Received 28 November 2023, accepted 11 December 2023, date of publication 15 December 2023,
date of current version 21 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3343383

RESEARCH ARTICLE

F-ROADNET: Late Fusion-Based Automotive Radar Object Detection

GULBADAN SIKANDER[1], SHAHZAD ANWAR[1,2], GHASSAN HUSNAIN[3], AND SANGSOON LIM[4], (Member, IEEE)

[1]Department of Mechatronics Engineering, University of Engineering and Technology, Peshawar, Peshawar 25000, Pakistan
[2]Artificial Intelligence in Healthcare, Intelligent Information Processing Laboratory, National Center of Artificial Intelligence, University of Engineering and Technology, Peshawar, Peshawar 25000, Pakistan
[3]Department of Computer Science, CECOS University of IT and Emerging Sciences, Peshawar 25100, Pakistan
[4]Department of Computer Engineering, Sungkyul University, Anyang, Gyeonggi 14097, South Korea

Corresponding author: Sangsoon Lim (slim@sungkyul.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government [Ministry of Science and ICT (MSIT)] under Grant 2021R1F1A1063319.

ABSTRACT Road user categorization is essential for autonomous driving perception. In challenging traffic situations including unfavorable weather (such as fog, snow, and rain) and dim lighting. There are several kinds of sensors that need to be researched in order to achieve the precision and resilience that autonomous systems demand. Currently, to create a depiction of the environment surrounding the vehicle, principally cameras and laser scanners (LiDAR) are commissioned. Despite their enticing qualities, Radar sensors are currently underutilized for autonomous driving, even though, they have been employed in the automobile industry for a long time. Radar's ability to measure the relative speed of obstacles and to operate even in adverse weather conditions makes it a front line contender for road user detection. This study proposes F-ROADNET, a multi-object classification method for vulnerable road users based on raw Radar data. The model is trained on Range Angle and Range Doppler maps based on a late fusion architecture. F-ROADNET has a detection accuracy of 99.01%, precision of 99.3% and recall of 99% on the CARRADA dataset and detection accuracy of 91.62%, precision of 87.2% and recall of 90.2% on the RADDet dataset. The findings exhibit that F-ROADNET outperforms established methods in terms of average precision.

INDEX TERMS ADAS, automotive radar, road user safety, late fusion.

I. INTRODUCTION

The development of advanced driver assistance systems (ADAS) during the past decade has increased the number of in-car sensors [1], including Radar, LiDAR, and cameras [2]. Together, these sensors provide the ability to visualize its surroundings and adjust its function accordingly. Nowadays, camera and LiDAR are employed for ADAS applications in the majority of intelligent vehicles as camera and LiDAR produce high-resolution output and perform well in 3D object identification and classification tasks.

Radar sensors have been mostly employed for blind spot detection or autonomous cruise control due to their subpar angular resolution, which has prevented them from being employed for object classification and detection tasks. Radars are unaffected by light or weather and offer information including range and velocity of the surrounding objects, Radar sensors seem particularly well-suited for important and real-time automotive applications such as autonomous emergency braking. Many self-driving or assisted-driving LiDARs rely on sensor fusion to improve the accuracy and reliability of perception findings, with Radar serving as a supplement to cameras or LiDARs. It is mostly due to the fact that most fusion systems only use robust location information in Radar signals and rich semantic information is less explored.

Object identification and classification is one of the most important jobs in computer vision, and deep neural networks

The associate editor coordinating the review of this manuscript and approving it for publication was Fabrizio Santi.

have made significant progress in this area over the last decade. Such approaches have been deployed successfully with LiDAR and cameras [3], [4], however a lack of publicly available annotated Radar datasets has delayed research in object detection and segmentation from Radar data. Although vision sensors may offer deeper semantic understandings of visual environments, they are not a robust sensor under adverse settings for instance, low/high lighting or inclement weather, which results in little/high exposure or blurry/occluded images. LiDAR, on the other hand, is an alternate sensor with point cloud data that could be utilized for accurate object detection and localization. Following the pioneering work on feature extraction from point cloud [5], following object detection from LiDAR point cloud [6], [7] has been addressed. However, for detailed semantic information, these approaches require a relatively dense LiDAR point cloud, not to mention the high equipment and processing expenses. Moreover, Radar is typically more dependable in difficult conditions. The following characteristics apply to frequency modulated continuous wave (FMCW) Radar, which works in the millimetre-wave (MMW) band (30-300GHz), below the visible light spectrum: 1) MMW has excellent ability to pass through dust, smoke, and fog; 2) FMCW Radar has excellent range detection capabilities due to enormous bandwidth and high working frequency. The FMCW Radar typically uses two types of data representations: RF images and Radar points. Fast Fourier transforms (FFTs) [8] are incorporated to create the RF images from the raw Radar signals, and a peak detection technique is employed to create the Radar points from these frequency images.

Despite the fact that Radar points could be directly incorporated into methods created for LiDAR point cloud [9], however, Radar points are typically more sparser (less than 5 points on a nearby car) than the LiDAR point cloud [10], so the information is insufficient to complete object detection. In contrast, RF images could preserve detailed Doppler and object motion information, enabling the comprehension of the semantic significance of a given object. It is possible to replace and enhance conventional methods for object recognition, classification, and segmentation without sacrificing information by using raw data tensors and deep neural networks. Recent research on new deep learning techniques for automobile Radar, spanning from object detection [11], [12], to object segmentation [13], has been made possible by Radar datasets and challenges like CARRADA [14], RADDet [15], or CRUW [16], where Radar data is offered as raw data tensors.

This study is motivated by developing a novel Radar based object detection and classification method which is capable of accurately detecting vulnerable road users well in advance. The main contributions of this research are proposing F-ROADNET, a novel late fusion based automotive Radar object detection method based on Range Doppler (RD) and Range Angle (RA) maps. The use of RD and RA maps fusion rather than Range-Angle-Doppler (RAD) tensors is prompted by the fact that producing RAD tensors is more computationally intensive and single map do not provide suitable detection accuracy. The study was evaluated on the well established CARRADA and RADDet dateset, it was then compared to established methods to prove its performance capability.

/hlThe novelty of this studies lies in the fact that the Radar data is treated like images, hence a simple convolutional model could be used to classify the data, even though some preprocessing is involved in converting the radar data to image format but it greatly reduced the complexity of the model.

The remaining of this manuscript is organized as following. Section II presents a review of existing road user based detection methods and establishes the need for this study. Section III presents the overall methodology of the late fusion based automotive Radar object detection method referred to as F-ROADNET here after. Section IV elaborates the dataset and training parameters incorporated in this study. Section V discusses the results. Finally, Section VI concludes this article.

II. RELATED WORK

Pedestrian detection has been a vital part of ADAS. Significant research has been performed on pedestrian detection, where particular emphasis is given to timely and accurate detection. Originally camera was the sensor of choice, however, recently LiDAR and Radar sensors are also under consideration. Sensor fusion has also been incorporated to enhance detection accuracy. In this section, pedestrian detection methods are categorized as camera, LiDAR, Radar and multi-modal fusion based methods.

A. VISION BASED METHODS

In the past decade, the availability of benchmark datasets, namely, KITTI [17] and Caltech [18] has driven the performance of vision based pedestrian detection to new heights. Traditionally, RGB image data has been utilized for pedestrian detection, while recent research additionally explore depth information [19], [20]. Vision based pedestrian could be dated back to 2003 when Viola and Jones (VJ) [21] incorporated their infamous Viola-Jones detector to detect walking pedestrians in frames. The VJ detector extracts features for each frame and integrates intensities with motion information for efficient pedestrian detection. Vision based pedestrian detection took a stride forward with the consideration of the people model [22], where the model considers the human body as a combination of deformable parts. Plentiful of research have incorporated the deformable parts model of accurate pedestrian detection [23], [24]. With the advent of deep learning, researchers also shifted gears towards learning features instead of hand crafting features. Region based Convolutional neural networks (RCNN) [25] and its variants Fast RCNN [26]and Faster RCNN [27] have been widely popular for pedestrian detection.

Many challenges are encountered in vision based pedestrian detection including camera motion [28], illumination variation [29], shadows [30], pedestrian occlusion [31] and abrupt motion [32]. overcoming these challenges are computationally expensive hence other sensors are potentially explored.

B. LIDAR BASED METHODS

LiDAR (Light Detection and Ranging) technologies give cost-effective and straightforward solutions to difficulties left unanswered by image processing. LiDARs are devices that emit layers of high-directivity lasers that reflect on all surrounding objects and provide information about their relative distance. These characteristics make LiDARs ideal for pedestrian detection applications in autonomous cars, where accurate distance estimate is a crucial element for collision prevention. The Velodyne LiDAR is commonly utilized for pedestrian detection and dataset including NUSCENE [33] provide relevant data for method training and testing. In [34], it was suggested to use streams of LiDAR point clouds to identify smartphone zombies using a tracking improved segmentation-based technique. The developed system uses tracking-enhanced SVM-based feature extraction to learn a distinctive pattern of the lateral profile for smartphone zombie detection. However, significant bottlenecks on segmentation and tracking must be overcome to improve the robustness of the system, and temporal information and additional features must be investigated to improve the detection accuracy. A more time efficient method is presented in [35], where the pedestrian recognition module uses a cutting-edge U-shaped CNN framework to input a specially built voxel feature; nonetheless, accuracy is degraded while confronting a throng of people or if there are frequent and thick occlusions in the environment.

With increasing range, the LiDAR point cloud gets sparse, and the ability to detect objects falls considerably. Weather conditions, for instance, severe rain, snow, or fog could also have an impact on camera and LiDAR systems [36]. These constraints enhance the hazards connected with autonomous driving and severely limit the development of related technology.

C. RADAR BASED METHODS

Radar sensors have lately gained popularity in driving applications, which use representations that vary depending on the task at hand: Doppler spectrograms for classifying vehicles [37], range-angle for classifying objects [38], and range-angle and range-Doppler for object identification [39]. S. Capobianco et al. [37] examine how well Convolutional Neural Networks (CNN) could identify different types of vehicles while analyzing an FM-CW Radar signal. Based on short time fourier transformation, convert a two-dimensional Radar wave into a three-dimensional tensor. The DeepRadarNet convolutional architecture is then trained using the computed tensors. A more detailed CNN based classifier is presented in [38], this research study Radar range-angle information for object detection. In [40] the feasibility of integrating heterogeneous lidar PC and radar raw data was examined. The suggested method used enhanced Doppler contexts to not only increase object identification performance for vulnerable road users, but also to precisely classify object motion status. However, incorporation of a larger data set and further comparisons with developed methods have to performed for verification.

A more recent study [41] propose a self-attention-based 4D-Radar 3D object detection network called RPFA-Net. The point cloud features in the pillar are extracted using the self-attention method, and RPFAlayer is suggested as a way to remotely record the data. Resulting in enhanced 3D object identification accuracy and network regression heading angle precision.

Although some improvements were reported, the algorithm was applied solely to Radar point cloud and did not fully use data from the Radar.

D. MULTI MODAL FUSION BASED METHODS

One of the fundamental issues in multi-modal fusion is maintaining a strong link between various aspects of the same item from each modality. While LiDAR offers great range, reflectance information, and day-long availability, it lacks enough deterministic information to allow some classifiers to correctly identify pedestrians. An Imaging device could potentially facilitate correct pedestrian identification. Such a method is presented in [42], where through point-wise and ROI-wise feature fusion, multi-task learning for map previous and geometric cues, and denser feature fusion, multi-task multi-sensor detection is carried out. Wang et al. achieved the best performance for the KITTI 3D object detection benchmark in [43] by offering a new viewpoint for converting images into a point cloud. However, when there were few LiDAR points and fusion was constrained. Additionally, unfavorable weather conditions including snow, rain, and fog greatly reduced the quality of the data produced by imaging device and LiDAR.

Fusion of vision and Radar data for object recognition in adverse weather should result in improved outcomes and increase the resilience of 3D object detection. The camera data is often employed in Vision-Radar fusion algorithms to extract area suggestions or 2D bounding boxes for detection, while 3D Radar gives depth information for the final detection results. One such method is RODNET [44], a deep Radar object detection network. The Radar-based object identification technique that could be more reliable than vision in challenging lighting situations. The suggested RODNet is based on a completely systematic cross-modal supervision scheme from an efficient vision-Radar fusion method. Even though RODNET performs fairly, however, its computationally expensive. RAMP-CNN's [45] extra Doppler channel enhanced the precision by 1% over RODNet, however this performance advantage comes at the

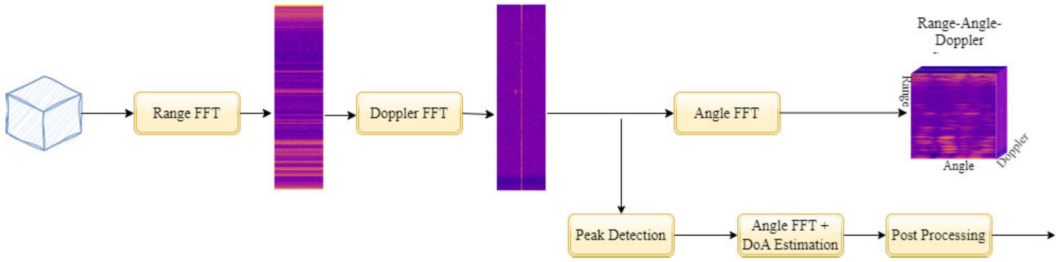


FIGURE 1. Radar data acquisition block diagram.

expense of more (3x) model parameters. To address issue of computational cost the T-RODNET [46] utilizes transformer to obtain global features causing increased model parameters. A more recent study [47] presents an image radar cross modal technique with radar customised panoptic segmentation in urban, rural, and highway situations using just range-Doppler matrix data.

However, in this instance, 3D Radar just serves as an adjunct, and there is still a sizable gap in the estimation of object depth.

Sparsity brought on by the filtering methods applied on the raw Radar signal and the post-processing procedures causes loss of important data included in the raw Radar signal shown in Fig. 1. Several publications investigate lower level representations, primarily the Range-Doppler (RD), Range-Angle (RA), or Range-Angle-Doppler (RAD) tensors, to overcome this. As RAD tensors combine distance, velocity, and angle information, this representation is being adopted in an increasing number of works. Even though RAD tensors provide the most information, they are computationally expensive for Radar processors. In this study we present a late fusion based Deep learning method for accurate pedestrian detection incorporating raw Radar data.

III. MATERIALS AND METHOD

This section presents the proposed novel late fusion Radar object detection deep learning method (F-ROADNET) for accurate pedestrian detection. Three five-layer CNN models are trained on RD, RA and AD respectively. Late (Decision level) fusion is performed to achieve the final decision. Late fusion is a merging approach that takes place outside of monomodal classification models. It integrates the results of each classifier to generate new, more accurate and trustworthy decisions.

A. OVERVIEW OF RADAR SENSOR

A Radar sensor produces electromagnetic waves using one or more transmitter antennas (Tx). The waves are reflected by an object and picked up by the Radar using one or more receiving antennas (Rx). The comparison of the sent and received waveforms yields the distance, radial velocity, azimuth angle, and elevation of the reflector in relation to the Radar position [8]. The majority of car Radars employ

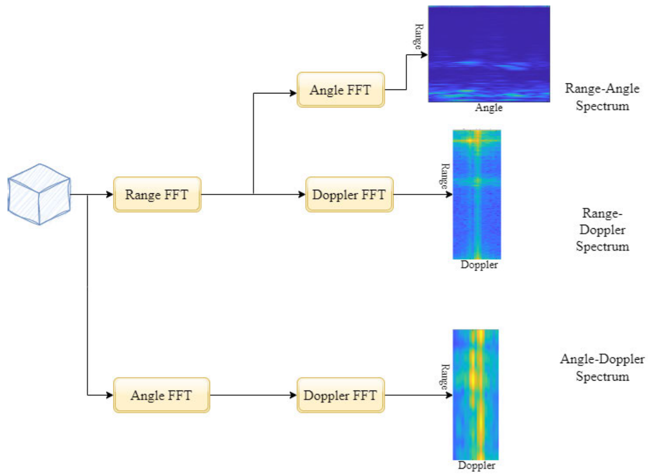


FIGURE 2. Raw radar data format.

Multiple Input Multiple Output (MIMO) systems, in which each Tx/Rx pair receives the reflected signal allocated to a specific Tx broadcasting a waveform. Frequency-Modulated Continuous Wave (FMCW) Radar sends out a signal known as a chirp [5], the frequency of which is linearly modulated throughout the sweeping period Ts. The received signal has a phase shift of $\phi(t)$. Radial velocity is determined by the frequency shift between the two signals, often known as the Doppler effect. The doppler frequency can be expressed as:

$$f_d = \frac{1}{2\pi}\frac{d\phi(t)}{dt} = \frac{2v_r}{c}f_s \tag{1}$$

The radial speed of the object could be recovered using equation. 1. The time delay between the received signals of each Rx provided by a particular Tx in a MIMO system with multiple Rx antennas conveys the object's orientation information. Depending on the location of the antennas, the azimuth angle and the elevation of the object are respectively deduced from the horizontal and vertical pairings of Tx/Rx. The Fast Fourier Transform (FFT) technique converts recorded data from the time domain to the frequency domain using a Discrete Fourier Transform (DFT). A 3D-FFT is utilized to process the 3D tensor: a Range-FFT along the rows to resolve the object range, a Doppler-FFT along the columns to resolve the object radial velocity, and an

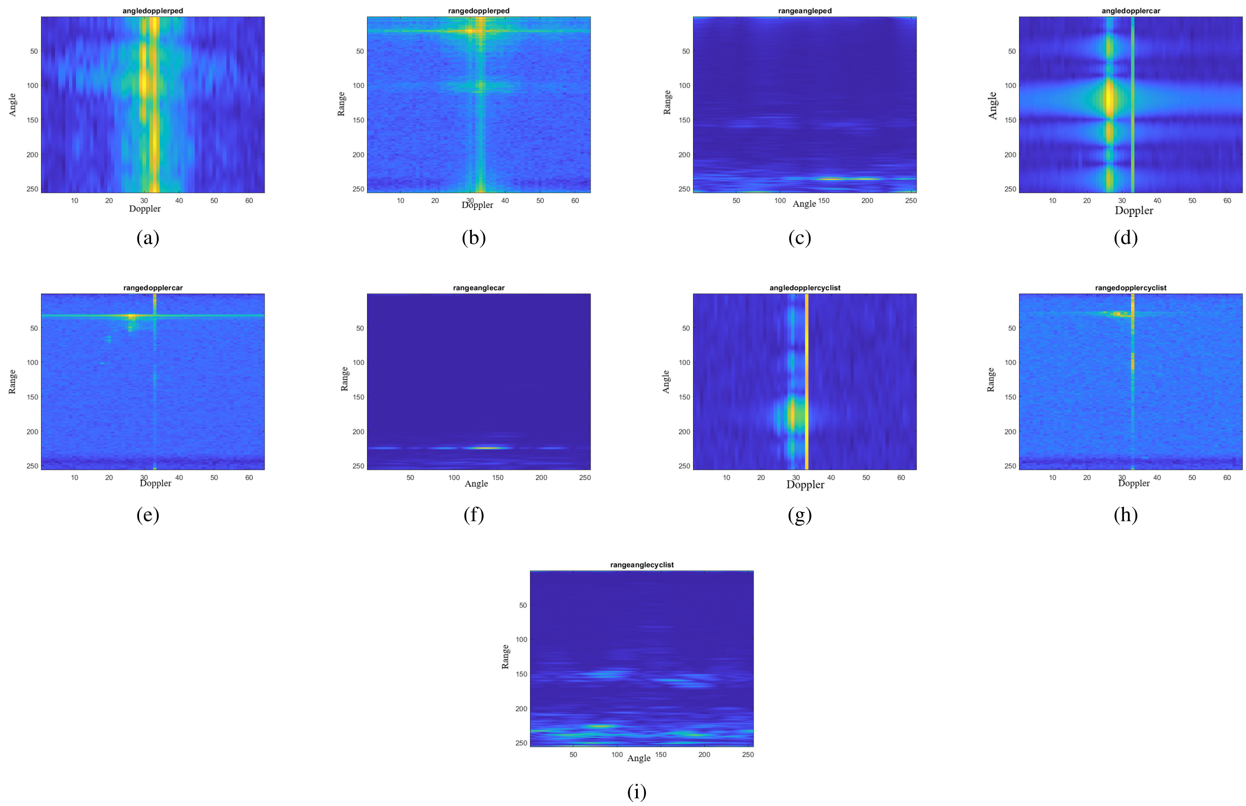


FIGURE 3. Radar maps (a) Pedestrian Angle Doppler Representation (b) Pedestrian Range Doppler Representation (c) Pedestrian Range Angle Representation (d) Car Angle Doppler Representation (e) Car Range Doppler Representation (f) Car Range Angle Representation (g) Bicyclist Angle Doppler Representation (h) Bicyclist Range Doppler Representation (i) Bicyclist Range Angle Representation.

Angle-FFT along the depth to resolve the angle between two objects as shown in fig. 2.

As shown in Fig. 2, fast Fourier transform (FFT) is applied to the ADC data within one chirp to retrieve distance information. A range-Doppler spectrum is obtained by applying a second FFT over the chirp index to measure the phase difference between chirps and deduce the Doppler shift. Finally, in the antenna dimension, a third FFT (or angle FFT) or more complex methods are incorporated to extract angle information and construct the Range-Angle-Doppler tensor. The RAD tensor is too complex to compute, targets on the RD spectrum are often discovered using peak detection methods. The Radar reflections are then obtained by the use of angle FFT or beam forming algorithms, as well as various post-processing procedures.

B. RAW RADAR DATA

Raw data is noisier, however, contain information on every reflected object in the scene. Raw Radar data is advantageous for the purpose of detection as no information is lost. Raw Radar information could be represented as RD, RA and AD maps. RD maps represents the velocity information of moving objects around the vehicle. RA maps offer angle information, allowing you to identify targets surrounding vehicle. AD presents the angle information in respect to the velocity of moving objects. The RD, RA and AD maps reveal distinctive patterns for pedestrian, cars and bicyclists, as can be observed in the Fig 3. Deep learning based model could be trained to detect vulnerable road users, for instance pedestrian and bicyclist. The proposed CNN models is elucidated in the next section.

C. NETWORK ARCHITECTURE DESIGN

The CNN model is trained for RD, RA and AD maps respectively as shown in fig. 4. Five convolutional layers and one fully connected layer make up the CNN network. A batch normalization layer, a rectified linear unit (ReLU) activation layer, and a max pooling layer are placed after the first four convolution layers. The max pooling layer is swapped out for an average pooling layer in the final convolution layer. After softmax activation, the output layer is a classification layer. The same network architecture is utilized for all three of the CNN models, the only difference being the size of the input layer. The input layer size of the RD and AD network model is 256×64 and the input layer size for RA is 256×256. The kernel size of the first convolutional layer is 6×6, while the kernel size of the other convolutional layers is 3×3.

In late fusion approach, the data for each modality is processed separately, with each classifier returning a decision for the respective modality. In the proposed method RD-

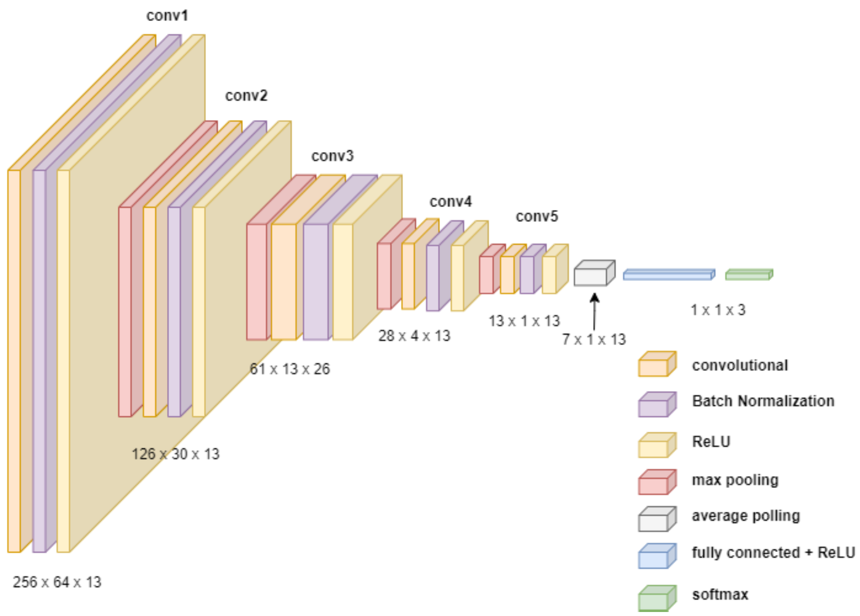


FIGURE 4. Proposed deep network architecture.

CNN, RA-CNN and AD-CNN return a decision regarding the object detected, eventually decision level fusion is performed to fuse the decision taken by a combination of the CNNs for best result. The late fusion is applied of the predicted response of RD-CNN and RA-CNN, RD-CNN and AD-CNN, RA-CNN and AD-CNN and RD-CNN, RA-CNN and AD-CNN. The fusion function can be expressed as in eqs. 2-5.

$$F_a(d) = Res - RA - CNN(d) * Res - AD - CNN(d) \quad (2)$$
$$F_b(d) = Res - RD - CNN(d) * Res - AD - CNN(d) \quad (3)$$
$$F_c(d) = Res - RA - CNN(d) * Res - RD - CNN(d) \quad (4)$$
$$F_d(d) = Res - RA - CNN(d) * Res - AD - CNN(d) * Res - RD - CNN(d) \quad (5)$$

Equation 2 represents the fusion function for RA and AD, where the decision layers are fused to combine the result by the RA network and the AD network. Equation 3 represents the fusion function for RD and AD, where the decision layers are fused to combine the result by the RD network and the AD network. Equation 4 represents the fusion function for RA and RD, where the decision layers are fused to combine the result by the RA network and the RD network. Equation 5 represents the fusion function for RD, RA and AD, where the decision layers are fused to combine the result by the RD network, RA network and the AD network.

Fig. 5 represents all possible combinations of RD-CNN, RA-CNN and AD-CNN under consideration. The F-ROADNET model aims at achieving the highest accuracy with the least amount of complexity. The different models for F-ROADNET is RA-RD fusion network, RA-AD fusion network, RD-AD fusion network and RD-RA-AD fusion network. Experimentation would prove the best possible architecture for road object detection.

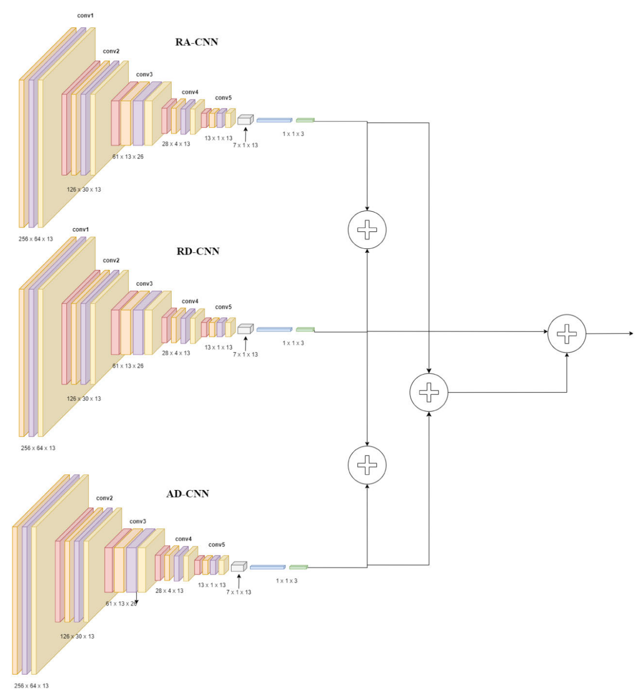


FIGURE 5. Proposed fusion based automotive radar object detection deep network (F-ROADNET) architecture.

IV. EXPERIMENTATION

A. DATASETS

The developed method was evaluated on the CARRADA [14] and RADDet [15] datasets. Scenarios involving automobiles, pedestrians, and cyclists have been documented. All vehicles

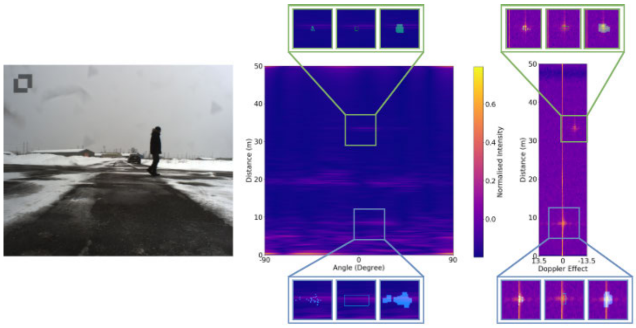


FIGURE 6. CARRADA [14] data representation.

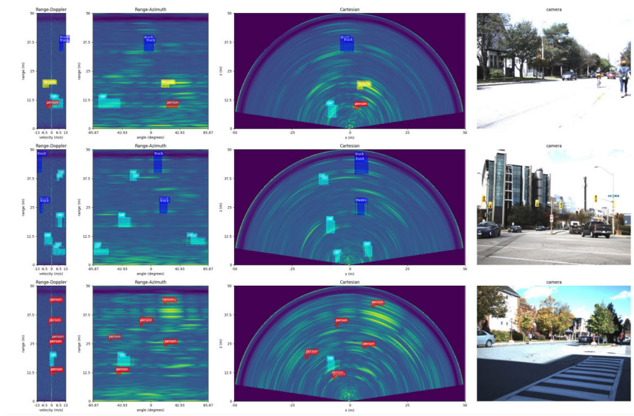


FIGURE 7. RADDet [15] data representation.

such as car, truck and bus have been combined under the class automobile. The RD and AD representations are saved as 2D matrices of size 256 × 64 and RA representation is saved as 256 × 256 matrix, example of CARRADA and RADDet data is shown in fig 6 and 7 respectively. In the CARRADA dataset, to mimic urban driving scenarios, one or two objects move in the area at the same time with different trajectories. The RADDet dataset is complex in comparison to the CARRADA dataset due to multiple objects in each frame and higher chances of occlusion.

The CARRADA and RADDet datasets are developed on python platform and our proposed method is developed on MATLAB. Hence, for this study the data set has been made compatible, so it could be accessed through the training platform. Secondly the Radar tensor data is stored in image format for training purpose. The dataset is randomly divided into training and testing sets, 70% of the data is reserved for training and 30% for testing.

B. TRAINING PARAMETERS

To put the developed method to the test, the model was trained using a system equipped with an Intel i7 7700HQ quad processor, 16 GB of RAM, and an NVIDIA GeForce GTX 1050 GPU. The Adam optimizer is employed with the suggested settings. All the experiments employ a learning rate of 1e-2. The minimum batch size is set to 128 and thirty epochs were used to train the model. The learn rate is schedule as piece-wise, learning rate drop factor is set 0.1 and learning rate drop period is 10 with data shuffle at every epoch.

V. RESULT AND DISCUSSION

Table 1 represents the accuracy, recall and precision of ROADNET and F-ROADNET. ROADNET is trained on singular Radar representation on the other hand F-ROADNET is fusion network and implements decision level fusion.

ROADNET trained on CARRADA dataset shows that RA outperforms the other features with an accuracy, precision and recall of 97.5%, 96.33% and 95.9% respectively. This proves that Range Angle carries information to accurately classify cars, pedestrians and bicyclists. ROADNET trained on either AD or RD perform similarly with an accuracy of around 93% and precision of around 89%. ROADNET trained on RADDet dataset shows similar results, RA maps have the highest accuracy of 87.574%, followed by RD with an accuracy of 85.293%.

As expected, F-ROADNET outperformed ROADNET in terms of accuracy, precision and recall for both CARRADA and RADDet. However, to achieve this performance model complexity is introduced. On the CARRADA dataset, in terms of accuracy F-ROADNET trained on RD, RA and AD performs the best with an accuracy of 99.18%, while F-ROADNET trained on RA and RD comes a close second with accuracy of 99.02%. Similar results are demonstrated by F-ROADNET trained on RADDet. In terms of accuracy network trained on RD, RA and AD has an accuracy of 92.213%, the highest accuracy achieved on the RADDet dataset. Network trained on RA and RD achieve an accuracy of 91.622%, Even though highest accuracy is achieved by the fusion of RA, RD and AD maps but, F-ROADNET trained on RD and RA is preferred due it credible performance compared to its complexity.

As pedestrian and bicyclist are more vulnerable road users, it of utmost importance to detect them accurately and precisely. Table 2 represents the average precision and recall of pedestrian, bicyclist and cars for both ROADNET and F-ROADNET. As could be observed in Table 2 that cars are easier to detect in Radar data, however, pedestrian and bicyclist are comparatively challenging to detect accurately.

ROADNET trained on RD has the lowest precision and recall, while ROADNET trained on RA maps performs considerably better. However due to the vulnerable nature of the road users higher precision is needed. F-ROADNET trained on RD and RA maps shows the most promising results with precision of 0.993, 0.951 and 0.99 for pedestrians, bicyclist and cars respectively on the CARRADA dataset and precision of 0.913 for pedestrians on the RADDet dataset.

Bicyclists prove to the most challenging class with the least precision reported on ROADNET trained on RD maps. On the other hand, F-ROADNET show promising results with precision of 0.951 on CARRADA dataset and 0.846 on RADDet dataset.

TABLE 1. Validation accuracy, recall and precision of ROADNET and F-ROADNET.

Dataset	Network	Input	Accuracy (%)	Precision	Recall
CARRADA	ROADNET	RD	93.4973	0.893	0.902
		RA	97.4863	0.9633	0.959
		AD	93.8251	0.8977	0.9127
	F-ROADNET	RD & RA	**99.0164**	0.9463	0.97
		RD & AD	97.377	**0.9813**	**0.987**
		RA & AD	98.5792	0.9727	0.9817
		RD & RA & AD	**99.1803**	**0.9827**	**0.991**
RADDet	ROADNET	RD	85.293	0.824	0.861
		RA	87.574	0.839	0.855
		AD	84.491	0.809	0.855
	F-ROADNET	RD & RA	**91.622**	**0.872**	**0.902**
		RD & AD	86.984	0.819	0.825
		RA & AD	90.295	0.856	0.841
		RD & RA & AD	**92.213**	**0.912**	**0.918**

TABLE 2. Class-wise recall and precision of ROADNET and F-ROADNET.

Dataset	Method	Input	Precision			Recall		
			pedestrian	bicyclist	car	pedestrain	bicyclist	car
RADDet	ROADNET	RD	0.841	0.612	0.873	0.743	0.629	0.852
		RA	0.874	0.724	0.844	0.751	0.721	0.849
		AD	0.878	0.631	0.859	0.764	0.721	0.849
	F-ROADNET	RD & RA	**0.913**	**0.846**	0.878	**0.792**	**0.742**	**0.884**
		RA & AD	0.889	0.748	**0.884**	0.783	0.722	0.872
		RD & AD	0.887	0.759	0.879	0.779	0.734	0.865
		RD RA AD	**0.928**	**0.851**	**0.902**	**0.798**	**0.749**	**0.892**
CARRADA	ROADNET	RD	0.938	0.754	0.987	0.903	0.806	0.997
		RA	0.97	0.928	0.992	0.976	0.907	0.994
		AD	0.951	0.761	0.981	0.912	0.841	0.985
	F-ROADNET	RD & RA	**0.993**	**0.951**	**0.999**	**0.981**	**0.98**	**0.999**
		RA & AD	0.99	0.928	**0.999**	0.975	0.972	0.998
		RD & AD	0.988	0.852	0.999	0.946	0.966	0.998
		RD RA AD	**0.997**	**0.951**	**0.999**	**0.981**	**0.992**	**0.999**

TABLE 3. Comparison with existing methods.

Reference	Input	Average Precision			
		Pedestrian	Bicyclist	Car	Overall
RODNET [44]	RA	0.88	0.84	0.93	0.8
Ramp CNN [45]	RAD	0.89	0.86	0.93	0.89
Bivar Cross [48]	RAD	0.89	0.93	0.95	0.93
MLP-22 [49]	RAD	0.90	0.84	0.91	0.88
ROADNET (This Study)	RA	0.97	0.928	0.992	0.9633
F-ROADNET (This Study)	RA&RD	**0.993**	**0.951**	**0.999**	**0.987**

It can be seen in the results that the complex nature of the RADDet dataset effects the performance of the developed method, still the developed methods show promising results in the presence of occulusions, as can be experienced in the RADDet dataset.

Comparisons with already-developed techniques [44], [45], [48], [49] were done to assess the developed method's potential. The developed method performs better than the existing method as can be seen in table 3. Pedestrian are detected with average precision of 0.993 by the developed F-ROADNET whereas RODNET [44], Ramp CNN [45], Bivariant Cross attention model [48] and MLP-22 [49] achieved average precision of 0.88, 0.89, 0.91 and 0.90 respectively. Bicyclist have been the most challenging to correctly classify, even so, F-ROADNET achieve average precision of 0.951, exceeding the average precision accomplished by the other methods. Car have been the easiest to detect (due to its unique signatures) with an average precision of 0.999, almost perfect classification. It can be observed in table 3 that RODNET [44], Ramp CNN [45], Bivariant Cross attention model [48] and MLP-22 [49] also perform considerably better on car classification. However, cars are lower on the vulnerable road user scale and higher emphasis has to be given to pedestrians and bicyclists. Even though Bivariant Cross attention model [48] show comparable results on bicyclist detection with an average precision of 0.93, the developed method outperforms it with an average precision of 0.951. In the existing methods MLP-22 [49] shows decent performance, with an average precision of 0.90, F-ROADNET has an improvement of 9.3% over MLP-22. In the context of average overall precision, the developed method has an improvement of 18% over [44], 9.7% over [45], 5.7% over [48] and 10.7% improvement over [49].

The developed novel method proposed here offers late fusion based Radar object detection. RA and RD maps are incorporated with computer vision based deep neural networks to achieve remarkable results. Vulnerable road users were detected with high precision.

VI. CONCLUSION AND FUTURE WORK

Object detection is critical in many fields, including autonomous driving. For decades, the computer vision community has focused on this subject and has produced several useful solutions. However, vision-based detection is still plagued by a number of serious issues. This research offered a brand-new and innovative object recognition system based solely on Radar data that is more robust than vision. The suggested ROADNet could recognize objects correctly and robustly in a variety of autonomous driving scenarios, including at night or in bad weather. Furthermore, this article developed a unique fusion-based model that may advance the function of Radar in autonomous driving applications.

Occlusion phenomena make tracking difficult as Doppler representation may vanish. It might be better to find these occlusions in the video frames and incorporate them into the tracking procedure. It is also challenging to distinguish when objects have similar radial velocities and are close to one another. Investigation is in progress on optimization of detection method. Computational load and complexity analysis for real time application are a vital part of the future work.

REFERENCES

[1] G. Sikander and S. Anwar, "A novel machine vision-based 3D facial action unit identification for fatigue detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2730–2740, May 2021.

[2] X. Zhao, J. Zhang, and W. Song, "A radar-nearest-neighbor based data-driven approach for crowd simulation," *Transp. Res. C, Emerg. Technol.*, vol. 129, Aug. 2021, Art. no. 103260. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X21002722

[3] J. Ku, M. Mozifian, J. Lee, A. Harakeh, and S. L. Waslander, "Joint 3D proposal generation and object detection from view aggregation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 1–8.

[4] Y. Zhou and O. Tuzel, "VoxelNet: End-to-end learning for point cloud based 3D object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2018, pp. 4490–4499.

[5] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NIPS)*. Red Hook, NY, USA: Curran Associates, 2017, pp. 5105–5114.

[6] S. Shi, X. Wang, and H. Li, "PointRCNN: 3D object proposal generation and detection from point cloud," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 770–779.

[7] S. Shi, Z. Wang, J. Shi, X. Wang, and H. Li, "From points to parts: 3D object detection from point cloud with part-aware and part-aggregation network," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2647–2664, Aug. 2021.

[8] B. R. Mahafza, *Radar Signal Analysis and Processing Using MATLAB*. London, U.K.: Chapman & Hall, 2016.

[9] N. Scheiner, F. Kraus, F. Wei, B. Phan, F. Mannan, N. Appenrodt, W. Ritter, J. Dickmann, K. Dietmayer, B. Sick, and F. Heide, "Seeing around street corners: Non-line-of-sight detection and tracking in-the-wild using Doppler radar," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2020, pp. 2065–2074.

[10] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Gläser, F. Timm, W. Wiesbeck, and K. Dietmayer, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1341–1360, Mar. 2021.

[11] M. Meyer, G. Kuschk, and S. Tomforde, "Graph convolutional networks for 3D object detection on radar data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 3053–3062.

[12] C. Decourt, R. VanRullen, D. Salle, and T. Oberlin, "DAROD: A deep automotive radar object detector on range-Doppler maps," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2022, pp. 112–118.

[13] A. Ouaknine, A. Newson, P. Pérez, F. Tupin, and J. Rebut, "Multi-view radar semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*. Los Alamitos, CA, USA: IEEE Computer Society, Oct. 2021, pp. 15651–15660.

[14] 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)*, Jan. 2021, pp. 5068–5075.

[15] 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)*. Los Alamitos, CA, USA: IEEE Computer Society, May 2021, pp. 95–102.

[16] Y. Wang, G. Wang, H.-M. Hsu, H. Liu, and J.-N. Hwang, "Rethinking of radar's role: A camera-radar dataset and systematic annotator via coordinate alignment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2021, pp. 2809–2818.

[17] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 3354–3361.

[18] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 304–311.

[19] C. G. Keller, M. Enzweiler, and D. M. Gavrila, "A new benchmark for stereo-based pedestrian detection," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 691–696.

[20] A. Eitel, J. T. Springenberg, L. Spinello, M. Riedmiller, and W. Burgard, "Multimodal deep learning for robust RGB-D object recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 681–687.

[21] P. Viola, M. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, vol. 2, Oct. 2003, pp. 734–741.

[22] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[23] M. Pedersoli, J. Gonzàlez, X. Hu, and X. Roca, "Toward real-time pedestrian detection based on a deformable template model," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 355–364, Feb. 2014.

[24] T. Liu, W. Luo, L. Ma, J.-J. Huang, T. Stathaki, and T. Dai, "Coupled network for robust pedestrian detection with gated multi-layer feature extraction and deformable occlusion handling," *IEEE Trans. Image Process.*, vol. 30, pp. 754–766, 2021.

[25] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Mali. Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2014, pp. 580–587.

[26] J. Li, X. Liang, S. Shen, T. Xu, J. Feng, and S. Yan, "Scale-aware fast R-CNN for pedestrian detection," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 985–996, Apr. 2018.

[27] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, vol. 28, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2015.

[28] I. Delibasoglu, I. Kosesoy, M. Kotan, and F. Selamet, "Motion detection in moving camera videos using background modeling and FlowNet," *J. Vis. Commun. Image Represent.*, vol. 88, Oct. 2022, Art. no. 103616.

[29] Y. Zhang, W. Zheng, K. Leng, and H. Li, "Background subtraction using an adaptive local median texture feature in illumination changes urban traffic scenes," *IEEE Access*, vol. 8, pp. 130367–130378, 2020.

[30] M. Sultana, A. Mahmood, and S. K. Jung, "Unsupervised moving object detection in complex scenes using adversarial regularizations," *IEEE Trans. Multimedia*, vol. 23, pp. 2005–2018, 2021.

[31] J. Xie, Y. Pang, M. H. Khan, R. M. Anwer, F. S. Khan, and L. Shao, "Mask-guided attention network and occlusion-sensitive hard example mining for occluded pedestrian detection," *IEEE Trans. Image Process.*, vol. 30, pp. 3872–3884, 2021.

[32] F. Wang and M. Lu, "Hamiltonian Monte Carlo estimator for abrupt motion tracking," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 3066–3069.

[33] W. K. Fong et al., "Panoptic nuScenes: A large-scale benchmark for LiDAR panoptic segmentation and tracking," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 3795–3802, Apr. 2022, doi: 10.1109/LRA.2022.3148457.

[34] J. Wu, Y. Tamura, Y. Wang, H. Woo, A. Moro, A. Yamashita, and H. Asama, "Smartphone zombie detection from LiDAR point cloud for mobile robot safety," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2256–2263, Apr. 2020.

[35] X. Chen, J. Liu, J. Wu, C. Wang, and R. Song, "LoPF: An online LiDAR-only person-following framework," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–13, 2022.

[36] K. Qian, S. Zhu, X. Zhang, and L. E. Li, "Robust multimodal vehicle detection in foggy weather using complementary LiDAR and radar signals," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 444–453.

[37] S. Capobianco, L. Facheris, F. Cuccoli, and S. Marinai, "Vehicle classification based on convolutional networks applied to FMCW radar signals," in *Traffic Mining Applied to Police Activities*, F. Leuzzi and S. Ferilli, Eds. Cham, Switzerland: Springer, 2018, pp. 115–128.

[38] K. Patel, K. Rambach, T. Visentin, D. Rusev, M. Pfeiffer, and B. Yang, "Deep learning-based object classification on automotive radar spectra," in *Proc. IEEE Radar Conf. (RadarConf)*, Apr. 2019, pp. 1–6.

[39] B. Major, D. Fontijne, A. Ansari, R. T. 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 *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 924–932.

[40] Y. Jin, Y. Kuang, M. Hoffmann, C. Schüßler, A. Deligiannis, J.-C. Fuentes-Michel, and M. Vossiek, "Radar and LiDAR deep fusion: Providing Doppler contexts to time-of-flight LiDAR," *IEEE Sensors J.*, vol. 23, no. 20, pp. 25587–25600, Oct. 2023.

[41] B. Xu, X. Zhang, L. Wang, X. Hu, Z. Li, S. Pan, J. Li, and Y. Deng, "RPFA-Net: A 4D RaDAR pillar feature attention network for 3D object detection," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, Sep. 2021, pp. 3061–3066.

[42] M. Liang, B. Yang, Y. Chen, R. Hu, and R. Urtasun, "Multi-task multi-sensor fusion for 3D object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7337–7345.

[43] Y. Wang, W.-L. Chao, D. Garg, B. Hariharan, M. Campbell, and K. Q. Weinberger, "Pseudo-LiDAR from visual depth estimation: Bridging the gap in 3D object detection for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8437–8445.

[44] 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.

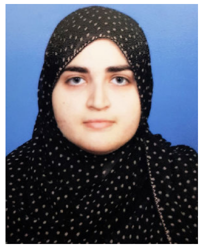
[45] X. Gao, G. Xing, S. Roy, and H. Liu, "RAMP-CNN: A novel neural network for enhanced automotive radar object recognition," *IEEE Sensors J.*, vol. 21, no. 4, pp. 5119–5132, Feb. 2021.

[46] T. Jiang, L. Zhuang, Q. An, J. Wang, K. Xiao, and A. Wang, "T-RODNet: Transformer for vehicular millimeter-wave radar object detection," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.

[47] Y. Jin, A. Deligiannis, J.-C. Fuentes-Michel, and M. Vossiek, "Cross-modal supervision-based multitask learning with automotive radar raw data," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 4, pp. 3012–3025, Jan. 2023.

[48] R. Kothari, A. Kariminezhad, C. Mayr, and H. Zhang, "Raw radar data based object detection and heading estimation using cross attention," 2022, *arXiv:2205.08406*.

[49] J. Bai, S. Li, B. Tan, L. Zheng, L. Huang, and L. Dong, "Traffic participants classification based on 3D radio detection and ranging point clouds," *IET Radar, Sonar Navigat.*, vol. 16, no. 2, pp. 278–290, Feb. 2022, doi: 10.1049/rsn2.12182.



GULBADAN SIKANDER received the Ph.D. degree in machine vision-based intelligent transportation. She is currently a Lecturer with the Department of Mechatronics Engineering, University of Engineering and Technology, Peshawar, Pakistan. Her research interests include intelligent transportation systems, machine vision, artificial intelligence, and machine learning.

SHAHZAD ANWAR received the Ph.D. degree from UWE Bristol, Frenchay Campus, U.K. He is currently an Associate Professor with the Department of Mechatronics Engineering, University of Engineering and Technology, Peshawar, Pakistan. His work focuses on computer vision and artificial intelligence with particular attention to innovative intelligent system techniques. He is a C.Eng.

GHASSAN HUSNAIN received the B.E. degree, the M.Sc. degree in network systems from The University of Sunderland, England, in 2011, and the Ph.D. degree in intelligent transportation systems (ITS) from the University of Engineering and Technology, Peshawar, Pakistan, in 2021. He is currently an Associate Professor with the Department of Computer Science, CECOS University of IT and Emerging Sciences, Peshawar. His research interests include intelligent systems, ad hoc networks, evolutionary computation, bio-inspired algorithms, soft computing, and artificial intelligence.

SANGSOON LIM (Member, IEEE) received the Ph.D. degree in electrical engineering and computer science from Seoul National University, South Korea, in 2013. From 2013 to 2017, he was a Senior Engineer with Samsung Research, Seoul, South Korea. He is currently an Assistant Professor with the Department of Computer Engineering, Sungkyul University, Anyang, South Korea. His current research interests include wireless/mobile networks, the Internet of Things, network management, resource optimization, machine learning, and big data analysis.