IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 April 2023, accepted 14 May 2023, date of publication 29 May 2023, date of current version 2 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3280544

RESEARCH ARTICLE

A Novel Method for Improving Point Cloud Accuracy in Automotive Radar Object Recognition

GUOWEI LU[1], ZHENHUA HE[1], SHENGKAI ZHANG[1], YANQING HUANG[2], YI ZHONG[1], ZHUO LI[3], AND YI HAN[1], (Member, IEEE)

[1]School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China
[2]Bigdata Operation and Information Technology Department, SAIC-GM-Wuling Automobile Company Ltd., Liuzhou 545007, China
[3]Plan and Operation Department, SAIC-GM-Wuling Automobile Company Ltd., Liuzhou 545007, China

Corresponding authors: Yi Zhong (zhongyi@whut.edu.cn) and Yi Han (hanyi@whut.edu.cn)

This work was supported in part by the Research Project of the Wuhan University of Technology Chongqing Research Institute under Grant YF2021-06, and in part by the National Natural Science Foundation of China under Grant 61801341.

ABSTRACT High-quality environmental perceptions are crucial for self-driving cars. Integrating multiple sensors is the predominant research direction for enhancing the accuracy and resilience of autonomous driving systems. Millimeter-wave radar has recently gained attention from the academic community owing to its unique physical properties that complement other sensing modalities, such as vision. Unlike cameras and LIDAR, millimeter-wave radar is not affected by light or weather conditions, has a high penetration capability, and can operate day and night, making it an ideal sensor for object tracking and identification. However, the longer wavelengths of millimeter-wave signals present challenges, including sparse point clouds and susceptibility to multipath effects, which limit their sensing accuracies. To enhance the object recognition capability of millimeter-wave radar, we propose a GAN-based point cloud enhancement method that converts sparse point clouds into RF images with richer semantic information, ultimately improving the accuracy of tasks such as object detection and semantic segmentation. We evaluated our method on the CARRADA and nuScenes datasets, and the experimental results demonstrate that our approach improves the object classification accuracy by 11.35% and semantic segmentation by 4.88% compared to current state-of-the-art methods.

INDEX TERMS Automotive radar, point clouds, GAN, object recognition.

I. INTRODUCTION

Advanced driving assistance systems (ADAS) rely on onboard sensors to acquire environmental data and generate assisted decisions. Optical cameras and LIDAR sensors are commonly used for scene recognition [1], [17]. But optical cameras may suffer from darkness, glare, rain, and fog, while LIDAR sensors cannot penetrate small particulate matter, leading to degraded accuracy in adverse weather conditions. In contrast, millimeter-wave radar has a larger wavelength and greater signal penetration, making it highly robust in harsh environments, unaffected by light, and capable of detecting obscured objects. Therefore, millimeter-wave radar can operate in all weather conditions and offer better detection capability than optical cameras and LIDAR sensors.

Conventional millimeter-wave radars use two types of data representation: radio frequency (RF) images and point clouds. To generate RF images, the raw radar signal undergoes a series of Fast Fourier Transforms (FFTs), and point clouds are derived from these images using peak detection algorithms such as Constant Virtual Alert (CFAR) [6]. A millimeter-wave radar point cloud consists of two-dimensional spatial coordinates (x, y) and the Doppler velocity in the third dimension. While LiDAR and

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

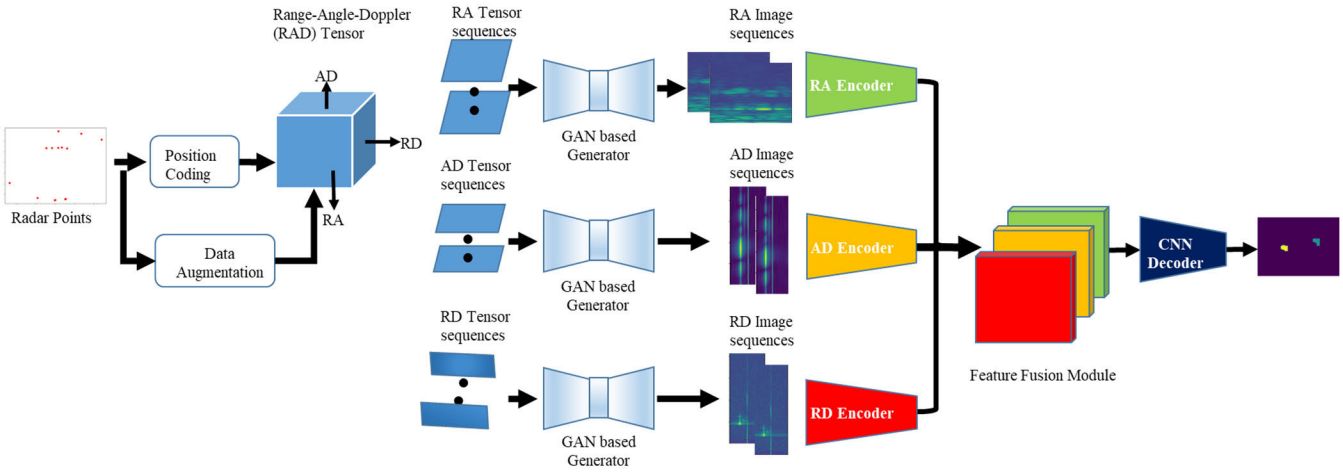

FIGURE 1. Overview of our RF-GAN approach to semantic segmentation of radar point clouds. The RF image corresponding to the radar point cloud is generated by fitting the GAN network, and then the features of each RF image are extracted by different encoders for feature fusion. Finally, the result is output through the CNN decoder (the background is purple, the cyclist is yellow, and the car is green).

millimeter-wave radar point clouds share a similar format, the latter tends to be sparser. For instance, a vehicle may have less than 10 reflection points, as opposed to the 200-300 points typically generated by LiDAR. Consequently, applying algorithms such as PointNet [19], Voxelnet [31], and PointPillars [34], which work well with dense LiDAR point clouds, to millimeter-wave radar solutions presents challenges. For instance, PointNet can accurately capture the local structure and geometric shape of the vehicle from the dense point clouds collected by LiDAR and make judgments accordingly. However, the point cloud data collected by millimeter-wave radar only contains a corner of the vehicle, and other positions are unknown. Therefore, PointNet cannot accurately infer the shape and category of the vehicle from sparse point clouds. Moreover, due to the influence of multipath effects, there are many noise points in the point clouds of millimeter-wave radar, which results in many false positives in the prediction results of PointNet. Researchers have noted that RF images from automotive radar provide richer information for object detection and semantic segmentation than point clouds [2], [8], [9]. However, RF images contain significant noise, which can increase neural network complexity and slow down processing speed. Furthermore, RF image acquisition and preservation require specialized equipment. Conversely, radar point clouds offer benefits such as simpler data acquisition, lower noise, and faster processing due to peak detection algorithms like CFAR. As a result, large-scale millimeter-wave radar datasets, such as the nuScenes dataset [16], the Radar Robot Car dataset [3], and Astyx [4], predominantly contain point cloud data.

We propose a new method called RF-GAN to enhance the accuracy of existing point cloud datasets for tasks such as object recognition and semantic segmentation. RF-GAN is based on the relationship between radar RF images and point clouds. With RF-GAN, we convert sparse point clouds into RF images with low noise and richer semantic information, thus increasing the amount of information available. These RF images address the issue of sparse and noisy radar point clouds and are then processed by an image-based neural network to improve the accuracy of tasks such as object detection and semantic segmentation. The RF-GAN comprises two modules: a point cloud encoding module and an RF image generation network. Fig.1 provides an overview of our RF-GAN approach to the semantic segmentation of radar point clouds.

This paper presents the following main contributions:

- Design of a novel millimeter-wave radar point cloud enhancement method called RF-GAN that significantly improves the accuracy of point clouds for tasks such as object recognition and semantic segmentation.
- Proposal of a data augmentation method suitable for radar point clouds to prevent overfitting during the training phase.
- Validation of the proposed RF-GAN model to work robustly in various driving environments, including curbside scenarios and on-road scenarios.

II. RELATED WORK

A. POINT CLOUD-BASED OBJECT DETECTION

In recent years, automotive radar (also known as single-chip millimeter-wave radar) has been widely used in autonomous driving, robotics, and other fields because of its robustness and signal penetration capability. Initially, Schumann et al. [21] used a random forest classifier to classify dynamic objects, while Prez et al. [27] utilized CNN neural networks to classify objects on the road, achieving good accuracy and real-time processing. However, their methods only considered single-frame data, neglecting the significance of temporal information. In dynamic scenarios, such as

interactions between vehicles and pedestrians, utilizing temporal information can help distinguish objects more easily. To accomplish tasks such as object detection and semantic segmentation, some researchers have borrowed networks from LIDAR, optimized them, and input millimeter wave radar point clouds into them [11], [26]. Danzer et al. [33] proposed a 2D vehicle detection method based on PointNet architecture using sparse radar point clouds, which showed promising results in detecting and localizing vehicles in challenging scenarios with varying lighting conditions. However, this approach may have limited generalizability to objects other than cars and may struggle to detect objects with irregular shapes or partial occlusions in sparse point clouds. Despite efforts to overcome hardware limitations, such as the lack of height information in its point cloud, it is still difficult to accurately infer the 3D position of an object using single radar sensors alone. Among the researchers trying to overcome these difficulties, Bansal et al. [18] used dual radars to increase the number of point clouds and accomplished 3D bounding box prediction of the vehicle. However, the manually set thresholds in this algorithm may result in missing many small objects, thus further optimization is required to improve its accuracy. There has also been some work using the fusion of millimeter-wave radar point clouds and vision to increase the accuracy and robustness of object detection [24]. Reference [10] proposed a multimodal sensor fusion and semantic segmentation-based approach to stabilize and accurately detect the 3D position of objects. However, in this approach, the millimeter-wave radar serves as an auxiliary role for distance and speed measurement, which weakens its superior object detection and tracking capabilities. Reference [28] proposes a multi-sensor fusion algorithm for object detection and recognition that fully exploits the strengths of different sensors. By integrating a camera and millimeter-wave radar at the decision-making level, the algorithm enhances the accuracy and robustness of object detection and recognition. However, it is sensitive to factors such as camera accuracy and noise, indicating that further optimization and improvement are needed.

Although information fusion from multiple sensors can improve detection accuracy, it is important to ensure the good performance of individual sensors to maintain robustness in adverse weather conditions or lighting. Millimeter-wave radar can provide high-resolution object information and detect objects that are not visible to other sensors or have low reflectivity, thereby improving the accuracy of the entire information fusion system. Additionally, millimeter-wave radar is not affected by adverse weather conditions, which enhances the robustness of the system.

B. RF IMAGES-BASED OBJECT DETECTION

After researchers identified the sparsity problem of point clouds, they turned to the use of millimeter-wave radar raw RF images for object detection and other related work. Reference [7] illustrates a solution for vehicle detection based on the raw RF images, which operates on the range-velocity-azimuth radar tensor, using a CNN to predict the object. Since the raw RF images are a 3D range- azimuth -Doppler representation, which is cumbersome to process, slicing along a certain dimension is often used to process the raw RF images. Major et al. [29] create 2D views by summing over each axis of the tensor. Their multi-view representation is processed by a single network dedicated to radar object detection. Similarly, Gao et al. [8] preprocessed the RAD tensor into views. The Autoencoder then extracts features from each view, and these features are fused to locate objects in the range-azimuth view.

Range-azimuth has the most intuitive view, using a polar coordinate representation, and is often used for tasks such as object classification, and semantic segmentation. Reference [9] shows a real-time radar object detection network (RODNet) for detecting objects from radar data in range-azimuth image sequence format. And for the range-Doppler representation. Recent engineers have used Doppler spectrograms for vehicle classification [22] and range-angel views for object detection [13]. These studies show the direct use of raw RF images, which exhibit higher performance in object detection and semantic segmentation than the use of point clouds.

C. RADAR DATA SETS

The nuScens dataset [16] is the first large-scale dataset to provide millimeter-wave radar data as well as camera data, however, the radar dataset in it contains only a few dozen unannotated points per frame. the Oxford RobotCar dataset uses 360° scanning radar. However, as with conventional radar, it has limited angular resolution and does not provide Doppler information. Radarscens [20] uses a single camera and high-resolution radar to capture radar point clouds and visual images of multiple real driving scenes. With the intensive research on raw RF images for millimeter-wave radar, most of the newly released datasets contain raw RF images. Rebut et al. [23] have recently released datasets named RADIal includes raw ADC sampling and can be represented as a range-azimuth-Doppler tensor, range-azimuth, and range-Doppler views or point clouds. The CRUW [9] dataset uses radar data in RF image format for road vehicle and pedestrian detection.

The CARRADA dataset [15] uses a camera and car radar to record data from 30 sequences of multiple moving objects in a variety of scenarios. They use a 77 GHz automotive radar with a detection range of up to 50 m, mounted on the front of the vehicle, and each frame of point cloud data is time-stamped to synchronize the data from the camera and the radar. The automotive radar not only provides data on the position, velocity, time, and reflected intensity of the point cloud, but also three-dimensional range-azimuth-Doppler data. This is then annotated using the manual, which is the only publicly available dataset that provides semantic segmentation annotations and bounding box annotations in a

dense, sparse format for both RA, RD and AD views. The data used in our work is derived from this dataset.

In summary, although many scholars have made significant progress in radar point cloud target detection, the sparsity of point cloud remains one of the biggest bottlenecks in its sensing field. Inspired by the use of Conditional GANs by Lu et al. [25] and others to accomplish indoor map reconstruction, we utilized GANs to deal with radar point clouds. Unlike indoor scenarios, outdoor environments pose additional challenges such as increased complexity and point cloud sparsity. Our proposed method, RF-GAN, considers these challenges and provides a solution based on deep GANs.

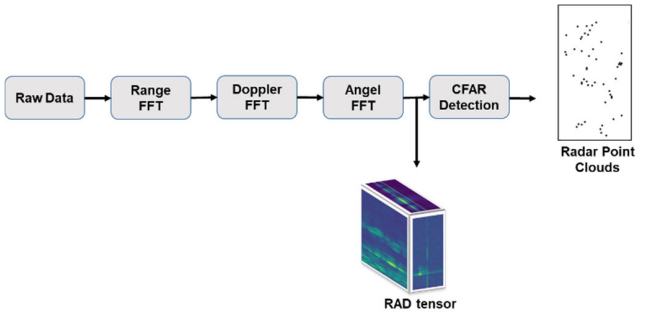

FIGURE 2. Radar signal processing flow.

III. BASIC THEORY

Millimeter-wave radar technology is capable of achieving precise detection capabilities due to its use of a signal with a wavelength of only a few millimeters. Specifically, radar systems operating at frequencies between 76 and 81 GHz can detect movements as small as a fraction of a millimeter. One of the commonly used millimeter-wave technologies is frequency-modulated continuous wave (FMCW), which measures the distance, direction, and speed of an object using a continuous FM signal.

To achieve this, a synthesizer creates a linear FM pulse that is sent out by the transmitter (TX) antenna. The pulse bounces off the object and generates a reflected linear FM pulse that is received by the receiver (RX) antenna. The mixer combines these two signals to produce an intermediate frequency (IF) signal. In a radar system, multiple transmit and receive antennas are typically used, resulting in multiple IF signals that contain information about the object's distance, speed, and direction. The IF signal is then sampled by an ADC and processed by a 3D-FFT algorithm, as illustrated in Fig. 2. The 3D-FFT algorithm utilizes three separate Discrete Fast Fourier Transforms (DFFT) to calculate the distance, Doppler velocity, and azimuth spectra of the object. The angular FFT is performed at the receiver for each cell of the range-velocity spectrum, which is the output of the velocity FFT. In traditional FMCW radars, obtaining the range-Doppler velocity-azimuth tensor is typically computationally complex. Therefore, the Constant False Alarm Rate (CFAR) algorithm is often used to detect objects in the range-Doppler domain, and a sparse point cloud is obtained by beamforming.

However, according to [5], the CFAR method has two major issues. Firstly, it may fail to detect the real object, causing a loss of the original RF image information. Secondly, objects with high reflection intensity, caused by multipath reflections or ground reflections, may not be filtered out, resulting in the presence of ''ghost points'' in the point cloud.

IV. METHOD

To extract more useful information from data containing only point clouds, we propose a new method for generating RF images for automotive radar. Our method consists of two modules the first is a point cloud encoding module and the second is an RF images generation network.

A. THE CHALLENGE: NOISE AND SPARSITY ISSUES

Before delving into the technical aspects, we will first explore the difficulties encountered while converting CFAR-processed point clouds into RF images.

Automotive radar is subject to multipath indoors and outdoors due to beam extension, diffraction, and reflection from surrounding objects so that the receiver antenna receives reflected wave signals from multiple paths, which is the main source of noise and 'ghost spots' in the automotive radar point cloud. As conventional CFAR algorithms rely solely on the intensity of the reflections from the range-Doppler medium cell. However, ground reflections and multipath effects can also cause some high-intensity cells in the range-Doppler to be detected, resulting in many additional noise points in the point cloud. During the process of converting single-frame point clouds into corresponding RF images using the GAN network, we observed that the noise points in the point cloud remained unfiltered and appeared as distinct shapes in the resulting RF images. This phenomenon has the potential to significantly impact the accuracy of our subsequent object detection efforts.

Automotive radar point clouds also have a very serious sparsity problem due to the specular reflection effects of automotive signals and the hardware limitations of automotive radars. Due to the highly specular reflection of automotive signals, only a small fraction of the reflected signal on the surface of the obstacle propagates to the millimeter-wave receiving antenna, resulting in a limited portion of the point cloud representing the obstacle. Furthermore, hardware costs can restrict the ability of automotive radar to differentiate between objects, exacerbating the sparsity issue in the resulting point cloud. Furthermore, to reduce bandwidth and filter out noise, the final point cloud is obtained by processing the RF image using the CFAR algorithm, which further reduces the point cloud density. Converting the sparse point cloud into a more feature-rich RF image poses an even greater challenge to our network.

B. RF-GAN NETWORKS

1) POINT CLOUD POSITION CODING

The point cloud data is inherently disordered and irregularly sampled in 3D space, making it challenging to extract useful

information from the sparse and unstructured point cloud data, which is further exacerbated by the ambiguous ordering problem caused by noise. The resolution in pitch angle is negligible due to the limited number of antennas in automotive radar. To address these challenges, we encode the point cloud into a 2D image by setting the z-axis height to zero and using (1).

$$\begin{aligned} angel &= \frac{arctan(\frac{x}{y})}{\Delta\alpha} \\ range &= \frac{\sqrt{x^2+y^2}}{\Delta\beta} \end{aligned} \tag{1}$$

where (x, y) are the coordinates of the point cloud and α is the angle of arrival of the point cloud, β is the distance of the point cloud, $\Delta\alpha$ is the angular resolution, and $\Delta\beta$ is the distance resolution, these two parameters depend on the hardware of the automotive radar. (range, angel) is the position of the point cloud projected onto the 2D map. To match the height and width of the range-azimuth images stored in the CARRADA dataset, we set the size of the (range, angel) matrix to 256×256, and the values in the matrix corresponding to that point are expressed using the reflection intensity of that point. After projection, we normalized the reflection intensities in it to between [0,1] to facilitate subsequent convergence of the network. As the range-Doppler images contain rich information about the motion of the reflected object, we also project the point cloud into the range-Doppler images, where $doppler = d/\Delta d$, d is the Doppler velocity, and Δd is the velocity resolution. To ensure that the dataset maintains consistency concerning the same range-Doppler images, we have set the size of the matrix representing (range, Doppler) to 256×64. The value of each point in the matrix is represented using the normalized reflection intensity of that point. It is worth noting that we can obtain the third view (AD) from the range-azimuth and range-Doppler images. Therefore, there is no need to encode the azimuth-Doppler images.

2) DATA AUGMENTATION

Data augmentation techniques can significantly improve the size and quality of training datasets, leading to better deep learning models. For LiDAR point clouds, common processing, and enhancement methods include point cloud normalization, random disruption, translation, rotation, scaling, and discarding. However, it's important to note that not all of these methods are directly applicable to radar point clouds. The velocity measured by radar must remain correlated with the angle of the observed object, so certain enhancement techniques, such as point cloud translation, may not be suitable for radar point cloud imaging. Therefore, in radar point cloud imaging, we need to focus on data enhancement operations that are consistent with the unique characteristics of millimeter wave radar imaging. In particular, we should consider three data enhancement operations: superposition, global scaling, and point cloud normalization. These methods can help improve the quality and quantity of the training dataset, leading to more accurate and reliable deep-learning models.

Millimeter-wave radar data is typically sparse, with an average of only 30-40 points per frame and a significant number of interference points. To address this issue, we employ a multi-frame superposition method to obtain denser point clouds. Specifically, we accumulate the multi-frame point cloud from the previous scan into the coordinate system of the current scan each time a new frame is acquired. This method enables us to obtain dense radar point clouds, which in turn improve the performance of object detectors.

We are scaling every point $p(x, y)$ in every direction by a scalar k drawn from a uniform distribution $U(1-t, 1)$, where $t \in \{0.05, 0.1, 0.25\}$ such that an augmented point p^* has the form $p^*(k \cdot x, k \cdot y)$. The scaled point cloud is then projected into the RF image. At the same time, we scale the corresponding RF image labels to the same multiple for training. Both increase the amount of data for training and improve the generalization ability of the model.

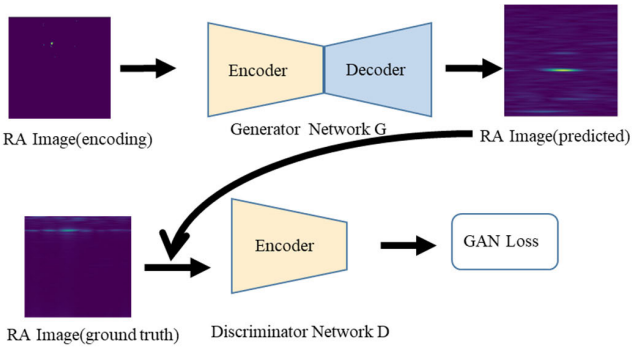

FIGURE 3. Process for generating RA Image. The input of the model is two-dimensional point clouds, and the corresponding RA Image is obtained after location coding and data enhancement.

3) RF-GAN NETWORK ARCHITECTURE

The advancement in GANs has led to a significant improvement in the photorealism of synthesized images in recent years, GANs can now generate high-resolution images of human faces, bodies, animals, cars, and other object classes that are almost indistinguishable from real photographs [32]. GANs work by simultaneously training two neural networks, a generator G and a discriminator D. The generator G takes a noise vector as input and is trained to generate data samples, while the discriminator D is trained to distinguish between real samples and those generated from G. The feedback from the discriminator is then used to improve the generator's performance, leading to better samples, and more effective counteractions against the discriminator. Both networks compete against each other, leading to an iterative process of improving their respective tasks. In our study, we utilize a network of GANs to generate dense RF images from sparse point cloud inputs, as illustrated in Fig. 3. By leveraging the power of GANs, we can achieve more realistic and accurate RF images, which can enhance the performance of our object detection system.

a: GENERATOR NETWORK G

The generator G in our study, as shown in Fig. 4, is structured as an encoder-decoder network. Since the input RF images are stacked along the time dimension, we start by processing the features in the time dimension using a 1×1 convolution. Next, we extract the features of the RF images through multiple convolutional layers with a kernel size of 3×3 and apply batch normalization and activation layers. To compress features between every two convolutional layers, we use a pooling layer with a kernel size of 2×2 and a sliding of 1. In the decoder stage, we use a transposed convolution with a kernel size of 2×2 and a sliding of 1 to gradually convert the feature map to the size of the input RF image. The activation function used in the network is RELU.

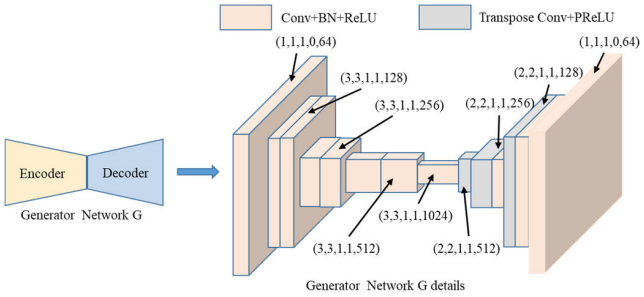

FIGURE 4. Details of Generator network G, The notation (1,1,1,0,64) represents the hyper-parameters of a convolutional layer, with a kernel size of 1 × 1, a stride of 1, a padding of 0, and 64 output channels.

b: DISCRIMINATOR NETWORK D

The discriminator network D is designed similarly to the encoder of the generator G network, with a series of convolutional layers using a 4×4 kernel size to extract deep features from the input. Through the GAN loss function, the discriminator D network learns to distinguish between factual label matrices and predictive label matrices, providing crucial feedback to the generator G network.

c: LOSS FUNCTION

Given an input millimeter-wave shot range-azimuth RF images s transformed from a sparse point cloud, we use GANs to model the distribution of the real RF images d. The loss function of GANs can be expressed as (2)

$$L_{GAN}(G, D) = E_{(s,d)}[\log D(s, d)] + E_s[\log(1 - D(s, G(s)))] \quad (2)$$

where D is the discriminator model and G is the generator model. From the loss function of GAN, we can see that the generator G of GAN wants the output data distribution to be closer to the distribution of the real data, while the discriminator D of GAN needs to make a judgment between the real data and the data output by the generator to find the real data and the data generated by G.

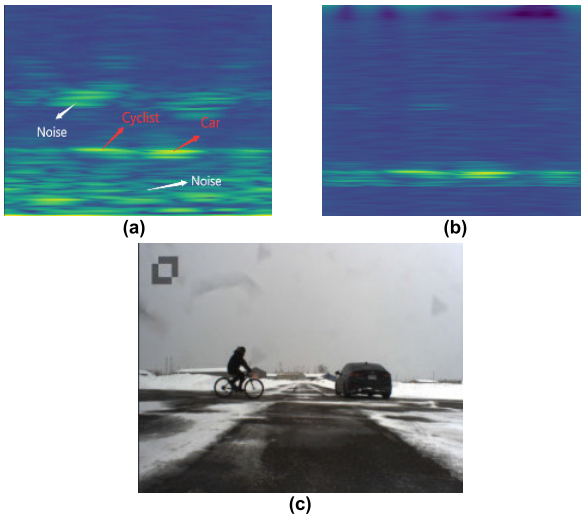

FIGURE 5. (a) RA image (original) with a lot of noise, including ground reflections, interference caused by snow piles (b) RA image (label) with only the reflected part of the object being detected retained (c) Camera image of the current scene.

4) TRAINING LABEL GENERATION

For supervised learning of GANs, it is crucial to generate accurate real labels. However, due to the presence of noise caused by the multipath effect in RF images captured by automotive radars, using the original RA images directly as training labels is not feasible, as shown in Fig. 5. In the presence of snow piles, the radar signals are reflected between the snow piles and the objects, leading to more interference and resulting in the detection of false points in the radar image. Conversely, in the absence of snow piles, the radar signals mainly interact with the objects of interest, resulting in reduced interference from other objects. To overcome this, we adopt a semi-automatic annotation method available in the CARRADA dataset. This method generates a bounding box for each object in the RA, AD, and RD images, and only the relevant part of the bounding box is retained. The resulting transformed RA and RD images are then utilized as training labels. We provide a detailed implementation of RF-GAN in Algorithm 1.

V. EXPERIMENT

A. DATASET AND ASSESSMENT INDICATORS

In this section, we perform an experimental evaluation of our model. We begin by describing the dataset we used and the evaluation metrics. Finally, we give details of the experiments and provide a qualitative and quantitative analysis of the results.

1) DATASET

The CARRADA dataset is a valuable resource as it provides synchronized camera and automotive radar recordings for 30 different sequences, each containing one or two moving objects. This publicly available dataset includes multiple annotations for RD, AD, and RA images, as well as point

Algorithm 1 RF-GAN Algorithm 1mplementation

Requirements: learning rate α, batch size b, number of iterations of discriminator in each generator iteration n, number of training epochs

Requirements: initial parameters δ_{d_0} of the discriminator network D, initial parameters δ_{g_0} of the generator network G

Input: point cloud P(x, y, velocity, energy)

Output: RF images G_{image}

z = **Encoding**(P) // Encoding point cloud
for epochs **do**
  //Train the discriminator network
  for t=0,,,,,n **do**
    Sample $\{l^{(i)}\}_{i=1}^{m} \sim \mathrm{D}_{label}$ a batch from the label
    Sample $\{\mathrm{z}^{(i)}\}_{i=1}^{m} \sim \mathrm{G}_Z$ a batch from the data z
    $\nabla_{\delta_d} \leftarrow \frac{\partial}{\partial \delta d}[\frac{1}{m}\sum_{i=1}^{m} Loss_D(l^{(i)}, z^{(i)}, \delta_d, \delta_g)]$
  //Update discriminator network parameters
    $\delta_d \leftarrow \delta_d - \alpha \cdot Optimizer(\nabla_{\delta_d}, \delta_d)$
  end for
  // Train the generator network
  Sample $\{\mathrm{z(i)}^{(i)}\}_{i=1}^{m} \sim \mathrm{G}_Z$ a batch from the data z
  $\nabla_{\delta g} \leftarrow \frac{\partial}{\partial_{\delta g}}[\frac{1}{m}\sum_{i=1}^{m} Loss_G(z^{(i)}, \delta_d, \delta_g)]$
  //Update the generator network parameters
  $\delta_g \leftarrow \delta_g - \alpha \cdot Optimizer(\nabla_{\delta g}, \delta_g)$
end for
$G_{image} = \mathrm{G(z)}$

cloud coordinates. The objects in the dataset are classified into three categories: pedestrians, cyclists, cars, and others in the background. The RA images in the annotation set have a size of 256 × 256, while the RD and AD images have a size of 256 × 64. To improve the quality of the data for training, we apply data augmentation techniques such as superposition and global scaling to the radar point clouds. The augmented data is saved locally and only used during training to prevent overfitting. The distribution of the data for the training, augmented, and test sets are shown in Table 1.

TABLE 1. Distribution for training and test.

	Augmented data	Training set	Testing set
Frames	6162	7265	5401

2) EVALUATION METRIC

The GAN network used in this study transforms the input point cloud into the corresponding RA and RD images. To assess the quality and diversity of the generated RA and RD images, we use the Fréchet Inception Distance (FID), a classical performance metric that measures the distance between the real image and the Inception feature vector of the generated image. FID provides an integrated characterization of the similarity between the real and generated images, and is calculated using (3).

$$FID(x, g) = ||\mu_x - \mu_g|| + Tr(\sum\nolimits_x + \sum\nolimits_g - 2\sqrt{\sum\nolimits_x \sum\nolimits_g}) \quad (3)$$

where g is the GAN network-generated image, x is the real image, and μ_x is the mean of the features of the real image, Tr is the sum of the elements on the diagonal of the matrix. $\sum_x$ and $\sum_g$ are the covariance matrices of the feature vectors of the real images and the generated images. The FID is more robust to noise and gives a better evaluation of the quality of the generated image, its score is more consistent with human visual judgment, and the computational complexity of the FID is relatively low.

In the image-based semantic segmentation task, we use the joint intersection (IoU) to determine how well the detection results match the actual situation. The joint intersection (IoU): Given the annotated test input, the specified class of IoU is defined as the percentage $\frac{|A \cap B|}{|A \cup B|}$, where A is the set of pixels predicted to originate from that class and B is the true set of pixels at locations in the same class. This can be readily applied to radar point clouds by (4)

$$\mathrm{IoU} = \frac{|\text{predicted points} \cap \text{true points}|}{|\text{predicted points} \cup \text{true points}|} \quad (4)$$

We use average precision (AP) and mean average precision (mAP) to evaluate the performance of object classification as (5) and (6).

$$AP = \frac{tp}{tp + fp} \quad (5)$$

$$mAP = \frac{1}{class}\sum AP \quad (6)$$

where true positive (tp) denotes the result of correct classification, false positive (fp) denotes the result of incorrect classification, class is the number of object classes, and class is 3 in the experiment of this paper.

B. TRAINING AND RESULTS

1) BASELINES

We used TMVA-Net [2] as our backbone network for the object detection task and compared the experimental results of TMVA-Net with the following methods using pure radar point clouds: (1) object classification based on clustering and decision trees [30]. (2) the PointNet-based object classification method proposed by Danzer et al. [33].

2) TRAINING

Due to the different sizes of RA and RD images in the CARRADA dataset, with RA images being 256 × 256 and RD images being 256 × 64, we trained two separate GAN models for each dataset. The training approach for both models was the same, using an alternating iterative training approach where we trained the discriminator network first

while keeping the generator network fixed. For the discriminator network, we treated the problem as a supervised binary classification problem and fed both sets of samples directly into the discriminator network for training. The discriminator was trained to accurately distinguish true samples, outputting a value as close to 1 as possible, from fake samples, outputting a value as close to 0 as possible. Subsequently, the generator network was trained while keeping the discriminator network fixed. During this phase, we used feedback from the discriminator network to guide the generator network in generating images that were as close to the true labels as possible. In the training process using GAN loss, a total of 13427 frames of point clouds from the CARRADA dataset were selected as the training set, while 5401 frames were reserved for the test set. The RMSProp optimizer [14] was used along with a batch size of 32, a learning rate of 0.0001, a learning rate decay of 0.9 per 10 iterations, and 600 rounds of training. Using the pre-training model of the GAN network, we generated RA and RD images corresponding to the 5401 frames in the test set. We split 3544 of these images for TMVA-Net training, while the rest were used for testing. The training process in TMVA-Net utilized the Adam optimizer [12] and recommended parameters ($\beta_1 = 0.9, \beta_2 = 0.9, \varepsilon = 10^{-8}$). The above network was built using the PyTorch framework and implemented using the Python programming language. We conducted all training on a single GeForce RTX 3080 (GPU) graphics card.

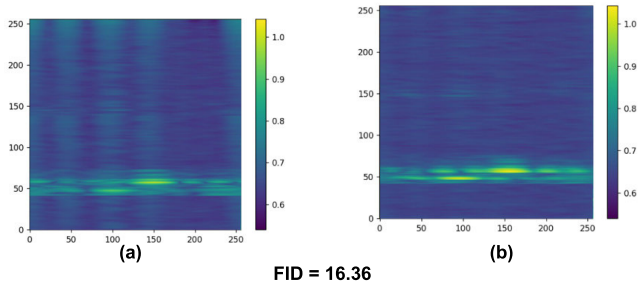

FIGURE 6. (a): RA image(predicted) (b): RA image (ground truth).

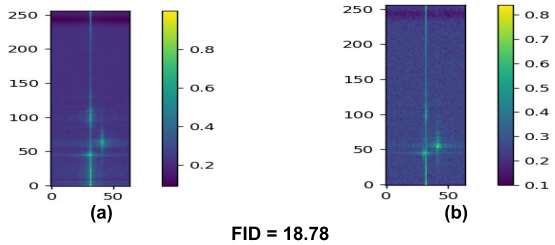

FIGURE 7. (a): RD image(predicted) (b): RD image (ground truth).

3) EVALUATION

In this section, we evaluate each module comprehensively, first we assess the feasibility and accuracy of the designed GAN network, and then compare our results with the PointNet based clustering and decision tree approach in the object classification and semantic segmentation tasks. Fig. 6 and Fig. 7 show the results of predicting RA and RD images using GAN networks with FID evaluation. The camera image presented in Fig. 8 provides context to the scene and aids in understanding the performance of the proposed method. In Table 2, we can see the results of the object detection task on the CARRADA-test set for both point cloud and RAD image data. Noisy points in the radar point cloud can lead to incorrect detection, while the sparsity of the point cloud can result in missed detection of some real objects. Our proposed module converts the point cloud into a RAD image(predicted) for the object detection task and compares it with the direct use of the point cloud. The results show that our method achieves a mAP metric of 62.46% for the RAD image(predicted), which is a 11.35% improvement for the classification task and a 4.88% improvement for the semantic segmentation task compared to the method used by Danzer et al [33]. These results demonstrate the effectiveness and accuracy of our proposed GAN-based approach for object detection tasks in the autonomous driving domain.

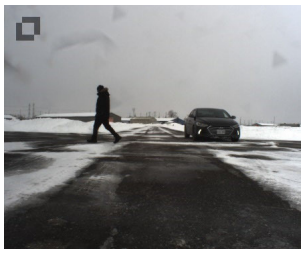

FIGURE 8. Camera image of the current scene.

a: QUALITATIVE RESULTS

In Fig. 9, we present the qualitative results for both methods on the CARRADA test set. Our predicted RAD image (b) generated using the GAN network, after the TMVA-Net output prediction, shows high accuracy in terms of object localization and classification, as can be observed in (d). On the other hand, the result (e) obtained from Danzer et al. indicates that the point cloud mapping cyclists are too sparse, resulting in the misidentification of pedestrians. The predicted result from Decision Tree [30] directly fails to identify cyclists. These experiments clearly demonstrate that our method, which generates automotive radar RF images, significantly improves the accuracy of object detection and semantic segmentation.

To evaluate the generalizability of our method to other automotive radar datasets containing point clouds, we conducted experiments on the nuScenes dataset [16]. Our RF-GAN network is a pre-trained model that can directly process point cloud data from the nuScenes dataset to generate corresponding RA and RD images, which are then fed into the TMVA-Net network to obtain prediction results. As shown in Table 3, the accuracy of all three methods has decreased

TABLE 2. Comparison with state-of-the-art radar-based object detection methods in the carrada dataset.

Method	Radar Input	Classification				Semantic Segmentation			
		AP_{car}	$AP_{cyc.}$	$AP_{ped.}$	mAP	IoU_{car}	$IoU_{cyc.}$	$IoU_{ped.}$	mIoU
Decision Tree[30]	Point Cloud	38.55%	21.67%	31.55%	30.59%	29.81%	14.2%	20.69%	21.57%
Danzer *et al.*[33]	Point Cloud	66.28%	32.42%	54.63%	51.11%	42.1%	19.74%	35.97%	32.60%
Our method	RAD Image(predicted)	79.17%	42.95%	65.26%	62.46%	47.21%	22.55%	42.68%	37.48%
Our method	RAD Image(Raw)	81.42%	46.6%	68.7%	65.57%	50.84%	25.38%	44.7%	40.30%

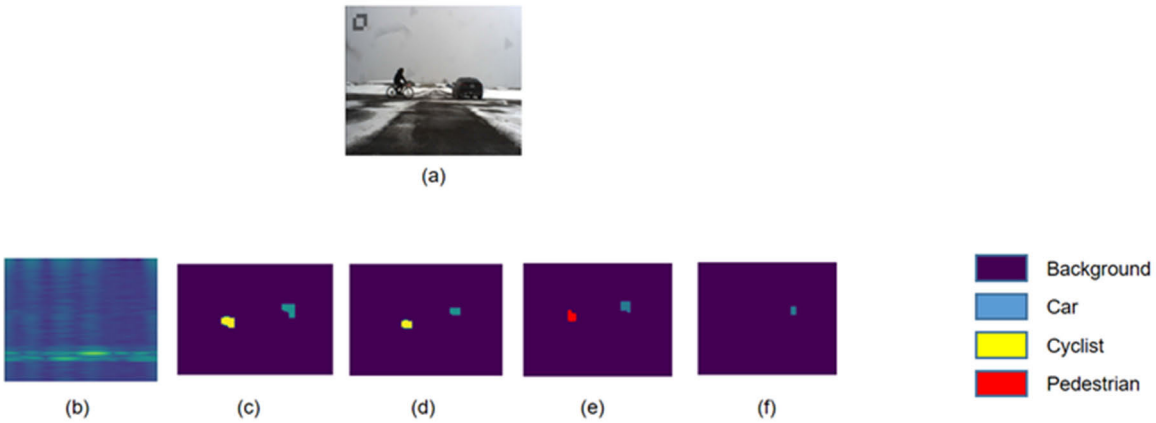

FIGURE 9. Qualitative results in the CARRADA test set. (a) Camera image of the scene, (b) RA view (predicted). (c) Prediction results for real labels, (d) TMVA-Net prediction results (using RAD Image (predicted)) (e) Danzer et al. [33] prediction results (f) Decision Tree [30] prediction results.

due to the more complex test scenarios and the larger number of objects in the nuScenes dataset compared to the CARRADA dataset. As the nuScenes dataset uses a more accurate automotive radar with higher point cloud density, Danzer's method shows a 6% decrease in mAP for the object detection task. RF-GAN network is a pre-trained model with a relatively homogeneous training scenario using the CARRADA dataset, and in the face of complex driving environments, our method can further increase the point cloud information and obtains a 7.31% improvement in mAP relative to the method used by Danzer et al. [33].

TABLE 3. Comparison with state-of-the-art radar-based object detection methods in the nuscens dataset.

Method	Radar Input	AP_{car}	$AP_{cyc.}$	$AP_{ped.}$	mAP
Decision Tree[30]	Point Cloud	33.58%	18.75%	28.70%	27.01%
Danzer *et al.* [33]	Point Cloud	60.38%	28.70%	43.13%	44.07%
Our method	Point Cloud	69.07%	30.66%	54.41%	51.38%

b: TIME AND SPACE COMPLEXITY ANALYSIS

To determine the viability of our proposed autonomous driving method, we analyzed the space and time complexity of RF-GAN, which we summarize in Table 4. Additionally, we compared RF-GAN to two other methods. As shown in Table 4, the decision tree-based approach has the shortest processing time, taking only 10ms due to its reliance on a manually crafted feature vector that does not require neural networks. In contrast, our method, which utilizes point cloud-to-RF image conversion and TMVA-Net for classification, takes 26ms with a total runtime of 55ms. Although our method is 2.5 times longer than the method proposed by Danzer et al. [33], it boasts an increased classification accuracy of 11.35%, which we deem acceptable. Notably, all three object detection algorithms have an inference time of less than the sensor cycle time of 60ms, making real-time processing feasible.

TABLE 4. Computional costs on different methods.

	#params	Runtime (ms)
Decision Tree [30]	0.2M	10
Danzer *et al.* [33]	3.7M	21
RF-GAN(Our method)	4.9M	26
TMVA-Net [2]	5.6M	29

VI. CONCLUSION

In this paper, we proposed the RF-GAN model to address the challenges associated with millimeter wave radar point

clouds in localization and target detection. Our model converts sparse point clouds into rich-information RF images and achieves accurate classification of road targets. Experimental results on multiple datasets demonstrate that our model can effectively enhance sparse radar point clouds in different road environments, with a 11.35% improvement in target classification accuracy compared to traditional methods. This improvement is due to our preprocessing approach, which enhances the semantic information in point clouds based on their unique characteristics. Additionally, we analyzed the temporal and spatial complexity of the RF-GAN model and showed that it can be applied to the field of autonomous driving.

In future work, we will continue to study this enhancement algorithm for radar point clouds, including the time complexity and generalization of the model. Furthermore, we plan to explore the fusion of enhanced RF images with other sensors to achieve more accurate and stable target recognition systems.

REFERENCES

[1] Y. Li and J. Ibanez-Guzman, "LiDAR for autonomous driving: The principles, challenges, and trends for automotive LiDAR and perception systems," *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 50–61, Jul. 2020.

[2] 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)*, Montreal, QC, Canada, Oct. 2021, pp. 15651–15660.

[3] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner, "The Oxford radar RobotCar dataset: A radar extension to the Oxford RobotCar dataset," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Paris, France, May 2020, pp. 6433–6438.

[4] M. Meyer and G. Kuschk, "Automotive radar dataset for deep learning based 3D object detection," in *Proc. 16th Eur. Radar Conf. (EuRAD)*, Paris, France, Oct. 2019, pp. 129–132.

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

[6] A. Coluccia, A. Fascista, and G. Ricci, "CFAR feature plane: A novel framework for the analysis and design of radar detectors," *IEEE Trans. Signal Process.*, vol. 68, pp. 3903–3916, 2020.

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

[8] 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.

[9] Y. Wang, Z. Jiang, Y. Li, J. 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.

[10] M. P. Muresan, I. Giosan, and S. Nedevschi, "Stabilization and validation of 3D object position using multimodal sensor fusion and semantic segmentation," *Sensors*, vol. 20, no. 4, p. 1110, Feb. 2020.

[11] A. Palffy, E. Pool, S. Baratam, J. F. P. Kooij, and D. M. Gavrila, "Multi-class road user detection with 3+1D radar in the view-of-delft dataset," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 4961–4968, Apr. 2022.

[12] P. K. Diederik and J. B. Adam, "A method for stochastic optimization," 2014. [Online]. Available: https://arxiv.org/abs/1801.01489

[13] 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, 2022.

[14] G. Hinton, N. Srivastava, and K. Swersky. (2012). *Neural Networks for Machine Learning Lecture 6a Overview of Mini-Batch Gradient Descent*. Cou-rsera. [Online]. Available: https://www.coursera.org/lecture/neural-networks-deep learning/mini-batch-gradient-descent-KjEKa

[15] 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)*, Milan, Italy, Jan. 2021, pp. 5068–5075.

[16] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "NuScenes: A multimodal dataset for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 11618–11628.

[17] Z. Xu, W. Yang, W. Zhang, X. Tan, H. Huang, and L. Huang, "Segment as points for efficient and effective online multi-object tracking and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6424–6437, Oct. 2022.

[18] K. Bansal, K. Rungta, S. Zhu, and D. Bharadia, "Pointillism: Accurate 3D bounding box estimation with multi-radars," in *Proc. 18th Conf. Embedded Networked Sensor Syst.*, Nov. 2020, pp. 340–353.

[19] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 77–85.

[20] O. Schumann, M. Hahn, N. Scheiner, F. Weishaupt, J. F. Tilly, J. Dickmann, and C. Wöhler, "RadarScenes: A real-world radar point cloud data set for automotive applications," in *Proc. IEEE 24th Int. Conf. Inf. Fusion (FUSION)*, Sun City, South Africa, Nov. 2021, pp. 1–8.

[21] O. Schumann, C. Wöhler, M. Hahn, and J. Dickmann, "Comparison of random forest and long short-term memory network performances in classification tasks using radar," in *Proc. Sensor Data Fusion: Trends, Solutions, Appl. (SDF)*, Bonn, Germany, Oct. 2017, pp. 1–6.

[22] W. Ng, G. Wang, Siddhartha, Z. Lin, and B. J. Dutta, "Range-Doppler detection in automotive radar with deep learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Glasgow, U.K., Jul. 2020, pp. 1–8.

[23] J. Rebut, A. Ouaknine, W. Malik, and P. Pérez, "Raw high-definition radar for multi-task learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 17000–17009.

[24] J.-J. Hwang, H. Kretzschmar, J. Manela, S. Rafferty, N. Armstrong-Crews, T. Chen, and D. Anguelov, "CramNet: Camera-radar fusion with ray-constrained cross-attention for robust 3D object detection," in *Proc. 17th Eur. Conf. Comput. Vis. (ECCV)*, Tel Aviv, Israel, 2022, pp. 388–405.

[25] C. X. Lu, S. Rosa, P. Zhao, B. Wang, C. Chen, J. A. Stankovic, N. Trigoni, and A. Markham, "See through smoke: Robust indoor mapping with low-cost mmWave radar," in *Proc. 18th Int. Conf. Mobile Syst., Appl., Services*, Toronto, ON, Canada, Jun. 2020, pp. 14–27.

[26] O. Schumann, M. Hahn, J. Dickmann, and C. Wöhler, "Semantic segmentation on radar point clouds," in *Proc. 21st Int. Conf. Inf. Fusion (FUSION)*, Cambridge, U.K., Jul. 2018, pp. 2179–2186.

[27] R. Pérez, F. Schubert, R. Rasshofer, and E. Biebl, "Single-frame vulnerable road users classification with a 77 GHz FMCW radar sensor and a convolutional neural network," in *Proc. 19th Int. Radar Symp. (IRS)*, Bonn, Germany, Jun. 2018, pp. 1–10.

[28] T. Liu, S. Du, C. Liang, B. Zhang, and R. Feng, "A novel multi-sensor fusion based object detection and recognition algorithm for intelligent assisted driving," *IEEE Access*, vol. 9, pp. 81564–81574, 2021.

[29] 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)*, Seoul, South Korea, Oct. 2019, pp. 924–932.

[30] X. Gao, G. Xing, S. Roy, and H. Liu, "Experiments with mmWave automotive radar test-bed," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2019, pp. 1–6.

[31] 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.*, Salt Lake City, UT, USA, Jun. 2018, pp. 4490–4499.

[32] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, Sydney, NSW, Australia, 2017, pp. 214–223.

[33] A. Danzer, T. Griebel, M. Bach, and K. Dietmayer, "2D car detection in radar data with PointNets," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Auckland, New Zealand, Oct. 2019, pp. 61–66.

[34] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast encoders for object detection from point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, Jun. 2019, pp. 12689–12697.

GUOWEI LU was born in Suzhou, Anhui, China, in 1998. He received the B.S. degree in information engineering from the Wuhan University of Technology, in 2020, where he is currently pursuing the M.S. degree in information engineering. His research interests include target tracking and recognition based on radar sensors.

ZHENHUA HE received the B.S. and M.S. degrees from the School of Materials Science, Wuhan University of Technology, in 2009, and the Ph.D. degree from the Institute of Metal Materials, Tohoku University, Japan. He is currently a Lecturer with the School of Information Engineering, Wuhan University of Technology. His research interests include electromagnetic fields and electromagnetic waves and optoelectronic films and devices.

SHENGKAI ZHANG received the Ph.D. degree from the School of Electronic Information and Communications, Huazhong University of Science and Technology, China, in 2021. He is currently an Associate Professor with the Wuhan University of Technology, China. His research interests include wireless sensing, localization, mobile computing, multi-sensor fusion, and robot control and planning.

YANQING HUANG received the master's degree in information technology and administration from Central Washington University, in 2015. She is currently the General Manager of the Operation Big Data and Information Technology Department, SAIC-GM-Wuling Automobile Company Ltd., fully responsible for the digital environment and digital operation of the company. She has published more than 30 technical articles and patents. Her main research interests include the digitization of intelligent manufacturing, the application of blockchain technology, big data technology, and other aspects. She has won ten district-level technical achievement awards and more than 20 individual honor awards.

YI ZHONG received the Ph.D. degree from the Wuhan University of Technology, China, in 2007. He was a Visiting Scholar with Stony Brook University, in 2011 and 2012. He is currently a Professor with the School of Information Engineering, Wuhan University of Technology. He has directed over 30 research projects. His research interests include data perception and big data analysis, digital signal processing, embedded system theory and design, and intelligent control theory research.

ZHUO LI received the Ph.D. degree from the School of Mechanical and Carrier Engineering, Hunan University, China, in 2020. He is currently the Manager of the Excellent Operation Digital Platform, SAIC-GM-Wuling Automobile Company Ltd. He has published more than ten academic papers, including three SCI papers as the first author and corresponding author. His main interests include the digital platform of automobile intelligent manufacturing, automobile sales prediction, and digital twin application.

YI HAN (Member, IEEE) received the B.Eng. degree from Wuhan University, China, in 2010, the M.S. degree in telecommunication from Dublin City University, in 2011, and the Ph.D. degree from the Performance Engineering Laboratory, University College Dublin, Ireland, in 2016. He is currently an Associate Professor with the School of Information Engineering, Wuhan University of Technology. He has directed over ten research projects. In recent years, he authored over 30 technical publications, proceedings, editorials, and patents. His research interests include QoE-oriented adaptive multimedia deliveries, VR video transmission, and prediction models.