. [6] propose two different NNs, but both use the lidar point cloud as ground truth. Since lidar provides high-resolution 3-D point clouds, it seems a more reasonable choice to serve as ground truth. The proposed method in [7] uses an NN to detect targets only in the range–Doppler dimensions, followed by the angle estimation and a spatiotemporal filter to enhance the resulting point cloud. On the other hand, in [6], a novel sparse approach to use an NN to detect in the range–azimuth–elevation space is presented. However, the Doppler information is collapsed into a single value, preventing the network from learning the possible angular estimation enhancement due to its relationship with

Doppler [24], [25]. Moreover, only the top 10% power cells are used as input to the network, and therefore, a predetection step is used, which can potentially remove target cells. This may be critical in automotive scenarios, where the angular sidelobes of close-range targets may be even 20 dB higher than weakly-reflecting distant targets such as pedestrians.

IV. RaDelft Dataset

The dataset was recorded with the demonstrator vehicle presented in [42] with an additional Texas Instrument MMWCAS-RF-EVM [43] imaging radar mounted on the roof at 1.5 m from the ground. The details of the radar and the waveform used are provided in Section IV-A. The collection was performed driving in multiple real-life scenarios in the city of Delft with different scene characteristics, such as suburban, university campus, and Delft old-town locations. Four different camera frames are shown in Fig. 2 to illustrate the differences in the environments. The output of the following sensors was recorded: a RoboSense Ruby Plus Lidar (128 layers rotating lidar, 10 Hz) and the imaging radar board installed on the roof, a video camera (1936 × 1216 pixels, ~30 Hz) mounted behind the windshield, and the ego vehicle's odometry (filtered combination of real time kinematics (RTK) GPS, inertial measurement unit (IMU), and wheel odometry, ~100 Hz). The sensor setup can be seen in Fig. 3. All sensors were jointly calibrated following [44] and time synchronized. With a 10-Hz frame rate, each scene contains around 2500 radar frames, adding to a total of 16 975 frames.

Example code for loading and visualizing the data is provided in a repository[1] to facilitate the use of the dataset, which can be downloaded from [19]. Moreover, the radar data are specifically provided at different processing stages for researchers with different backgrounds and interests, including ADC data, radar cubes, and point clouds. The details of the radar processing applied to the data can be found in Section IV-A.

A. Radar Configuration and Processing

In terms of the specific details of the radar system, this is the MIMO frequency modulated continuous wave (FMCW) evaluation board MMWCAS-RF-EVM from Texas Instruments, with 12 transmitters and 16 receivers [43]. The resulting virtual array is an 86-dense uniform linear array (ULA) in the X-direction (as shown in Fig. 3) with half-wavelength spacing, allowing azimuth estimation without grating lobes and a theoretical resolution of 1.33° looking at boresight. However, from the point of view of the 2-D angular estimation problem in both azimuth and elevation, the resulting uniform rectangular array (URA) is very sparse, with only a few minimum redundancy arrays (MRAs) in the Z-direction (as shown in Fig. 3). Thus, the elevation estimation is very poor in terms of both resolution and ambiguity. The details of the array topology can be found in [43] with graphical representations of the positions of all the elements. Moreover, some elements are overlapped, which can be used to address some of the problems introduced by using time-division multiple access (TDMA) in transmission, as detailed later in this section.

TABLE II

Radar Waveform Parameters Used in the Data Collection

Waveform Parameters	Value
Start Frequency (GHz)	76
Effective Bandwidth (MHz)	750
Chirp Slope (MHz/μs)	35
Chirps Length (μs)	28
Idle time (μs)	5
Number ADC Samples per Chirp	256
Number of Chirps per Frame	128
Sampling Frequency (Msps)	12
Tx strategy	TDMA
Derived Quantities	**Value**
Range Resolution (m)	0.2
Maximum Unambiguous Range (m)	51.4
Velocity Resolution (m/s)	0.046
Maximum Unambiguous Velocity (without extension) (m/s)	2.48
Maximum Unambiguous Velocity (with extension) (m/s)	17.36

The radar waveform parameters used can be seen in Table II, with the derived resolution and ambiguity values. The complex baseband samples are saved in the dataset using the same format provided by the radar manufacturer, but MATLAB code is provided to parse it, reshape it into an $N_{\text{fast}} \times N_{\text{slow}} \times N_{\text{Vchan}}$ 3-D tensor, and process it to the radar cubes.

The first step of the processing is to apply a Hamming windowing and an FFT in the fast-time and slow-time dimensions to perform range and Doppler estimations. Then, the detrimental effects of the TDMA have to be compensated. The first effect is related to the extension of the pulse repetition interval (PRI) by a factor equal to the number of transmitters. Therefore, the maximum unambiguous Doppler and the corresponding maximum measurable velocity (without ambiguity) v_{max} is reduced, as can be seen in the following equation:

$$v_{\text{max}} = \frac{c}{4 f_c \text{PRI}} \tag{1}$$

where c is the speed of light and f_c is the carrier frequency. This effect is especially problematic in the automotive context, where targets can have high relative speeds. Moreover, the phase difference between signals received from different transmitters will depend on both the angle of arrival of the signal and the velocity of the targets, due to the target's movement between transmission times of different transmitters operating in TDMA mode [45]. This resulting phase migration term is shown in the following equation:

$$\phi_{\text{mig}} = \frac{4\pi}{\lambda} v \Delta t \tag{2}$$

where λ is the wavelength, v is the relative speed of the target, and Δt is the time difference between transmitters. This term must be compensated before performing angle estimation to avoid significant artifacts.

In this work, both undesirable effects of TDMA are solved by using the overlapped virtual antennas present in the radar system with the algorithms provided in [46]. However, it is important to take into account that the maximum unambiguous velocity extension only works when a single target is present in a range–Doppler cell. Therefore, if multiple targets are folded

[1] https://github.com/RaDelft/RaDelft-Dataset

Fig. 2. Different frames of different scenes of the *RaDelft* dataset. As it can be seen, there are city center environments, suburban, and different road infrastructures such as large bridges.

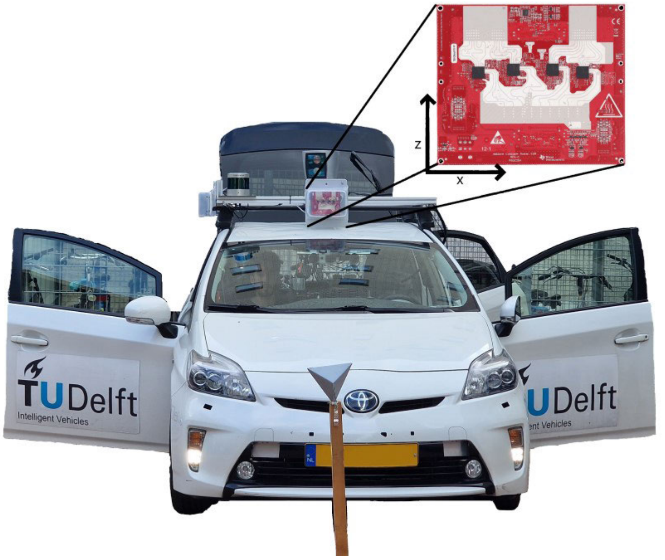

Fig. 3. Vehicle used to collect the dataset presented in this article, equipped with a high-resolution radar, lidar, camera, and odometry. The radar is shown in the top-right inset, with the defined X- and Z-coordinate axes assumed in this work.

into the same Doppler bin, or there are targets in different angles at the same range–Doppler bin, the algorithm will not be able to address the problem. Since this work does not aim to solve the Doppler ambiguity problem in TDMA, the aforementioned constraint is accepted as a limitation of the current commercial radar system. Nevertheless, it is assumed that making the ADC samples directly available in our dataset can be valuable for the research community, for example, to apply more advanced approaches for Doppler/velocity ambiguity in TDMA in the future.

The angle estimation can be performed once the TDMA effects have been compensated. It is important to remember that the resulting virtual array is a very sparse URA with some structures. While other research works deal with this type of array, for instance by trying to fill/interpolate the missing elements or applying compressive sensing techniques [22], [47], the core of this work is not to improve the angular estimation with sparse arrays. Therefore, a very simple approach of zero-filling and FFT processing has been applied. However, due to the sparseness of the radar antenna array in the Z-direction (as shown in Fig. 3), grating lobes and high sidelobes appear in elevation. To mitigate this problem, the FoV in elevation has been restricted to ±15°, and the elevation value with the highest power has been selected and saved, discarding the rest. Also, the azimuth estimation has been restricted to ±70° for two reasons. First, the angular estimation performance outside this region is rather poor, as

$$\Delta\theta \sim \frac{1}{\cos\theta} \tag{3}$$

being $\Delta\theta$ the angular resolution and θ the estimated angle. Second, the radiation power is almost 10 dB lower than at boresight outside this region, making target detection very challenging.

Subsequently, after zero padding, FFT processing, and FoV cropping, the resulting radar cubes have dimensions $N_r \times N_D \times N_a \times 2$ $(500 \times 128 \times 240 \times 2)$. This essentially means that for each range–Doppler–azimuth cell, there are two values: the elevation value with the highest detected power level, and the power level itself. Note that the 240 azimuth bins span the ±70° of the FoV after cropping, but not uniformly, due to the nonlinear relation in (3). For simplicity and to save storage space in the shared dataset, the aforementioned values are saved as different cubes since the elevation can be stored as an integer number (i.e., denoted as elevation bin), while the power value is a float.

Finally, a detection stage is applied to the radar cubes to generate a point cloud. This lower dimensionality representation of the data is also provided within the shared dataset to ease the process for researchers who want to use this highly processed data straightforwardly without going into the details of radar signal processing.

V. Proposed Data-Driven Detector

To address the aforementioned shortcomings of current detectors in automotive radar, a novel data-driven detector is proposed to generate 3-D occupancy grids only with radar data, using NNs and lidar data as ground truth. A visual summary of the method can be seen in Fig. 4.

The first step of the method is to adapt the lidar point cloud to serve as the ground truth. For each radar cube, the closest lidar point cloud in time is selected based on the timestamps for both radar and lidar data, assuming that a small error due to different start times may be present. Since the lidar system used in this work is mechanically rotating, it provides 360° coverage. Therefore, the first step is to crop this as to the same FoV of the radar, i.e., ±20° in elevation and ±70° in azimuth, and a maximum range of 50 m. To illustrate this difference in the FoV, Fig. 5(b) shows the cropped lidar point cloud compared to the original point cloud in Fig. 5(a). Moreover, removing all the lidar points from the road surface is essential as the road surface is hardly visible to the radars and could lead to noisy ground truth for the training process. The Patchwork++ algorithm is used to this end [48]. After

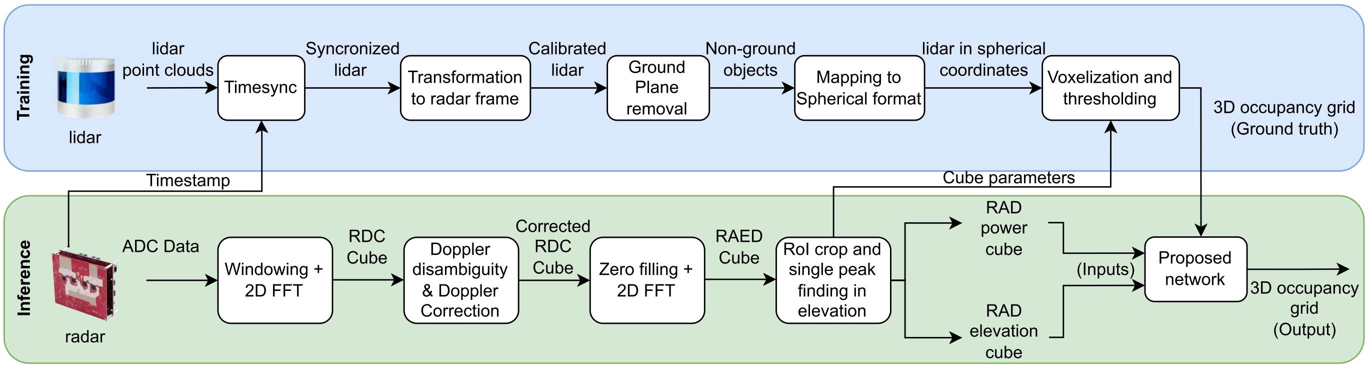

Fig. 4. Overview of the proposed data-driven detector. The steps to generate the 3-D lidar occupancy grid are shown on the top row, which will be then used as ground truth for training the NN. The radar signal processing pipeline is shown at the bottom of the figure and is needed to generate the input data for the network. RDC stands for range–Doppler–channel (no angle estimation), and RAED stands for range–azimuth–elevation–Doppler [18]. RoI stands the region of interest.

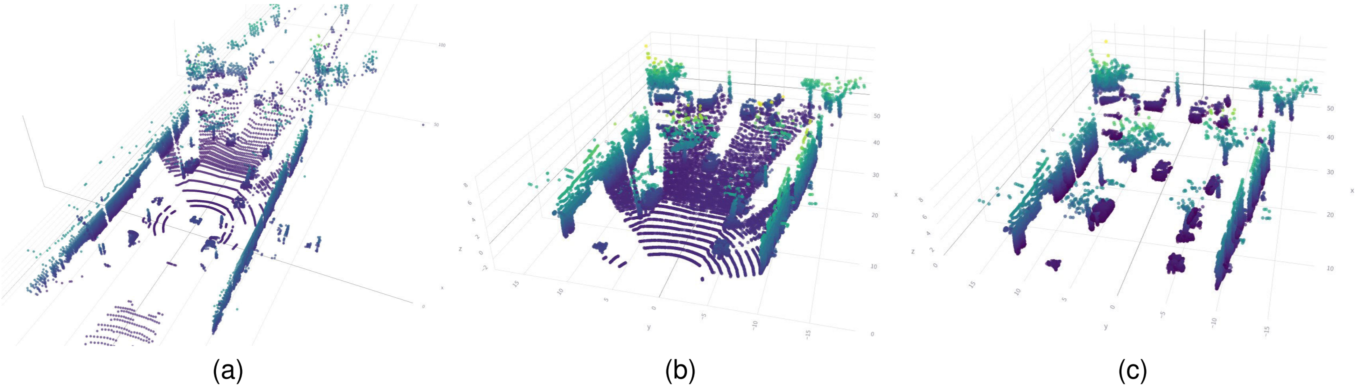

Fig. 5. (a) Original point cloud as provided by the lidar sensor. (b) Lidar point cloud after cropping to mimic the radar FoV (i.e., ±70° in azimuth and ±20° in elevation). (c) Lidar point cloud after the road surface removal using PatchWork++ [48], which will be used as ground truth to train the proposed data-driven detector.

removing the road surface points, the resulting lidar point cloud can be seen in Fig. 5(c).

Finally, the processed lidar point cloud has to be converted into a 3-D cube to serve as ground truth. This voxelization process can be understood as generating a 3-D occupancy grid, where each voxel contains “one” if at least one lidar point is inside, and “zero” otherwise. However, it is important to note that the radar cube grid is not uniform due to the Fourier transform processing for angular estimation and its relationship with the cosine of the estimated angle. This effect, which essentially makes the cells thinner at boresight and broader at the edge of the FoV, must be considered to generate the same nonuniform lidar 3-D occupancy grid. It is important to notice that all this process can be performed offline, outside the NN training loop, saving the processed lidar point clouds beforehand to speed up the training.

Once the ground truth has been appropriately generated as described above, the NN can be trained. The proposed NN is an evolution of the previous model validated in [18]. Specifically, in this case, the network is modified to use three frames of data as input to model temporal patterns, and the NN predicts the 3-D occupancy grid for the three frames simultaneously. This modification has been implemented to reduce the “flickering” usually present in the radar point clouds, where isolated points pass the detection threshold due to instantaneous high noise but disappear in consecutive frames. Therefore, the proposed NN tries to enforce some temporal consistency. A diagram of the complete network architecture is shown in Fig. 6. As it can be seen, the input is a $T \times 2 \times R \times A \times D$ tensor, where in practice, $T = 3$ (frames), $R = 500$ (range bins), $A = 240$ (azimuth bins), and $D = 128$ (Doppler bins). As explained in Section IV-A, these values are higher than the initial number of fast-time samples, slow-time samples, and virtual channels due to zero padding applied before the FFT processing. Moreover, the number of frames $T = 3$ has been chosen as a tradeoff between managing to capture temporal information and losing useful correlation between frames since the scene is often not static, and including too many frames will result in inconsistencies.

In terms of architecture, the first part of the proposed NN is the *DopplerEncoder* subnetwork. As the lidar cannot measure Doppler information, the detections on the Doppler dimension of the radar data cannot be directly utilized and compared to the ground truth. However, there is a known relationship between Doppler and angle in the case of moving platforms (or moving targets). Thus, the Doppler dimension is not simply removed from the radar data but rather encoded so that it can still be used in the overall detection process, as it may be beneficial for angular estimation. Specifically, here, the *DopplerEncoder* subnetwork extracts all the Doppler information in each range–azimuth cell and encodes it into the channel dimension. This is achieved by using two 3-D convolutional layers followed by a 3-D max pool layer, transforming the $2 \times R \times A \times D$ input tensor into a $64 \times R \times A$ tensor, where the 64-channel dimension contains the encoded information of Doppler and elevation.

The second part of the proposed network is an off-the-shelf 2-D CNN *backbone*, applied to estimate the final $R \times A \times E$ $(500 \times 240 \times 44)$ 3-D occupancy grid. The significant advantage of using such 2-D CNN backbones is their compatibility with hardware accelerators [e.g., GPUs and tensor processing units (TPUs)] and major machine learning frameworks (e.g., Tensor-Flow and PyTorch), leading to enhanced computational efficiency. While the current proposed implementation employs a feature pyramidal network (FPN) [49] with a Resnet18 backbone [50], our modular design allows for different architectures to be used for this purpose, enabling the system to be tailored to the specific memory and computational requirements of the intended platform.

These two parts of the proposed network are applied to each of the three considered frames independently, as shown in Fig. 6, but the weights of the layers are shared, and the output is concatenated into a $T \times R \times A \times E$ $(3 \times 500 \times 240 \times 44)$ tensor. Finally, to take into account the temporal relationship between the three frames, a third module composed of six 3-D convolutional layers is included (referred to as *TemporalCoherenceNetwork* in Fig. 6). It is important to notice that even if the output is a 3-D occupancy grid for each frame, the power information on each cell is not lost since the indices of the detected cells from such grid can be used to retrieve the corresponding intensity information from the original RAED cube.

One of the key characteristics of the radar data is the scene sparsity. Of all the voxels in the generated 3-D occupancy grid, only around 1% contain targets. Therefore, this must be considered when selecting the loss function for training the NN. In this work, the Focal loss [51] is used for this purpose, which handles class imbalances in a similar way to the weighted cross-entropy loss and adds an extra modulating factor to focus on the hard cases. The Focal loss [51] is defined as

$$\mathrm{FL}(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \tag{4}$$

with

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \tag{5}$$

where $y \in \{\pm 1\}$ is the ground-truth class (i.e., detection or not), α_t is the weighting factor to take into account data imbalance defined as $\alpha \in [0, 1]$ for class 1 and $1-\alpha$ for class -1, and $\gamma > 1$ is the focusing factor. This loss is especially interesting in radar data since high radar cross section (RCS) targets can be easily detected, but low RCS targets or targets located at a far distance are more challenging to detect, and this can be taken into account by the γ parameter. In terms of training–testing split, 90% of the data from five of the seven recorded scenarios have been used to train the network using Adam optimizer, leaving 10% for validation. The network was trained using the DelftBlue Supercomputer [52] from TU Delft. The remaining two recorded scenarios are used as a test set, i.e., with data completely new, unseen for the network.

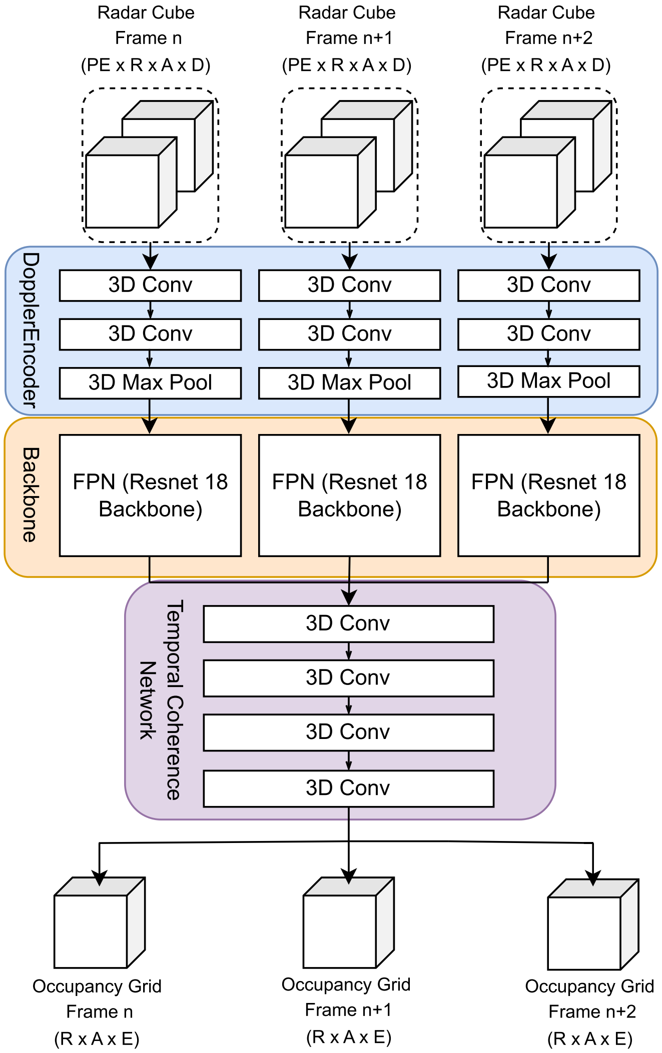

Fig. 6. Proposed network architecture for the data-driven detector composed of three subnetworks. First, the *DopplerEncoder* network aims to encode the Doppler information so that its information is retained even if not directly comparable with ground-truth lidar data. Then, a standard FPN with a Resnet *backbone* is used. Note that the three branches process separately three frames of data but share the same weights. Finally, the three outputs are concatenated to produce an input tensor to the *temporal coherence network*, which generates the final occupancy grid for each frame.

VI. Results

The trained NN can estimate the 3-D occupancy grid for each radar cube and thus act as a detector. It should be noted that all the results presented in this section are evaluated using only the test set, composed of the two scenes left out from the training process. This ensures data independence and, while still collected in the same geographical area, the capability of the proposed method to generalize to unseen data with different characteristics.

Two main performance metrics are used to evaluate the results of the proposed NN: the usual probability of detection (P_d) and probability of false alarm (P_{fa}) metrics, and the Chamfer distance (CD). In both cases, the lidar data are used as a reference, either in the occupancy grid format for the P_d and P_{fa} computation, or in the point cloud format for the

CD. While different definitions are given for the CD in the literature, in this work, the following is used:

$$\mathrm{CD}(S_1, S_2) = \frac{1}{|S_1|}\sum_{x\in S_1}\min_{y\in S_2}||x-y||_2 + \frac{1}{|S_2|}\sum_{y\in S_2}\min_{x\in S_1}||x-y||_2 \quad (6)$$

where S_1 and S_2 are the two sets of points being compared (e.g., the lidar points assumed as ground truth versus the points from the 3-D occupancy grid provided by the proposed data-driven detector), and $|S|$ is the cardinality of the set. The closer the two sets of points are the better, and so the lower the CD.

However, it is important to note that a caveat is needed when analyzing the P_d and the P_{fa} metrics. A small misalignment in the calibration of only a few centimeters in range or of a small angle will cause the probability of detection to fall drastically, while the probability of false alarms will rise, as can be seen in the examples presented in Fig. 7. For instance, considering the example in Fig. 7(b), even though the P_{fa} of this case is numerically the same as in the case represented in Fig. 7(a), the impact in terms of quality of the perceived environment can be very different, especially taking into account the small cell dimensions. While this is not a problem in the proposed method (as the network used as the data-driven detector can learn offsets such as those in this example), it may affect the other methods used in this section for benchmarking, such as different variants of CFAR detectors. Moreover, since the radar resolutions are worse than the lidar's, many targets will be overestimated in size, raising the P_{fa}. These false alarms are, in general, assumed to be less relevant for assessing the quality of automotive radar since a small overestimation of objects in the order of centimeters (i.e., few lidar resolution cells) may not be as bad as detecting isolated ghost targets. Nevertheless, all the false alarms are treated equally in the assessment in this article since an extra clustering or tracking stage may be needed to distinguish between these unfavorable cases in terms of P_{fa}. An example of this phenomenon can be seen in Fig. 7(c). On the other hand, it can be seen how the CD can capture these spatial relationships, yielding different values for the three different cases. Taking all this into account, a point cloud-level metric like the CD is considered to be a better evaluation metric for this work.

Table III shows the performance of the proposed method with the three aforementioned metrics averaged over the whole test set and compared with different alternative approaches for detection. Specifically, different rows on Table III are as follows.

1) *Proposed Method:* The results of the proposed method explained in Section V and with the overall architecture shown in Fig. 6.
2) *No Doppler and Quantile:* An approach similar to the one presented in [6], where only those power cells with values higher than the 0.9 quantile are kept and the rest are set to zero. Furthermore, the Doppler information is collapsed by taking the mean over the Doppler dimension. This is used to "sparsify" the data and speed up processing, with the risk of cutting out weakly reflecting targets. Since there is no Doppler data anymore, the *DopplerEncoder* subnetwork is removed from the general architecture of the proposed data-driven detector. It is important to note that the segmentation backbone has 13.2 million parameters, while the *DopplerEncoder* subnetwork is only 76.9k. Therefore, the comparison between the full network and the network without the *DopplerEncoder* is possible without adding extra layers.
3) *Quantile:* The proposed method, but with the predetection fixed threshold based on the 0.9 quantile inspired by Paek et al. [6].
4) *No Time:* In order to assess the impact of inputting several frames into the network and use the temporal evolution of the scene, this tests the proposed method without the *Temporal Coherence* subnetwork in the architecture, essentially an ablation study without interframe temporal information.
5) *OS-CFAR:* A 2-D ordered-statistics (OS)-CFAR in range–angle, followed by a 1-D OS-CFAR in Doppler. While multiple different CFAR alternatives have been tested (i.e., different combinations of CA and OS-CFAR detectors), only the best implementation is reported here for conciseness. An analysis with different variations has been presented in [18] for completeness. Following [53], the rank has been set to 0.75 times the number of training cells, and no guard cells have been used.

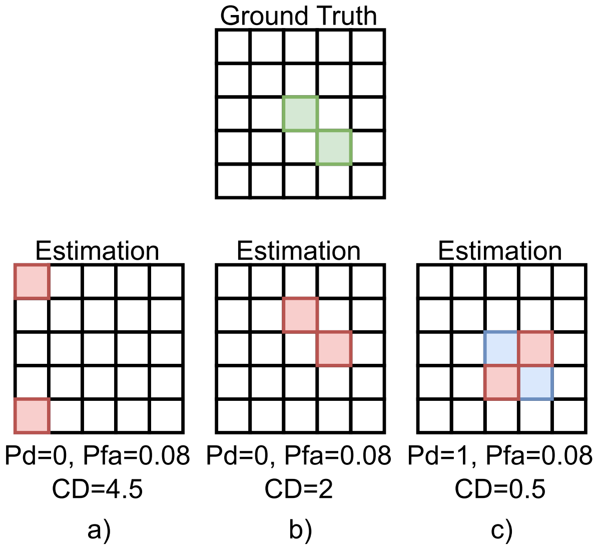

Fig. 7. Illustration of the problem in computing the P_d, P_{fa}, and CD as performance metrics. (a) Case where two ghost targets are created. (b) Calibration misalignment shifts the detection cells, raising the P_{fa} as if two ghost targets were created. (c) Problem of the overestimation of target size. These three cases have nominally the same P_{fa}, but the implications for overall scene perception are completely different. It can be seen how the CD captures the spatial relationships and yields a better value in (b) and (c), where the false alarms have less impact from an application point of view.

As it can be seen in Table III, the highest P_d and the lowest CD are achieved by the proposed method while maintaining a similar P_{fa}. On the other hand, applying the quantile cut and removing the Doppler information similar to [6] reduce the P_d from 62.13% to 52.97% and worsen the CD

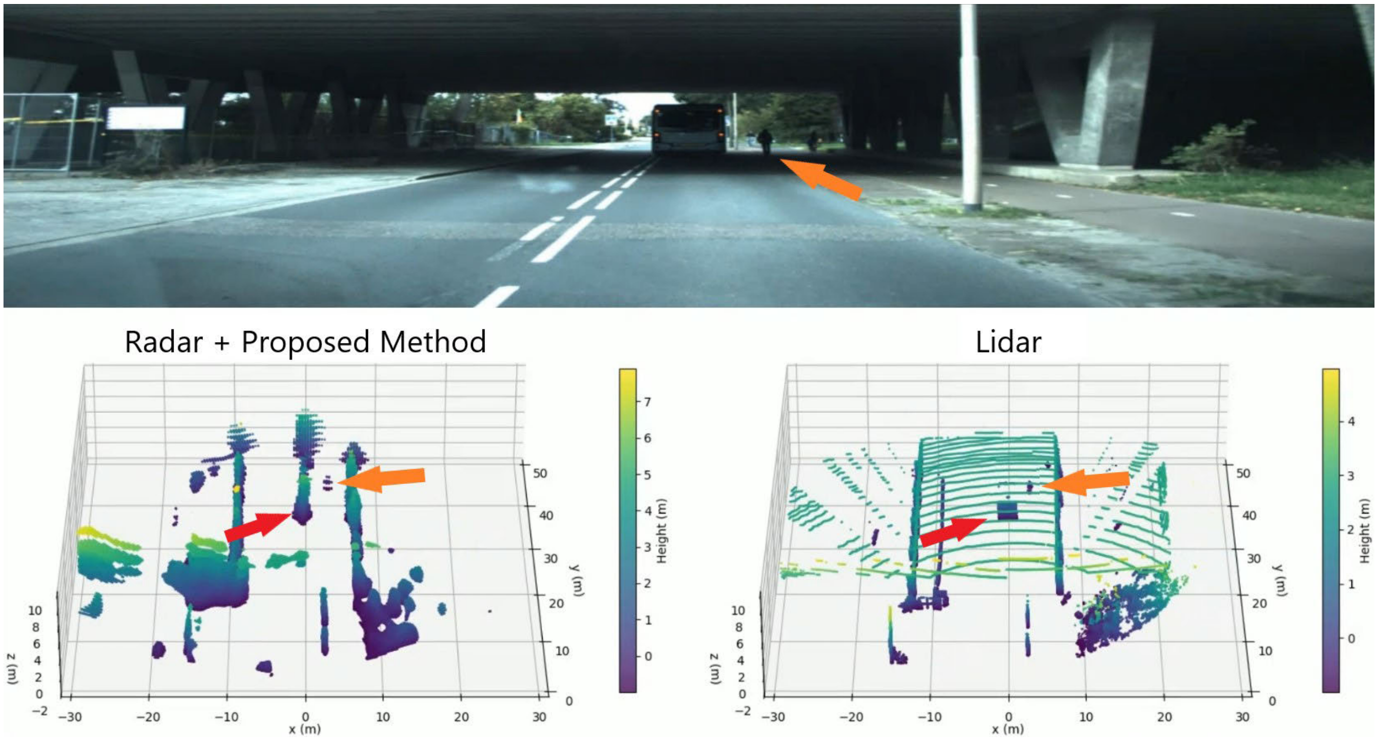

Fig. 8. Example frame in a challenging situation for the radar system, where the vehicle is going under a large bridge. In the top figure, the camera image is shown for reference. On the left, the point cloud generated with the proposed data-driven method is shown, and on the right, the original point cloud provided by the lidar. The red arrows point to the bus under the bridge and the orange arrow points to the pedestrian next to it. Note that the color in the point clouds refers to the height of the objects.

from 1.54 to 2.16 m. Looking at the results for the other versions, it can be seen that this drop in performance is mostly due to the removal of the Doppler information. Using only the quantile-based threshold may be a good tradeoff since the performance degradation is not substantial, but the computational cost is reduced. Looking at the version without the *Temporal Coherence* subnetwork, which is trained on single frames, it can be seen how all the metrics are worse than in the baseline. Thus, including temporal information in the network is a good strategy to boost performance, with the only downside of increasing slightly the training time due to the extra layers. Finally, it can be seen that the conventional OS-CFAR is the method that performs the worst, with a much higher CD of 6.73 m.

In order to have a fairer comparison against the conventional CFAR detector, a 2-D version of the proposed method has also been evaluated by disregarding the elevation information, as this can only be estimated rather poorly due to the unfavorable design of the radar array. To this end, the proposed NN has been trained without elevation information, discarding the virtual channels in the Z-direction and, thus, treating it as a ULA in the azimuth direction. For completeness, the implementation with a quantile-based threshold has also been assessed in this new analysis. The results are shown in Table III under the “No Elevation cases” rows. For these tests, the P_d of the OS-CFAR approach is increased to 11.5%, but the P_{fa} is also raised. This is mainly due to detections triggered in the adjacent angle bins of a target generating “ring like” patterns due to sidelobes, a phenomenon also mentioned in [6]. Both the proposed method and the proposed method with the quantile-based threshold are shown to outperform the conventional OS-CFAR in the three metrics.

TABLE III
PERFORMANCE RESULTS OF THE PROPOSED METHOD FOR DATA-DRIVEN DETECTION, DIFFERENT VARIATIONS OF THE METHOD, AND THE BEST-PERFORMING CFAR DETECTOR IMPLEMENTED

Method	P_d (%)	P_{fa} (%)	**Chamfer distance** (m)
Proposed Method	62.13	2.77	1.54
No Doppler & Quantile	52.97	2.63	2.16
No Doppler	50.44	2.50	2.13
Quantile	57.9	2.85	1.92
No Time (single frame)	58.08	2.63	2.16
OSCFAR	0.41	0.015	6.73
No Elevation cases			
Proposed Method	74.83	1.12	2.92
Quantile	74.09	1.11	2.78
OSCFAR	11.56	3.1	4.11

In addition to the quantitative results, some qualitative results are also presented to show the performance of the proposed method visually. In Fig. 8, a challenging frame from the radar point of view is shown, where the vehicle is going under a large but relatively not tall bridge. The 3-D point cloud generated with the proposed method is shown in the left plot, with the original lidar on the right plot. As it can be seen, the road is clear of false alarms, and the bus (in red arrow) and pedestrian (in orange arrow) are clearly detected. The bus and the ceiling merge due to the poor elevation resolution of the radar data, but they could be split and identified in Doppler.

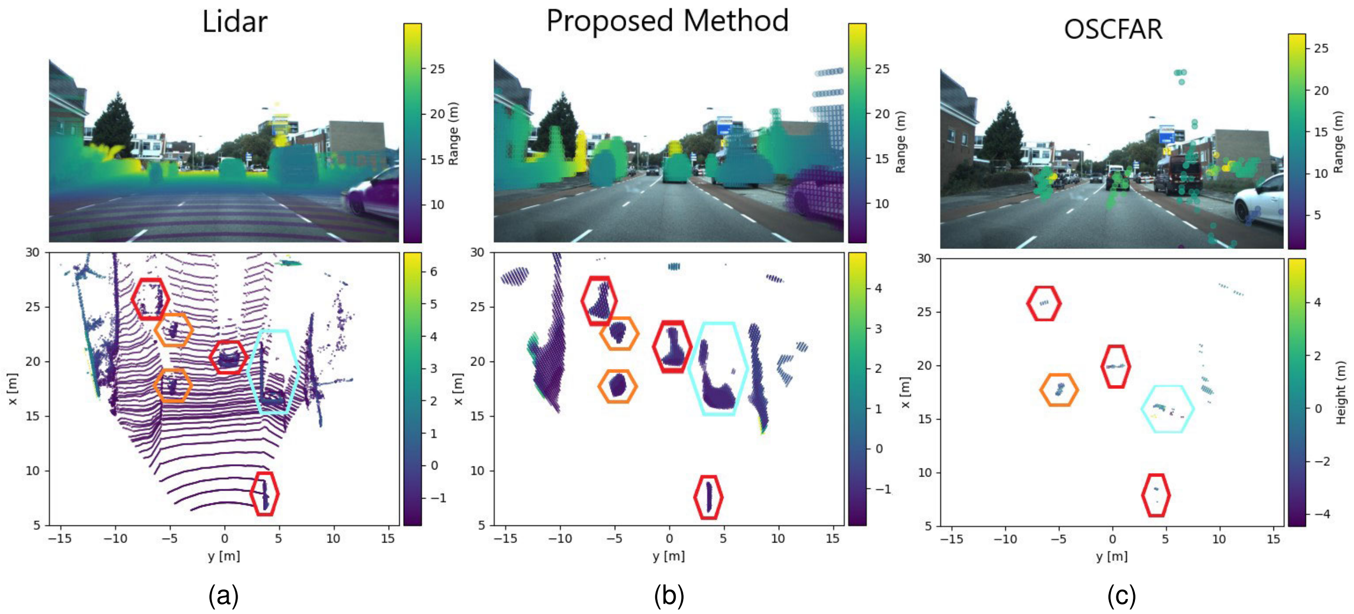

Fig. 9. Example of the data frame in the urban scenario with related detections. (a) Original lidar point cloud projected onto the camera as well as a bird's eye view. (b) Radar point cloud generated with the proposed data-driven detector. (c) Radar point cloud generated with the best-performing CFAR implemented (i.e., 2-D OS-CFAR in range–azimuth, followed by an OS-CFAR in Doppler).

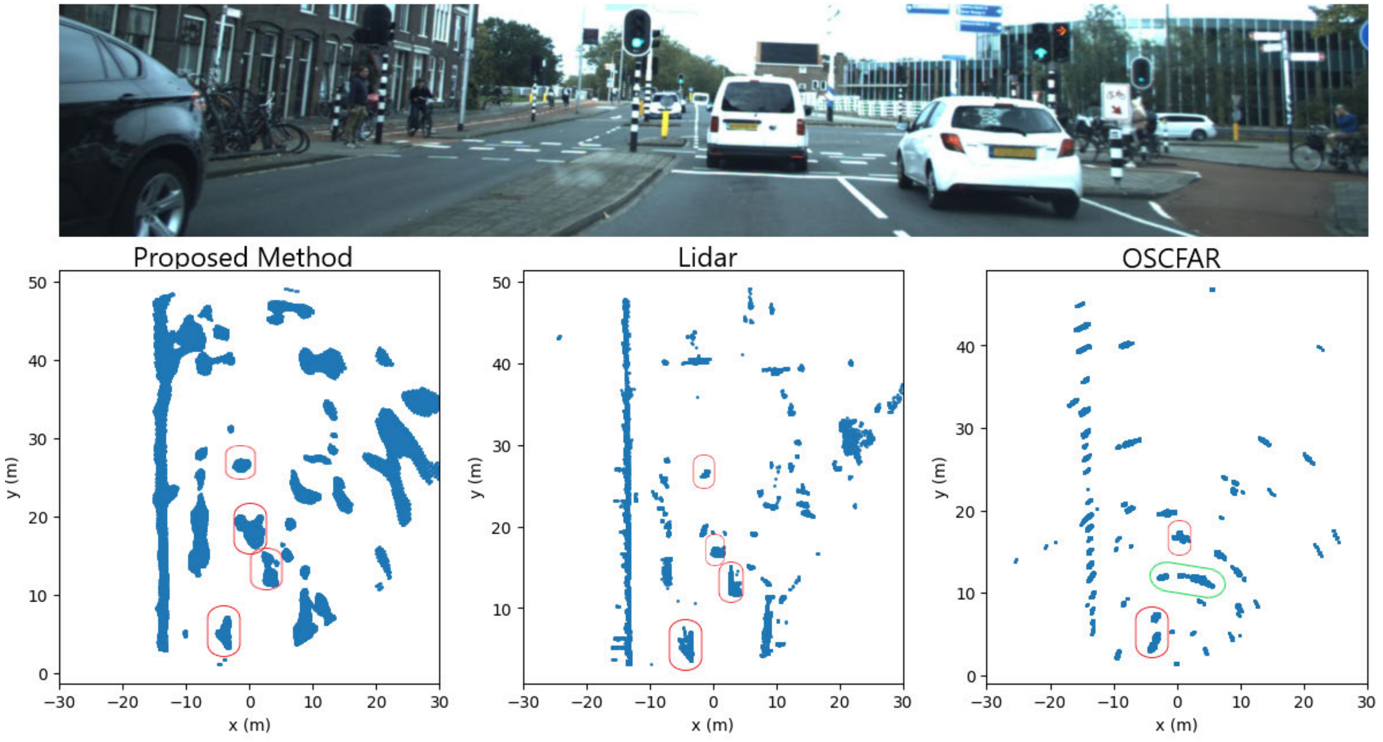

Fig. 10. Example of data frame where the elevation information is disregarded from the detection process. In the top figure, the camera image is shown for reference. In the bottom part of the figure, the original lidar point cloud is shown (center), with the point cloud generated by the proposed data-driven detector (left) and by the best-performing implemented CFAR (right).

Fig. 9 shows another scene where the resulting point clouds have been projected onto the camera image to provide a sense of the 3-D scene (top), but the bird's eye view projection is also shown (bottom). For simplicity, the point clouds have been cropped to a maximum range of 30 m. Moreover, as a visual aid in the bird's eye view, cyclists are highlighted with an orange hexagon, cars with a red hexagon, and a large van with a light blue hexagon. In Fig. 9(a), the original lidar point cloud is presented, where many details of the scene can be appreciated. Fig. 9(b) shows the detections generated using the proposed data-driven detector, and as it can be seen, most of the details of the relevant targets are preserved. Objects are slightly overestimated in size, but the overall scene is clear. Also, the shape of the objects is preserved, especially in the case of cars and large vans. Finally, Fig. 9(c) shows the output of the previously-mentioned best-performing CFAR detector, where it can be seen how the output is much sparser in terms of detected points, and also missing one of the cyclists in the scene.

Finally, an example of results where the elevation information is disregarded in the detection process is presented in Fig. 10. Fig. 10 shows the camera image for visual reference (top), and the comparison of the resulting point cloud from the radar data with the proposed data-driven detector (left), the original lidar data (center), and the point cloud from the radar data with the best-performing implemented CFAR. Note that cars are highlighted in red, and there are "ring-like" detections (highlighted in green) due to the high sidelobes of the van, which can be seen in Fig. 10 generated using the CFAR detector. This phenomenon raises the P_{fa} and is an expected behavior that has been reported in other automotive radar datasets [6] when using CFAR detectors. As also reported in the previous qualitative examples, the point cloud generated by the proposed data-driven detector is denser than the CFAR-generated one and conserves the correct location and shape of most objects.

VII. Conclusion

This work introduces an innovative data-driven detector for automotive radar and the *RaDelft* dataset, a newly collected multisensor real-world dataset. The proposed radar detector is trained exclusively from unlabeled synchronized radar and lidar data, thus eliminating the need for costly manual object annotations for the detection process. Two types of performance metrics were employed to validate the method, i.e., conventional probability of detection and probability of false alarm, alongside the CD, a point cloud-level metric designed to capture spatial relationships and similarities between point clouds. The proposed method reduces by 4.2 m (77% reduction) the CD when compared with conventional OS-CFAR detectors, and by 0.62 m (28% reduction) when compared with the state of the art. Also, it significantly increases the probability of detection. Moreover, an ablation study showed that including temporal information in the process is important, and Doppler information is especially crucial for our model's good performance. Results show that the probability of detection is increased from 50.44% to 62.13%, and the CD is reduced by 27% when using Doppler information.

For the experimental evaluation of the proposed approach, a comprehensive dataset encompassing over 30 min of actual driving scenarios was collected using a vehicle equipped with both lidar and radar sensors, resulting in 16 975 radar frames paired with corresponding lidar ground truth. Compared with other existing datasets, *RaDelft* provides raw data from a commercial 4-D imaging radar needed for radar practitioners for many research lines. Moreover, it contains data processed at other levels (e.g., radar cubes and point clouds) suitable for researchers with different backgrounds and interests. The dataset is publicly available, with code to parse, visualize, and process the data, as well as the code to reproduce the results reported in this work.

References

[1] F. Sezgin, D. Vriesman, D. Steinhauser, R. Lugner, and T. Brandmeier, "Safe autonomous driving in adverse weather: Sensor evaluation and performance monitoring," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2023, pp. 1–6.

[2] I. Bilik, O. Longman, S. Villeval, and J. Tabrikian, "The rise of radar for autonomous vehicles: Signal processing solutions and future research directions," *IEEE Signal Process. Mag.*, vol. 36, no. 5, pp. 20–31, Sep. 2019.

[3] 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.

[4] A. Srivastav and S. Mandal, "Radars for autonomous driving: A review of deep learning methods and challenges," *IEEE Access*, vol. 11, pp. 97147–97168, 2023.

[5] D. Brodeski, I. Bilik, and R. Giryes, "Deep radar detector," in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2019, pp. 1–6.

[6] D.-H. Paek, S.-H. Kong, and K. T. Wijaya, "K-radar: 4D radar object detection for autonomous driving in various weather conditions," in *Proc. 36th Conf. Neural Inf. Process. Syst. Datasets Benchmarks Track*, 2022, pp. 3819–3829. [Online]. Available: https://openreview.net/forum?id=W_bsDmzwaZ7

[7] Y. Cheng, J. Su, M. Jiang, and Y. Liu, "A novel radar point cloud generation method for robot environment perception," *IEEE Trans. Robot.*, vol. 38, no. 6, pp. 3754–3773, Dec. 2022.

[8] Y. Lin, X. Wei, Z. Zou, and W. Yi, "Deep learning based target detection method for the range-Azimuth-Doppler cube of automotive radar," in *Proc. IEEE 26th Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2023, pp. 2868–2873.

[9] F. E. Nowruzi et al., "Deep open space segmentation using automotive radar," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Nov. 2020, pp. 1–4.

[10] D. Gusland, S. Rolfsjord, and B. Torvik, "Deep temporal detection—A machine learning approach to multiple-dwell target detection," in *Proc. IEEE Int. Radar Conf. (RADAR)*, Apr. 2020, pp. 203–207.

[11] R. Zheng, S. Sun, H. Liu, and T. Wu, "Deep-neural-network-enabled vehicle detection using high-resolution automotive radar imaging," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 5, pp. 4815–4830, Oct. 2023.

[12] A. Palffy, J. Dong, J. F. P. Kooij, and D. M. Gavrila, "CNN based road user detection using the 3D radar cube," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1263–1270, Apr. 2020.

[13] O. Schumann, J. Lombacher, M. Hahn, C. Wöhler, and J. Dickmann, "Scene understanding with automotive radar," *IEEE Trans. Intell. Vehicles*, vol. 5, no. 2, pp. 188–203, Jun. 2020.

[14] Y. Wang, Z. Jiang, Y. Li, J.-N. Hwang, G. Xing, and H. Liu, "RODNet: A real-time radar object detection network cross-supervised by camera-radar fused object 3D localization," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 4, pp. 954–967, Jun. 2021.

[15] I. Roldan, F. Fioranelli, and A. Yarovoy, "Self-supervised learning for enhancing angular resolution in automotive MIMO radars," *IEEE Trans. Veh. Technol.*, vol. 72, no. 9, pp. 11505–11514, Sep. 2023.

[16] M. A. Richards, J. A. Scheer, and W. A. Holm, *Principles of Modern Radar: Basic Principles*, vol. 1. Rijeka, Croatia: SciTech, 2010.

[17] J. Yoon, S. Lee, S. Lim, and S.-C. Kim, "High-density clutter recognition and suppression for automotive radar systems," *IEEE Access*, vol. 7, pp. 58368–58380, 2019.

[18] I. Roldan, A. Palffy, J. F. P. Kooij, D. M. Gavrila, F. Fioranelli, and A. Yarovoy, "See further than CFAR: A data-driven radar detector trained by LiDAR," in *Proc. IEEE Radar Conf.*, Denver, CO, USA, May 2024, pp. 1–6.

[19] I. Roldan et al. (2024). *RaDelft Dataset: A Large-scale, Real-Life, and Multi-Sensor Automotive Dataset*. [Online]. Available: https://data.4tu.nl/datasets/4e277430-e562-4a7a-adfe-30b58d9a5f0a/1

[20] Y. Jin, M. Hoffmann, A. Deligiannis, J.-C. Fuentes-Michel, and M. Vossiek, "Semantic segmentation-based occupancy grid map learning with automotive radar raw data," *IEEE Trans. Intell. Vehicles*, vol. 9, no. 1, pp. 216–230, Jan. 2024.

[21] L. Xu, J. Lien, and J. Li, "Doppler–range processing for enhanced high-speed moving target detection using LFMCW automotive radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 1, pp. 568–580, Feb. 2022.

[22] M. Rossi, A. M. Haimovich, and Y. C. Eldar, "Spatial compressive sensing for MIMO radar," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 419–430, Jan. 2014.

[23] I. Roldan, F. Fioranelli, and A. Yarovoy, "Total variation compressive sensing for 3D shape estimation in short-range imaging radars," *IEEE Trans. Radar Syst.*, vol. 1, pp. 583–592, 2023.

[24] W. Zhang, P. Wang, N. He, and Z. He, "Super resolution DOA based on relative motion for FMCW automotive radar," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8698–8709, Aug. 2020.

[25] S. Yuan, F. Fioranelli, and A. G. Yarovoy, "3DRUDAT: 3D robust unambiguous Doppler beam sharpening using adaptive threshold for forward-looking region," *IEEE Trans. Radar Syst.*, vol. 2, pp. 138–153, 2024.

[26] J. Fuchs, M. Gardill, M. Lübke, A. Dubey, and F. Lurz, "A machine learning perspective on automotive radar direction of arrival estimation," *IEEE Access*, vol. 10, pp. 6775–6797, 2022.

[27] I. Bilik et al., "Automotive multi-mode cascaded radar data processing embedded system," in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2018, pp. 372–376.

[28] D. Kellner, M. Barjenbruch, J. Klappstein, J. Dickmann, and K. Dietmayer, "Tracking of extended objects with high-resolution Doppler radar," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 5, pp. 1341–1353, May 2016.

[29] M. Hassan, F. Fioranelli, A. Yarovoy, and S. Ravindran, "Radar multi object tracking using DNN features," in *Proc. IEEE Int. Radar Conf. (RADAR)*, Nov. 2023, pp. 1–6.

[30] 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, p. 4208, May 2022.

[31] 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. (CVPR)*, Jun. 2022, pp. 17000–17009.

[32] A. Kramer, K. Harlow, C. Williams, and C. Heckman, "ColoRadar: The direct 3D millimeter wave radar dataset," *Int. J. Robot. Res.*, vol. 41, no. 4, pp. 351–360, Apr. 2022.

[33] 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 (CVPRW)*, Jun. 2020, pp. 450–457.

[34] M. Sheeny, E. De Pellegrin, S. Mukherjee, A. Ahrabian, S. Wang, and A. Wallace, "RADIATE: A radar dataset for automotive perception in bad weather," 2020, *arXiv:2010.09076*.

[35] 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. (ICPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jan. 2021, pp. 5068–5075.

[36] 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. (CRV)*, May 2021, pp. 95–102.

[37] T. Lim, S. A. Markowitz, and M. N. Do, "RaDICaL: A synchronized FMCW radar, depth, IMU and RGB camera data dataset with low-level FMCW radar signals," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 4, pp. 941–953, Jun. 2021.

[38] J. Yang, J. Yi, T. Sakamoto, and X. Wan, "An extended target detector using image-processing techniques exploiting energy-spillover phenomenon in radar echoes," *IEEE Sensors J.*, vol. 23, no. 19, pp. 22919–22929, Oct. 2023.

[39] J. Guan and X. Zhang, "Subspace detection for range and Doppler distributed targets with rao and wald tests," *Signal Process.*, vol. 91, no. 1, pp. 51–60, Jan. 2011.

[40] J. Carretero-Moya, J. Gismero-Menoyo, A. Asensio-Lopez, and A. Blanco-Del-Campo, "Small-target detection in high-resolution heterogeneous sea-clutter: An empirical analysis," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 3, pp. 1880–1898, Jul. 2011.

[41] X. Gao, G. Xing, S. Roy, and H. Liu, "RAMP-CNN: A novel neural network for enhanced automotive radar object recognition," *IEEE Sensors J.*, vol. 21, no. 4, pp. 5119–5132, Feb. 2021, doi: 10.1109/JSEN.2020.3036047.

[42] A. Palffy, E. Pool, S. Baratam, J. F. P. Kooij, and D. M. 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.

[43] Texas Instruments. (2019). *Design Guide: TIDEP-01012-imaging Radar Using Cascaded mmWave Sensor Reference Design (Rev. A)*. [Online]. Available: https://www.ti.com/lit/ug/tiduen5a/tiduen5a.pdf

[44] J. Domhof, J. F. P. Kooij, and D. M. Gavrila, "A joint extrinsic calibration tool for radar, camera and LiDAR," *IEEE Trans. Intell. Vehicles*, vol. 6, no. 3, pp. 571–582, Sep. 2021.

[45] D. Zoeke and A. Ziroff, "Phase migration effects in moving target localization using switched MIMO arrays," in *Proc. Eur. Radar Conf. (EuRAD)*, Sep. 2015, pp. 85–88.

[46] 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. (EUCAP)*, Mar. 2012, pp. 1746–1750.

[47] 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.

[48] S. Lee, H. Lim, and H. Myung, "Patchwork++: Fast and robust ground segmentation solving partial under-segmentation using 3D point cloud," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2022, pp. 13276–13283.

[49] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.

[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.

[51] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," 2017, *arXiv:1708.02002*.

[52] Delft High Performance Computing Centre (DHPC). (2024). *DelftBlue Supercomputer (Phase 2)*. [Online]. Available: https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase2

[53] H. Rohling, "Radar CFAR thresholding in clutter and multiple target situations," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-19, no. 4, pp. 608–621, Jul. 1983.

Ignacio Roldan (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunication engineering and the M.Sc. degree in signal processing and machine learning from the Universidad Politécnica de Madrid, Madrid, Spain, in 2014, 2016, and 2018, respectively. He is currently pursuing the Ph.D. degree with the Microwave Sensing, Signals and Systems Group, Delft University of Technology (TU Delft), Delft, The Netherlands.

He has worked for more than five years with Advanced Radar Technologies, Madrid, a Spanish tech company focused on the design and manufacture of radar systems. During this period, he has been involved in several international projects developing state-of-the-art signal processing techniques for radars. In his last stage, he was focused on applying machine learning techniques to unmanned aerial vehicle (UAV) detection and classification. In September 2020, he joined the Microwave Sensing, Signals and Systems Group, TU Delft.

Mr. Roldan received the best student paper award at the 2024 IEEE Radar Conference held in Denver, USA, for his work on automotive radar target detection using neural networks.

Andras Palffy (Member, IEEE) received the M.Sc. degree in computer science engineering from Pazmany Peter Catholic University, Budapest, Hungary, in 2016, the M.Sc. degree in digital signal and image processing from Cranfield University, Cranfield, U.K., in 2015, and the Ph.D. degree from Delft University of Technology, Delft, The Netherlands, in 2022, focusing on radar-based vulnerable road user detection for automated driving.

From 2013 to 2017, he was an Algorithm Researcher with Eutecus, Budapest, a U.S.-based start-up developing computer vision algorithms for traffic monitoring and driver assistance applications. In 2022, he co-founded Perciv AI, Delft, a machine perception start-up developing AI-driven, next-generation machine perception for radars.

Julian F. P. Kooij (Member, IEEE) received the Ph.D. degree in artificial intelligence from the University of Amsterdam, Amsterdam, The Netherlands, in 2015.

In 2013, he joined Daimler AG, Ulm, Germany, and worked on path prediction for vulnerable road users. In 2014, he joined the Computer Vision Laboratory, Delft University of Technology (TU Delft), Delft, The Netherlands. Since 2016, he has been with the Intelligent Vehicles Group, part of the Cognitive Robotics Department, TU Delft, where he is currently an Associate Professor. His research interests include probabilistic models and machine learning techniques to infer and anticipate critical traffic situations from multimodal sensor data.

Dariu M. Gavrila (Member, IEEE) received the Ph.D. degree in computer science from the University of Maryland, College Park, MD, USA, in 1996.

In 1997, he joined Daimler R&D, Ulm, Germany, where he became a Distinguished Scientist. In 2016, he moved to Delft University of Technology, Delft, The Netherlands, where he is currently a Full Professor with the Intelligent Vehicles Group. His research interests include sensor-based detection of humans and analysis of behavior, recently in the context of self-driving cars in urban traffic.

Dr. Gavrila was a recipient of the Outstanding Application Award in 2014 and the Outstanding Researcher Award in 2019 from the IEEE Intelligent Transportation Systems Society.

Francesco Fioranelli (Senior Member, IEEE) received the Ph.D. degree from Durham University, Durham, U.K., in 2014.

He was a Research Associate with University College London, London, U.K., from 2014 to 2016, and an Assistant Professor with the University of Glasgow, Glasgow, U.K., from 2016 to 2019. He is currently an Associate Professor with Delft University of Technology (TU Delft), Delft, The Netherlands. He has authored over 190 peer-reviewed publications and edited the books on "Micro-Doppler Radar and Its Applications" and "Radar Countermeasures for Unmanned Aerial Vehicles" (IET-Scitech, 2020). His research interests include the development of radar systems and automatic classification for human signatures analysis in healthcare and security, drones and unmanned aerial vehicle (UAV) detection and classification, and automotive radar.

Dr. Fioranelli received four best paper awards and the IEEE AESS Fred Nathanson Memorial Radar Award in 2024.

Alexander Yarovoy (Fellow, IEEE) received the Diploma degree (Hons.) in radiophysics and electronics and the Candidate Phys. and Math. Sci. and Doctor Phys. and Math. Sci. degrees in radiophysics from Kharkov State University, Kharkiv, Ukraine, in 1984, 1987, and 1994, respectively.

In 1987, he joined the Department of Radiophysics, Kharkov State University, as a Researcher and became a Full Professor in 1997. From September 1994 to 1996, he was a Visiting Researcher with the Technical University of Ilmenau, Ilmenau, Germany. Since 1999, he has been with Delft University of Technology, Delft, The Netherlands, where he has been the Chair of the Microwave Sensing, Signals and Systems (MS3) Group, since 2009. He has authored or co-authored more than 600 scientific or technical articles and 14 book chapters, and holds 11 patents. His main research interests are in high-resolution radar, microwave imaging, and applied electromagnetics (in particular, ultra-wideband (UWB) antennas).

Prof. Yarovoy was a recipient of the European Microwave Week Radar Award for the article that best advances the state of the art in radar technology in 2001 (together with L. P. Ligthart and P. van Genderen) and 2012 (together with T. Savelyev). In 2023, together with Dr. I. Ullmann, N. Kruse, R. Gündel, and Dr. F. Fioranelli, he received the Best Paper Award at IEEE Sensor Conference. In 2010, together with D. Caratelli, he received the Best Paper Award of the Applied Computational Electromagnetic Society (ACES). From 2008 to 2017, he served as the Director of the European Microwave Association (EuMA). He is/has been serving on various editorial boards such as that of IEEE TRANSACTIONS ON RADAR SYSTEMS. From 2011 to 2018, he served as an Associate Editor for the *International Journal of Microwave and Wireless Technologies*. He has been a member of numerous conference steering and technical program committees. He served as the General TPC Chair for the 2020 European Microwave Week (EuMW'20), the Chair and a TPC Chair for the 5th European Radar Conference (EuRAD'08), as well as the Secretary for the 1st European Radar Conference (EuRAD'04). He also served as the Co-Chair and a TPC Chair for the Xth International Conference on GPR (GPR2004).