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

Ruxin Zheng1, Shunqiao Sun1, David Scharff2 and Teresa Wu2,3

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

ABSTRACT

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

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

I. INTRODUCTION

Automotive radar operating at millimeter-wave frequency, i.e., 76-81 GHz, plays an important role in autonomous driving systems due to its robustness in environment perception under all weather conditions [1], [2]. Existing automotive radar transceivers, such as NXP Semiconductors MR3003 and Texas Instruments AWR1243 [3], support up to 3 transmit and 4 receive antennas, yielding an angular resolution of around $10^\circ$, which is not capable for Level 4 and Level 5 autonomous driving where a vehicle drives itself in all conditions without any human interaction. High resolution imaging radar systems are highly desired for Level 4 and Level 5 autonomous driving to provide point clouds of the surrounding environment [1], [4]–[6]. Cascaded radar chips rendering 12 transmit and 16 receive antennas are being developed [7]–[9] to synthesize a large virtual array using multi-input multi-output (MIMO) radar technology at a low cost [1], [2], [10]. Several products are available with different array configurations, such as forward-looking full-range radar AR540 by Continental [11]. With imaging radar, it is of great interest to investigate environment perception using deep neural networks.

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

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

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

II. SYSTEM MODEL

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

II-A. FMCW Radar Principles

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

$$f_T(t) = f_c + \frac{B}{T} t,$$ (1)
where $f_c$ is carry frequency. The phase $\varphi_T(t)$ of the transmitted signal could be obtained after integration as $\varphi_T(t) = 2\pi \int_{-T/2}^{t} f_T(t) \, dt$. The noiseless received signal is a delayed version of transmit signal. For a target at range of $R$ with radial velocity of $v$, the round-trip delay can be expressed as $\tau = 2(R + vt)/c$. The received signal is mixed with the transmit signal, and the output of the mixer is called beat signal, whose phase could be approximated as

$$\varphi_B(t) = 2\pi \left[ \frac{2f_bR}{c} + \left( \frac{2f_v}{c} + \frac{2BR}{Tc} \right) t \right],$$

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

**III. RADAR MACHINE LEARNING FRAMEWORK**

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

Our field experiments included three multi-modal sensors, i.e., a TI imaging radar, stereo cameras of Teledyne FLIR Blackfly S, and Velodyne Ultra Puck VLP-32C LiDAR, as shown in Fig. 2 (a). The measurements of cameras and
LiDAR are used as ground truth for labeling the radar data. The sensor features are summarized in Table II.

### III-A. High Resolution Imaging Radar Dataset

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

### III-B. Target Detection with SpectraNet

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

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

### III-C. Performance Evaluation

Class-wise random sampling method was used to divide the dataset into training and testing sets with a ratio of 8 : 2, ensuring that training and testing sets follow the overall dataset distribution. Additionally, a 5-fold cross-validation method was employed during the training process to find the optimal model by reducing data bias due to the lack of bus data in our dataset. The performance of SpectraNet was evaluated with various backbones, such as VGG16, VGG19 [24], and ResNet18 [25]. All networks were trained on a Dell workstation with dual Intel Xeon Gold 6230 CPU and an Nvidia Quadro RTX 4000 GPU. Radar
Fig. 4: Examples of high resolution radar spectra representing target’s shape. (a) Field experiments at different locations, with the first row as the RGB images, and the second row as the corresponding range-azimuth heatmap in Cartesian coordinate. Buses, cars, and people are labeled with bounding boxes in green, red, and white colors, respectively. (b) Radar’s performance is comparable to LiDAR, with 3D LiDAR point clouds BEV at the first row, and range-azimuth spectra at the second row.

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

<table>
<thead>
<tr>
<th>Backbone</th>
<th>mAP</th>
<th>AP@0.5 Bus</th>
<th>AP@0.5 Car</th>
<th>AP@0.5 People</th>
<th>AR@0.5 Bus</th>
<th>AR@0.5 Car</th>
<th>AR@0.5 People</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv4</td>
<td>66.3%</td>
<td>56.4%</td>
<td>77.6%</td>
<td>64.9%</td>
<td>70.0%</td>
<td>86.7%</td>
<td>80.8%</td>
</tr>
<tr>
<td>VGG16</td>
<td>54.3%</td>
<td>36.7%</td>
<td>71.3%</td>
<td>49.6%</td>
<td>42.5%</td>
<td>83.0%</td>
<td>61.0%</td>
</tr>
<tr>
<td>VGG19</td>
<td>67.7%</td>
<td>63.0%</td>
<td>82.0%</td>
<td>58.1%</td>
<td>67.5%</td>
<td>87.3%</td>
<td>73.3%</td>
</tr>
<tr>
<td>ResNet18</td>
<td>61.9%</td>
<td>87.4%</td>
<td>89.2%</td>
<td>69.9%</td>
<td>87.5%</td>
<td>90.3%</td>
<td>72.1%</td>
</tr>
<tr>
<td>ResNet50</td>
<td>59.7%</td>
<td>52.2%</td>
<td>78.8%</td>
<td>48.1%</td>
<td>55.0%</td>
<td>82.1%</td>
<td>54.9%</td>
</tr>
</tbody>
</table>

Table III: Evaluation metrics of the SpectraNet.

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

IV. CONCLUSIONS

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


