Time-sensitive and Distance-tolerant Deep Learning-Based Vehicle Detection Using High-Resolution Radar Bird's-Eye-View Images

Ruxin Zheng[†], Shunqiao Sun[†], Hongshan Liu[†] and Teresa Wu[‡§]

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

2023 IEEE Radar Conference (RadarConf23) | 978-1-6654-3669-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/RADARCONF2351548.2023.10149671

***Abstract*—Advanced driver assistance systems (ADASs) and autonomous vehicles rely on different types of sensors, such as cameras, 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 low hardware cost and reliable object detection under poor weather conditions, such as snow, rain, or fog, and doesn't suffer from light condition variations, such as darkness. High-resolution radar bird's-eye-view (BEV) obtained from radar range-azimuth spectra through a polar-to-Cartesian coordinate transform contains targets' geometric information that can be learned by deep neural networks for object detection. Compared to radar point clouds, there is no information loss in radar BEV. Unlike RGB images, radar BEVs are single-channel grayscale images with unique characteristics such as inconsistent resolution and SNR. Therefore, directly implementing an image-based object detection network is not an optimal solution for object detection using radar BEV. We propose a Temporal-fusion, Distance tolerant single stage object detection Network, termed as, *TDRadarNet*, to robustly detect vehicles up to 100 meters under various driving scenarios. DRadarNet leverages historical radar frames to exploit temporal features and separates far and near fields to address inconsistent resolution in radar frames. With qualitative and quantitative results, we show that TDRadarNet achieves 68.9% in precision and 66.8% in recall, and 67.8% in F1-score, which outperforms the state-of-the-art image-based object detection networks by 10.6%, 17.1%, and 14.1%.**

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

I. Introduction

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, like 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 [1]–[4]. Although cameras allow us to better understand visual scenes, their performance suffers in poor weather conditions [5]. LiDAR produces three dimensional (3D) point clouds of the environment with high-resolution on a good day by reflecting laser beams off surrounding objects [6], [7]. However, its performance degrades significantly in fog, dust, rain or snow. Further, the average price of LiDAR products is high.

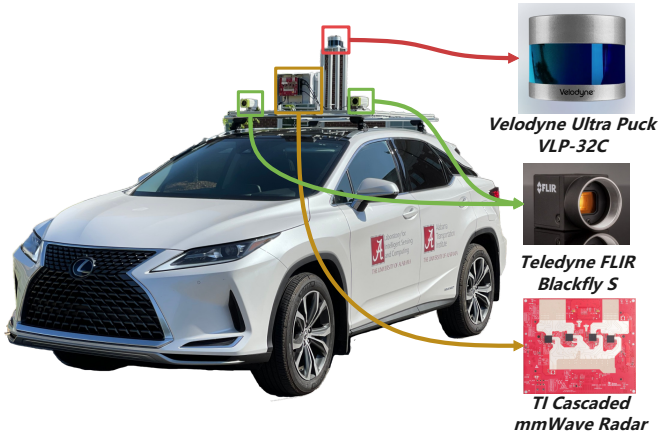


Fig. 1. The data acquisition vehicle platform of Lexus RX450h with high-resolution imaging radar, LiDAR, and stereo cameras, that is used to carry out field experiments at The University of Alabama.

Radar, on the other hand, is robust, inexpensive, and reliable even in harsh environments [5], [8], [9]. Automotive radar sensor is a fundamental part of advanced driver assistance systems (ADASs) and autonomous vehicles largely because of its inexpensive circuitry, ability to sense during inclement weather, and immunity to poor visibility conditions [5], [8]–[12]. Frequency-modulated continuous-wave (FMCW) has been widely adopted in automotive radars as transmit signals to achieve high range resolution sensing at a low-cost for both commercial vehicles with ADAS features and fully autonomous vehicles. [13], [14]. The wavelength of the millimeter-wave automotive radar operating at 76-81 GHz is in the millimeter range. A high bandwidth of total 4 GHz in carrier frequency 77-81 GHz is available for short-range 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

This work was supported in part by U.S. National Science Foundation (NSF) under Grant CCF-2153386 and Alabama Transportation Institute (ATI).

Authorized licensed use limited to: UNIV OF ALABAMA-TUSCALOOSA. Downloaded on June 24,2023 at 15:22:53 UTC from IEEE Xplore. Restrictions apply.

with optical sensors such as LiDAR and cameras, millimeter-wave automotive radar has strong penetration capabilities in fog, rain, snow, smoke, and dust, making it robust to bad weather conditions [9]. However, the potential of object detection and classification using automotive radar has not been fully exploited. Today, most radar devices used 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 high angular resolution close to LiDAR [9]. Therefore, high-resolution automotive imaging radars [13], [15] are of great interest to support object detection and classification. Some commercial imaging radar products are available with different array configurations, such as forward-looking full-range radar of ZF [16] and ARS540 of Continental [17]. Both provide radar point clouds only.

Deep learning has found wide application in radar systems [18], [19]. For example, low-cost radar, such as Soli radar [20], 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 [21]. Radar has a long application history in commercial automobiles [8] 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 roughly 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 (CFAR), on the radar range-azimuth map. In this way, radar produces sparse point clouds, and thus it can be viewed as a low-quality LiDAR. Point clouds based object detection networks, such as PointPillars [22], VoxelNet [23] and PointNets [24], can be directly used or adjusted [25] for radar point clouds. However, such thresholding algorithms in generating radar point clouds lead to significant information loss of objects.

2) Radar Data Tensor: To avoid loss of information, radar data can be processed in three-dimensional (3D) tensors, i.e., range-Doppler-azimuth for one-dimensional (1D) antenna array, or four-dimensional (4D) tensors, i.e., range-Doppler-azimuth-elevation for two-dimensional (2D) antenna array. Deep learning based radar detectors [19] can directly learn from 4D complex radar tensors for object detection and localization. It's also possible to project the 3D radar tensors along different views to extract 2D features for semantic segmentation [26] and object recognition [27]–[29].

3) Radar BEV: Radar bird's-eye-view (BEV) is 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 were proposed in [30]–[32], achieving relatively accurate object detection. However, only highway scene is considered in [30], which are considered as the clean and easy scenario in autonomous driving. In [31], the radar is placed at intersections, and only moving targets in the near field are of interest. Similarly, objects within ultra short range are considered in [32].

Taking advantage of temporal and spatial information of radar data can effectively improve radar detector performance [33]. Extensive studies have been conducted on the combination of different radar frames, such as summation among neighboring frames [26], concatenation in frame level [34], and stacking in feature level [27], [35]. In [27], a convolutional long short-term memory (LSTM) layer is adopted after the encoder network to extract temporal features from a sequence of feature maps. In [35], frame-level feature maps are concatenated and temporal features are extracted by a 3D convolutional neural network (CNN) layer. Other than using a CNN-based network, an isotropic graph convolution network (GCN) that leverages spatial information from neighboring nodes is proposed in [36] to boost radar detection performance.

In this paper, we propose a high-resolution radar object detection deep learning network that can robustly detect vehicles up to 100 meters under various driving scenarios. The proposed network is Temporal-fusion, Distance tolerance Radar object detection network, termed as *TDRadarNet*, for vehicle detection. In this work, we make efforts to leverage historical radar frames to exploit temporal features and separate far and near fields to address inconsistent resolution in radar frames, The evaluation results show that the proposed TDRadarNet outperforms the state-of-the-art image-based object detection network (baseline) with 10.6%, 17.1%, and 14.1% improvements in precision, recall, and F1-score, respectively and demonstrates the effectiveness of the proposed temporal fusion and far-near-field design, indicating the potential for robust radar object detection in the real-world driving scenes.

II. TDRadarNet

Unlike RGB images, radar BEVs are single-channel grayscale images with unique characteristics such as inconsistent resolution and SNR. Therefore, directly implementing an image-based object detection network is not an optimal solution for object detection using radar data. In this section, we first review the unique characteristics of radar BEVs and then develop a novel deep neural network, namely TDRadarNet to identify vehicles from radar BEVs.

A. Not Every Pixel is Created Equally

After applying 3D FFT on a radar data cube along fast-time, slow-time, and channel, a radar tensor in range-Doppler-azimuth is obtained, of which the range-azimuth spectra image is the BEV that can be utilized for machine learning. However, not every pixel in the BEV is generated with the same accuracy. In general, the pixels in BEV are sensitive to both range and angle.

1) Effective Antenna Aperture Relies on Angle: For a uniform linear array, the half power beamwidth (HPBW) [37] is $\theta_B \approx \frac{0.886\lambda}{Nd\cos\theta}$, which indicates that the effective array aperture decreases as the view angle increases. The maximum effective

Authorized licensed use limited to: UNIV OF ALABAMA-TUSCALOOSA. Downloaded on June 24,2023 at 15:22:53 UTC from IEEE Xplore. Restrictions apply.

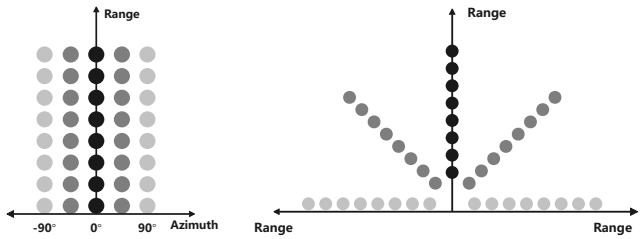
antenna aperture or equivalently best angular resolution is achieved along the boresight direction.

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 [38]:

$$P_r = \frac{P_t C \sigma}{(4\pi)^3 r^4}, \tag{1}$$

where C can be considered as a constant number for the same radar, which includes antenna gain, effective antenna area, and efficiency; 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 obtained radar range-azimuth spectra are in polar coordinates, as shown in Fig. 2 (a). Usually, the BEV in polar coordinates is first transformed or interpolated into a Cartesian coordinate system before being fed into deep neural networks. It can be found in Fig. 2 (b) that as the range increases, the distance between adjacent bins becomes larger. In other words, the variance of adjacent pixels' distance in Cartesian coordinate becomes large for pixels or targets with large ranges [27]. And that variance is further amplified by the relatively large DOA estimation error due to SNR drops at long ranges.



(a) (b)

Fig. 2. (a) Polar coordinate and its (b) Cartesian coordinate transformation.

In summary, the radar BEV image obtained using MIMO radar with non-rotate antenna arrays is hardly shift-invariant. Instead, it is shift-variant over both angles and ranges.

B. Network Architecture

Because of the above unique radar BEV characteristics, we propose TDRadarNet as shown in Fig. 3. The TDRadarNet consists of two identical networks that are trained to detect objects in far and near fields, respectively. The input sequences of radar frames are divided into overlapped sequences of far fields and near fields. For each sequence, we use a backbone to extract features frame-wise, and then we develop the temporal fusion stage to explore the frame relationship with the historical frames, lastly, with the predictions made by the detection head, we merge the far field and near field result. The proposed network is inspired by You Only Look Once v7 (YOLO) [39] and optimized for radar BEVs by learning different features of far and near fields and integrating temporal information of historical radar frames, the details are as follows:

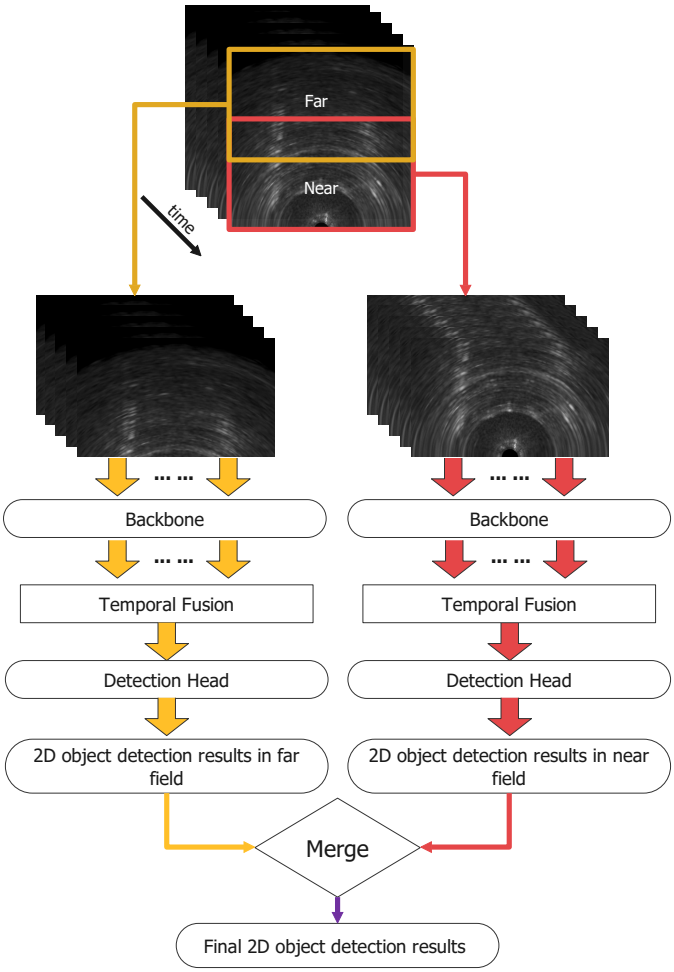


Fig. 3. TDRadarNet. Far-near fields are divided and used to train two sets of learnable parameters of the model. Temporal fusion works by extracting temporal features from a sequence of frames.

1) Far and Near Fields: Due to the resolution difference in the radar BEVs, objects present varying intensity, shape, and contrast, thus it is challenging for a single model to detect the same objects with a large dissimilarity. We divide the radar frame into two overlapping regions, the far field and the near field, as shown in Fig. 3. The overlapping dividing strategy ensures that no information is lost across boundaries. The divided regions are sent to separate tracks to train the deep learning model which learns two sets of parameters for predicting the object in far and near fields, respectively.

2) Temporal Fusion: Even though the predictions can be made with the current radar frame, using historical frames can contribute temporal information. As shown in Fig. 4, with a sequence of n frames, the backbone extracts feature maps of three scales for each frame, resulting $3 \times n$ feature maps. We concatenate the feature maps that belong to the same scale, and then use a convolution kernel followed by batch normalization and ReLu activation to extract the features along the temporal direction.

Authorized licensed use limited to: UNIV OF ALABAMA-TUSCALOOSA. Downloaded on June 24,2023 at 15:22:53 UTC from IEEE Xplore. Restrictions apply.

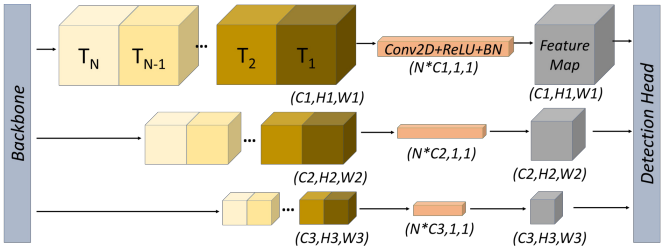


Fig. 4. Temporal fusion module. Feature maps extracted from N consecutive frames are concatenated in three scales separately. The temporal features are then extracted by applying a convolution kernel followed by batch normalization and ReLu activation.

3) Merging: After we collect the object detection results in the far field and near field, we merge the results into a single frame. For the duplicated detections in the common region of both fields, we apply the non-maximum suppression to filter out largely overlapped detections by setting the intersection of union (IoU) threshold. The detection with the highest confidence score will be kept if multiple detections share the same intersection.

III. Experiments

In this section, we collect our automotive radar dataset, BAMA, through field experiments. The BAMA dataset contains high-resolution radar BEVs with corresponding stereo camera images and LiDAR point clouds. To our best knowledge, there is no existing automotive radar dataset with high angular resolution and up to a 100 meter detectable range. We evaluate TDRadarNet with the BAMA dataset quantitatively and qualitatively by comparing it with a baseline model and performing ablation experiments.

A. BAMA Dataset

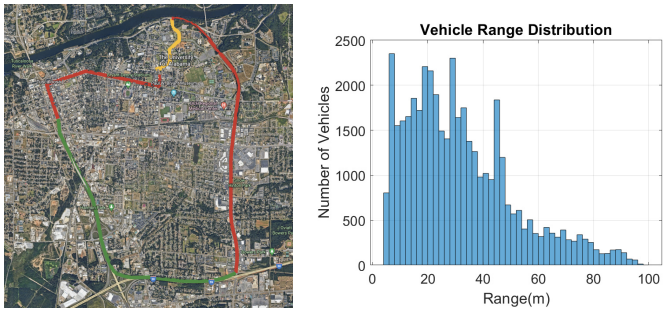
Our field experiments included three multi-modal sensors, i.e., a TI imaging radar, stereo cameras of Teledyne FLIR Blackfly S, and Velodyne Ultra Puck VLP-32C LiDAR, as shown in Fig. 1. The measurements of cameras and LiDAR are used as ground truth for labeling the radar data. The sensor features are summarized in Table I.

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

TABLE I
Multi-Modal Sensors.

We drove over 30 minutes to collect data around the city of Tuscaloosa, Alabama, USA. Our driving route is shown in Fig. 5 (a), which consists of three types of driving scenarios, such as campus road, urban street, and highway.

BAMA dataset contains $14,800$ radar BEV frames, which are generated by a deep learning aided TDM MIMO radar signal processing pipeline [40], with synchronized stereo camera images and LiDAR 3D point clouds. There are different types of objects of interest, including pedestrians, cars, trucks, and buses. For simplicity, in this paper, we focus on vehicle detection only [41]. A total number of $42,390$ vehicles at various ranges are labeled using camera images and LiDAR 3D point clouds as ground truth. Vehicle range distribution is shown in Fig. 5 (b). Examples under various driving scenarios are shown in Fig. 6.



(a) (b)

Fig. 5. (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.

B. Implementation Details

From $14,800$ high-resolution radar BEV frames, we use $11,500$ for training and $3,300$ for testing. The single frames dataset is used for the baseline training. For the proposed TDRadarNet, we use the sequence of radar frames as input and the last frame annotation 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 4 Nvidia RTX A6000 GPUs. The baseline model YOLO using 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 ablations experiments to evaluate the far-near field and temporal fusion design, with all the hyperparameters the same as the implementation of TDRadarNet.

C. Results and Discussion

We evaluate the detection performance using precision, recall, and F1-score with an IoU of 0.5. The metrics are defined below.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{3}$$

$$\text{F}_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}, \tag{4}$$

where the true positive (TP) is the correct positive prediction, the false positive (FP) is the incorrect positive prediction, and the false negative (FN) is the incorrect negative prediction.

The quantitative evaluation results are shown in Table. II. The proposed TDRadarNet leads to obvious improvement over

Authorized licensed use limited to: UNIV OF ALABAMA-TUSCALOOSA. Downloaded on June 24,2023 at 15:22:53 UTC from IEEE Xplore. Restrictions apply.

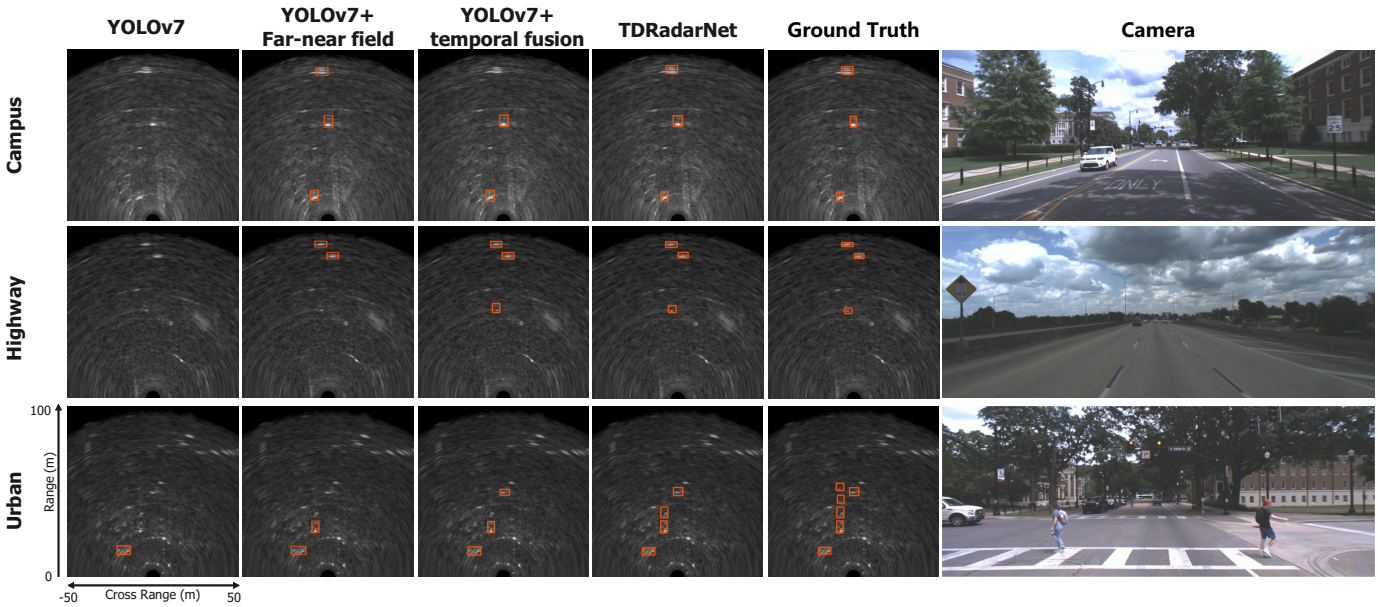


Fig. 6. Examples from our test set. Detection results are marked in red. (Note: The pixels represent locations beyond 100m from ego-vehicle (0,0) are set as zeros, corresponding to dark pixels.)

the baseline model by 10.6% in precision, 17.1% in recall, and 14.1% in F1-score. For the far-near field design and temporal fusion design, the ablated models generally outperform the baseline method.

The representative detection results over three scenarios, the corresponding original radar frames, and ground truth annotations are shown in Fig. 6. 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=$1,700$ and $2,300$, respectively) are available compared to the urban scenario (n=$7,500$).

D. Ablation Study

For the far-near field design and temporal fusion design, the ablated experiments are conducted as, YOLOV7, YOLOv7 plus far-near-field only, YOLOv7 plus temporal fusion only, and TDRadarNet with two designs. The evaluation results are shown in Table. II. With adding the far-near-field design or temporal fusion design, the improved performance is observed in all three metrics. In the second column of Fig. 6, the implementation of far-near field not only helps to detect vehicles that are far in distance and have lower resolution, but also improves the detection in the near field, by learning field-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.

Networks	Precision	Recall	F1-score
YOLOv7	58.3%	49.7%	53.7%
YOLOv7+far-near field	62.8%	57.9%	60.2%
YOLOv7+temporal fusion	67.9%	59.0%	63.2%
TDRadarNet	**68.9%**	**66.8%**	**67.8%**

TABLE II
OBJECT DETECTION RESULTS OF DEEP LEARNING MODELS IN PRECISION, RECALL, AND F-1 SCORE.

IV. CONCLUSIONS

The proposed TDRadarNet innovatively combines temporal fusion and far-near field designs into vehicle detection tasks for the radar BEV dataset. We investigate and compare the detection performance of TDRadarNet with baseline model. The result proves that the proposed model is superior to the baseline model in producing accurate results. The ablation experiments demonstrate the effectiveness of the two designs.

The potential limitation of TDRadarNet could be the extra computational cost introduced by the temporal fusion over a sequence of frames. Besides, 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 could 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] D. Feng, C. H. Schutz, L. Rosenbaum, H. Hertlein, C. Glaser, F. Timm, W. Wiesbeck, and K. Dietmayer, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1341–1360, 2021.

Authorized licensed use limited to: UNIV OF ALABAMA-TUSCALOOSA. Downloaded on June 24,2023 at 15:22:53 UTC from IEEE Xplore. Restrictions apply.

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

[3] Y. Xiao, F. Codevilla, A. Gurram, O. Urfalioglu, and A. M. López, "Multimodal end-to-end autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 537–547, 2022.

[4] Y. Cui, R. Chen, W. Chu, L. Chen, D. Tian, Y. Li, and D. Cao, "Deep learning for image and point cloud fusion in autonomous driving: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 722–739, 2022.

[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 Processing Magazine*, vol. 37, no. 4, pp. 98–117, 2020.

[6] Y. Li and J. Ibanez-Guzman, "Lidar for autonomous driving: The principles, challenges, and trends for automotive Lidar and perception systems," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 50–61, 2020.

[7] 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 Processing Magazine*, vol. 38, no. 1, pp. 68–86, 2021.

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

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

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

[11] 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 Processing Magazine*, vol. 34, no. 2, pp. 36–46, 2017.

[12] Z. Peng, C. Li, and F. Uysal, *Modern Radar for Automotive Applications*. London, UK: IET, 2022.

[13] S. Sun and Y. D. Zhang, "4D automotive radar sensing for autonomous vehicles: A sparsity-oriented approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 4, pp. 879–891, 2021.

[14] G. Duggal, S. Vishwakarma, K. V. Mishra, and S. S. Ram, "Doppler-resilient 802.11ad-based ultrashort range automotive joint radar-communications system," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 5, pp. 4035–4048, 2020.

[15] S. Sun and Y. D. Zhang, "Four-dimensional high-resolution automotive radar imaging exploiting joint sparse-frequency and sparse-array design," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2021, pp. 8413–8417.

[16] https://www.zf.com/products/en/cars/products_58368.html, May 2021.

[17] Continental Automotive, https://www.continental-automotive.com, accessed: Sept. 2022.

[18] E. Mason, B. Yonel, and B. Yazici, "Deep learning for radar," in *IEEE Radar Conference*, Seattle, WA, May 8-12, 2017.

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

[20] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Transactions on Graphics*, vol. 35, no. 4, pp. 1–19, 2016.

[21] C. Li, J. Cummings, J. Lam, E. Graves, and W. Wu, "Radar remote monitoring of vital signs," *IEEE Microwave Magazine*, vol. 10, no. 1, pp. 47–56, 2009.

[22] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "Pointpillars: Fast encoders for object detection from point clouds," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 12 697–12 705.

[23] Y. Zhou and O. Tuzel, "VoxelNet: End-to-end learning for point cloud based 3D object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[24] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, July 2017.

[25] A. Danzer, T. Griebel, M. Bach, and K. Dietmayer, "2D car detection in radar data with PointNets," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 61–66.

[26] A. Ouaknine, A. Newson, P. Pérez, F. Tupin, and J. Rebut, "Multi-view radar semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 15 671–15 680.

[27] B. Major, D. Fontijne, A. Ansari, R. Teja Sukhavasi, R. Gowaikar, M. Hamilton, S. Lee, S. Grzechnik, and S. Subramanian, "Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors," in *IEEE/CVF International Conference on Computer Vision (CVPR) Workshops*, 2019, pp. 0–0.

[28] X. Gao, G. Xing, S. Roy, and H. Liu, "Ramp-CNN: A novel neural network for enhanced automotive radar object recognition," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 5119–5132, 2020.

[29] A. Zhang, F. E. Nowruzi, and R. Laganiere, "Raddet: Range-azimuth-doppler based radar object detection for dynamic road users," in *2021 18th Conference on Robots and Vision (CRV)*. IEEE, 2021, pp. 95–102.

[30] X. Dong, P. Wang, P. Zhang, and L. Liu, "Probabilistic oriented object detection in automotive radar," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[31] R. Zheng, S. Sun, D. Scharff, and T. Wu, "Spectranet: A high resolution imaging radar deep neural network for autonomous vehicles," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Trondheim, Norway, June 20-23, 2022, pp. 301–305.

[32] S. Madani, J. Guan, W. Ahmed, S. Gupta, and H. Hassanieh, "Radatron: Accurate detection using multi-resolution cascaded MIMO radar," in *European Conference on Computer Vision (ECCV)*, 2022, pp. 160–178.

[33] 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, 2022.

[34] J. Peršić, L. Petrović, I. Marković, and I. Petrović, "Spatio-temporal multisensor calibration based on Gaussian processes moving object tracking," *arXiv preprint arXiv:1904.04187*, 2019.

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

[36] M. Meyer, G. Kuschk, and S. Tomforde, "Graph convolutional networks for 3D object detection on radar data," in *IEEE/CVF International Conference on Computer Vision (CVPR) Workshop*, 2021, pp. 3060–3069.

[37] C. Balanis, *Antenna Theory: Analysis and Design, 4th Edition*. Wiley, 2016.

[38] M. A. Richards, *Fundamentals of Radar Signal Processing*, 2nd ed. McGraw-Hill, 2014.

[39] 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," *arXiv preprint arXiv:2207.02696*, 2022.

[40] R. Zheng, H. Liu, and S. Sun, "A deep learning approach for Doppler unfolding in automotive TDM MIMO radar," in *IEEE 56th Annual Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Oct. 30 - Nov. 2, 2022.

[41] K. Qian, S. Zhu, X. Zhang, and L. E. Li, "Robust multimodal vehicle detection in foggy weather using complementary Lidar and radar signals," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, June 19-25, 2021.

Authorized licensed use limited to: UNIV OF ALABAMA-TUSCALOOSA. Downloaded on June 24,2023 at 15:22:53 UTC from IEEE Xplore. Restrictions apply.