Millimeter-wave radar object classification using knowledge-assisted neural network

Yanhua Wang[1,2,3,4], Chang Han[1,4], Liang Zhang[1,4]*, Jianhu Liu[5], Qingru An[5] and Fei Yang[2]

[1]Radar Research Laboratory, School of Information and Electronics, Beijing Institute of Technology, Beijing, China, [2]Beijing Institute of Technology Chongqing Innovation Center, Chongqing, China, [3]Advanced Technology Research Institute, Beijing Institute of Technology, Jinan, Shandong, China, [4]Electromagnetic Sensing Research Center of CEMEE State Key Laboratory, School of Information and Electronics, Beijing Institute of Technology, Beijing, China, [5]Beijing Rxbit Electronic Technology Co., Ltd., Beijing, China

To improve the cognition and understanding capabilities of artificial intelligence (AI) technology, it is a tendency to explore the human brain learning processing and integrate brain mechanisms or knowledge into neural networks for inspiration and assistance. This paper concentrates on the application of AI technology in advanced driving assistance system. In this field, millimeter-wave radar is essential for elaborate environment perception due to its robustness to adverse conditions. However, it is still challenging for radar object classification in the complex traffic environment. In this paper, a knowledge-assisted neural network (KANN) is proposed for radar object classification. Inspired by the human brain cognition mechanism and algorithms based on human expertise, two kinds of prior knowledge are injected into the neural network to guide its training and improve its classification accuracy. Specifically, image knowledge provides spatial information about samples. It is integrated into an attention mechanism in the early stage of the network to help reassign attention precisely. In the late stage, object knowledge is combined with the deep features extracted from the network. It contains discriminant semantic information about samples. An attention-based injection method is proposed to adaptively allocate weights to the knowledge and deep features, generating more comprehensive and discriminative features. Experimental results on measured data demonstrate that KANN is superior to current methods and the performance is improved with knowledge assistance.

KEYWORDS

millimeter-wave radar, object classification, knowledge-assisted, neural network, artificial intelligence

Introduction

Thanks to the complex structure and mechanisms of the brain, humans have the capability to continuously learn new knowledge, perceive complex environments, and make precise decisions (Cornelio et al., 2022; Kuroda et al., 2022). With the groundbreaking discovery of cells and continuous research in neuroscience, a variety of artificial neural networks have been proposed (van Dyck et al., 2022). Neural networks have promoted the development of artificial intelligence (AI) technologies in many fields, such as smart healthcare (Alsubai et al., 2022; Soenksen et al., 2022), intelligent transportation (Zhu et al., 2019; Zhu F. et al., 2020), etc. Similar to humans, networks acquire capabilities through learning. However, they learn things by brute force optimization based on input data, which limits their performance in various practical applications. To promote the next generation of AI technology, the neurology mechanism of the human brain learning process is studied, and the brain mechanism or knowledge is integrated into neural networks to help networks improve the perception and understanding of the world (Marblestone et al., 2016; Lindsay, 2020; Zhu J. et al., 2020).

This paper mainly focuses on the application of AI technology in advanced driving assistance system (ADAS), which has proved its effectiveness in safe driving and its evolution is in full swing. To elaborately capture the surroundings, multiple sensors are equipped on vehicles, such as cameras, LiDAR and millimeter-wave (MMW) radar. Cameras provide high-resolution optical images that are in line with human visual cognition and are widely applied in object detection (Redmon et al., 2016; Ren et al., 2017; Kim and Ro, 2019; Deng et al., 2022) and tracking (Danelljan et al., 2014; Smeulders et al., 2014; Nam and Han, 2016; Zhao et al., 2017; Han et al., 2019b) tasks. Although cameras offer optical images and give a semantic understanding of real-world scenarios, it is not robust facing adverse conditions, such as weak lighting or bad weather (Wang et al., 2021). As for LiDAR, it generates point cloud data and can be utilized for object detection and localization (Qi et al., 2018; Shi et al., 2019, 2021). However, these methods require dense point clouds to describe detailed information for accurate prediction, and LiDAR also has poor robustness to fog (Bijelic et al., 2018), rain or snow.

Compared with cameras and LiDAR, MMW radar is more reliable and robust in harsh environments. It is widely used in many practical scenarios, such as remote sensing target detection and classification (Liu et al., 2018; Wang et al., 2018; Liu et al., 2021; Tang et al., 2022) and intelligent transportation (Munoz-Ferreras et al., 2008; Felguera-Martin et al., 2012). Therefore, perception from pure radar data becomes a valuable alternative (Wang et al., 2021). Although it is widely used to obtain accurate location information about different objects (Prophet et al., 2019), it is still a challenge to extract discriminative semantic features from radar data

for precise object classification. Great efforts have been made to advance MMW radar object classification performance. Existing researches are mainly based on three kinds of radar data, including micro-Doppler signatures (Villeval et al., 2014; Angelov et al., 2018; Held et al., 2019), point clouds (Feng et al., 2019; Zhao et al., 2020) and range-Doppler (RD) maps or range-azimuth maps (Major et al., 2019; Palffy et al., 2020). Since RD maps can be easily obtained in engineering and maintain rich Doppler and object motion information, this paper focuses on object classification based on RD maps.

Typical feature-based approaches (Rohling et al., 2010; Heuel and Rohling, 2012) extract hand-crafted features from RD maps, such as velocity, extension in range dimension, etc., which are physically interpretable. Then, a support vector machine (SVM) classifier is trained to classify the features. To extract these features, humans constantly learn and summarize laws from various objective things and construct feature extraction algorithms based on accumulated knowledge and experience. Therefore, these methods heavily rely on human experience and algorithm design, and their performance may degrade in complex practical application scenarios.

Recent advances in deep learning have promoted the development of automatic object classification. By learning and optimizing details from pure input data, neural networks can accomplish various specific tasks. A convolutional neural network (CNN) has been established to extract valuable features for automotive radar object classification (Patel et al., 2019; Shirakata et al., 2019). Recently, a radar object detection method was proposed (Gao et al., 2019), which combines a statistical constant false alarm rate (CFAR) detection algorithm with a visual geometry group network 16 (VGG-16) classifier (Simonyan and Zisserman, 2015). After that, RadarResNet (Zhang A. et al., 2021) was constructed for dynamic object detection based on range-azimuth-Doppler maps. Ouaknine et al. (2021) utilized a fully convolutional network (FCN) to accomplish object detection and classification tasks. A RODNet (Wang et al., 2021) was proposed for radar object detection based on cross-modal supervision approach. These methods automatically learn features from training data and obtain good results. However, they discard human knowledge, which means the information they obtain may be not comprehensive.

In order to promote the intelligence of radar object classification and achieve more accurate and stable performance, it is a trend to introduce prior knowledge generated from human brains and experience into neural networks for assistance and guidance. Recently, similar ideas and methods have been studied in many fields. Reference (Chen and Zhang, 2022) presented the concept of knowledge embedding in machine learning and summarized the current research results. In the field of radar target classification, physics-aware features were obtained from synthetic aperture radar (SAR) images and injected into the layer of a deep network to provide abundant prior information for training and classification

(Huang et al., 2022). In Zhang et al. (2022), azimuth angle and phase information were extracted from SAR images and served as domain knowledge to improve the performance of SAR vehicle classification. For polarimetric high-resolution range profile classification, a feature-guided network was proposed with state-of-the-art results (Zhang L. et al., 2021). In the driving assistance system, the information obtained by the tracker has been studied to improve object classification accuracy (Heuel and Rohling, 2011). A state-aware method was proposed to model the discrimination and reliability information synchronously into the tracking framework to ensure robust performance (Han et al., 2019a).

Following the idea, a knowledge-assisted neural network (KANN) using RD map sequence for automotive MMW radar object classification is proposed. The primary intention is to inject knowledge into the neural network to supplement the network with physical information and to improve the classification performance. The network imitates the structure of neural mechanisms in human brains, however, it achieves learning tasks through brute force optimization of input data and lacks perception of the practical physical world. While knowledge is generated based on how and what the human brain thinks when accomplishing complex tasks. It conforms to human brain cognition and is an objective and physical description of the objects in the practical world. By fusing the knowledge and high dimensional data fitting, the network will have some physical cognition capability and be more similar to the way the human brain perceives, which will improve the network performance in practical driving assistance applications.

Specifically, in the method, the RD map sequence is served as input, which consists of several frames of region-of-interest (RoI) about an object based on CFAR algorithm. To improve the performance of the network, two kinds of prior knowledge of RD maps based on human expertise are extracted and hierarchically integrated into the network for assistance. The first one is image knowledge which describes the explicit spatial information of the RD map. It is obtained from the algorithms consistent with the human brain visual mechanism and applied to the attention mechanism to help the network more accurately concentrate on object regions. The second one is object knowledge which represents the semantic attribute information of objects. It includes the ranges, velocities, azimuths, and RD map extension features, which are important when humans are classifying objects. Additionally, RD maps of the same object may vary with different ranges, velocities, etc. Therefore, object knowledge is injected into the network to assist its training and classification. It is combined with the deep features extracted from the network adaptively through an attention-based injection method to provide more comprehensive and discriminative features. Experimental results on measured data of four kinds of objects
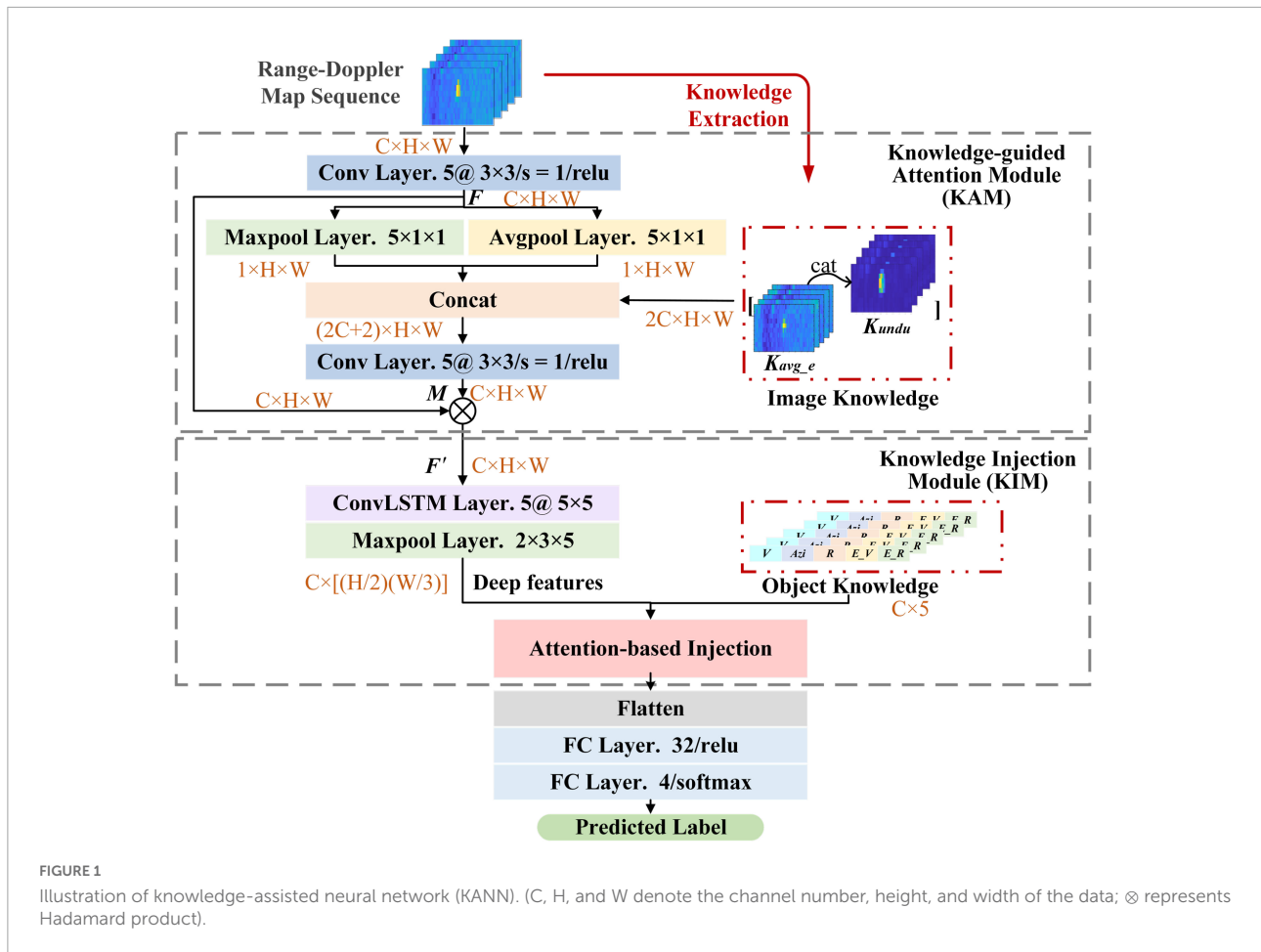
demonstrate that KANN can achieve advanced performance and the assistance of knowledge is helpful.

Knowledge-assisted neural network

The architecture of KANN is shown in **Figure 1**. KANN employs the RD map sequence containing several consecutive frames of RoIs in RD maps about an object as input data. The RoIs are cut out from RD maps based on CFAR algorithm. Different frames are fed into the network as different channels to provide temporal dimension information. Knowledge-guided attention module (KAM) and knowledge injection module (KIM) are proposed to generate the features for classification with knowledge assistance. The knowledge utilized is some prior information obtained from artificial algorithms, and it contains the physical cognition consistent with human brain when humans classify objects in traffic environments. Specifically, in KAM, an attention mechanism is established, and the prior image knowledge containing specific spatial information is applied to help make the attention assignment more reasonable and discriminative. KIM is utilized to extract spatiotemporal information about input data. Inspired by the human brain cognition when classifying objects, in this module, object knowledge containing semantic attribute information is adaptively injected into the network to provide more valuable information for classification. The rest of this section will first introduce the RD map sequence generation method in detail. Then, the specific contents of KAM and KIM are explained.

Range-Doppler map sequence generation

MMW radar dominantly transmits continuous chirps and receives the reflected echoes from objects. The workflow of RD map sequence generation is shown in **Figure 2**. First, the RD map is generated from the radar signal through the 2D-fast Fourier transform (FFT). The 2D-CFAR detection algorithm is then utilized to detect the objects. After that, inspired by Patel et al. (2019), a fixed-size RoI of the RD map is cut out for each detected object. Finally, an RD map sequence is constructed by stacking several RoIs about the same object in continuous frames of RD maps. The RoIs across frames are associated based on the range and velocity correlation, which means the detection results with the largest overlap are regarded as the same object. It should be noted that, to provide temporal information, in a sequence, the highest detected peak of the object is in the center of the first RoI, and the rest RoIs have the same location in the RD map as the first one. The ground truth categories of the RD map sequences are annotated according to the optical images. Specifically, before the data collection, the radar sensor and

FIGURE 1

Illustration of knowledge-assisted neural network (KANN). (C, H, and W denote the channel number, height, and width of the data; ⊗ represents Hadamard product).

camera are calibrated in typical scenarios. First, the range and azimuth measurement results of the radar sensor are calibrated based on angle reflectors. Then, some cooperative pedestrians, cyclists, and cars are employed as detection objects on a test road. The radar data and optical images are recorded separately, and the locations and other information of the objects from the two sensors are compared and calibrated. Finally, after collecting the measured data, the RoIs in RD maps are labeled based on the optical images.

Knowledge-guided attention module

Since an object only occupies a small region in the RD map, KAM establishes an attention mechanism that is inspired by the visual attention mechanism of human brains (Lindsay, 2020). It generates different weights to help networks focus on the discriminative regions in each RoI, while suppressing unnecessary ones. In KAM, as shown in **Figure 1**, image knowledge is prepared as the assistant knowledge to participate in the generation of the attention matrix for more precise attention assignment. Considering that the spatial information obtained by the network lacks the physical cognition of the

practical world, introducing image knowledge can make the network assign attention in a way more similar to the human brain. The image knowledge is obtained from algorithms based on human expertise and is composed of the average energy (K^{avg-e}) and undulation feature (K^{undu}) which delineate the exact spatial information and distinguish the target and background clutter. Given a pixel s_{ij} whose location is (i, j) in an RoI, we consider the pixels in the surrounding region with the size of 3×3 to calculate the features:

$$K_{ij}^{avg\_e} = \frac{1}{n}\left(\sum_{i-1}^{i+1}\sum_{j-1}^{j+1}s_{ij}^2\right), \qquad (1)$$

$$K_{ij}^{undu} = \frac{1}{n}\left(\sum_{i-1}^{i+1}\sum_{j-1}^{j+1}\left(s_{ij}-\bar{s}\right)^2\right), \qquad (2)$$

where $n = 9$ and $\bar{s}$ denotes the number of pixels and the mean amplitude of the RoI, respectively. By stacking the two feature sequences in channel dimension, image knowledge of the RD map sequence can be obtained.

Figure 3 shows the two features extracted over the same RoI in an RD map. It can be observed that K^{avg-e} and K^{undu} can represent the spatial information consistent with the visual

FIGURE 2
The workflow of range-Doppler (RD) map sequence generation.
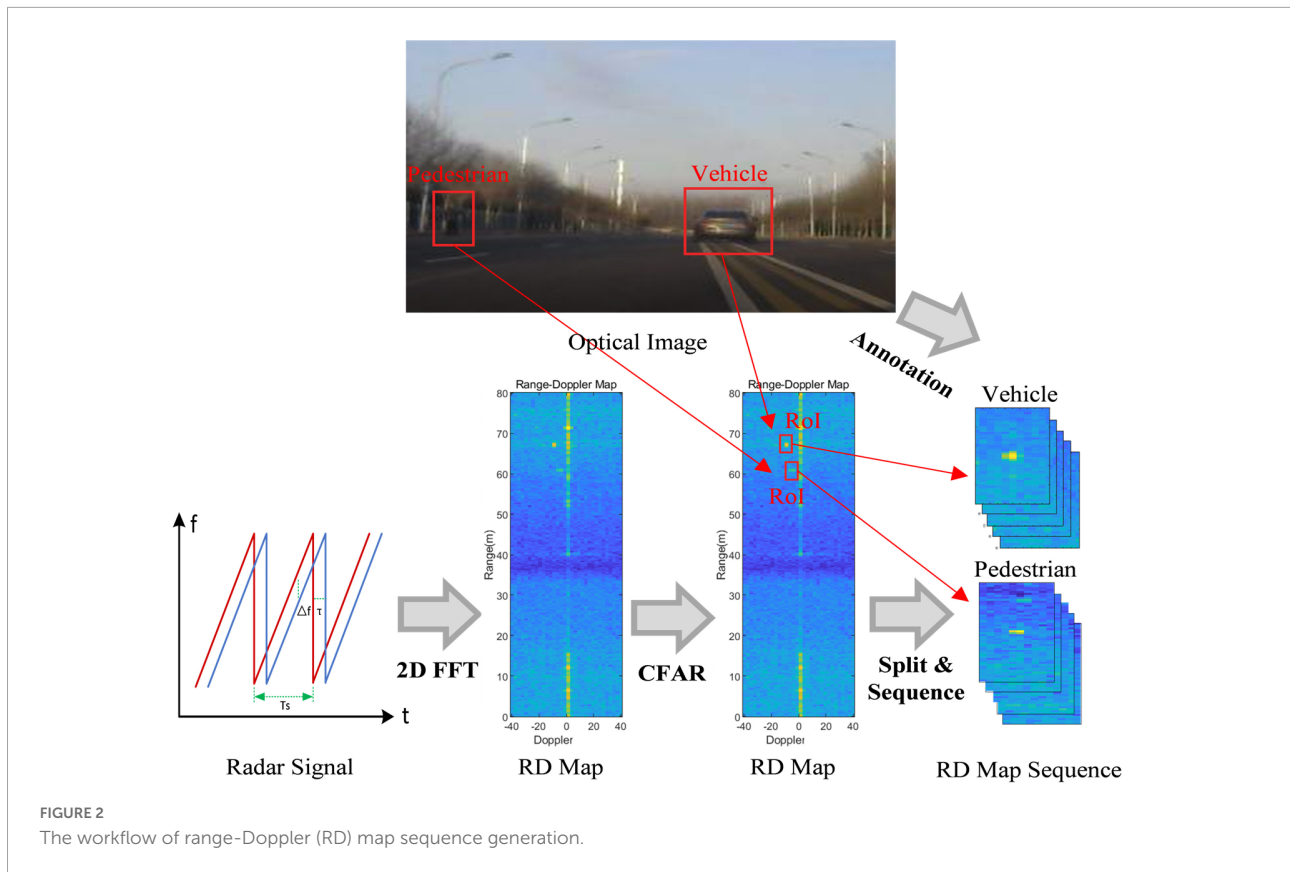
cognition of the human brain. Concretely, K^{avg_e} describes the average energy of the region and highlights the target regions, while K^{undu} describes the amplitude undulation information.

In this module, given an RD map sequence, it is first processed with a convolutional layer whose kernel size is 3×3 to obtain the feature map $\boldsymbol{F}$. Then, a max pooling layer and an average pooling layer are applied to down-sampled $\boldsymbol{F}$ in two aspects to capture spatial information autonomously. The size of the layers is configured to $5 \times 1 \times 1$ to obtain compact spatial information. At this time, image knowledge is introduced to concatenated with the pooling results in channel dimension to generate the weight matrix $\boldsymbol{M}$:

$$\boldsymbol{M} = \sigma \left(f \left(concat \left(MaxPool(\boldsymbol{F}); \; AvgPool(\boldsymbol{F}); \; IMK \right) \right) \right), \quad (3)$$

where σ denotes the "relu" activation function, f represents the convolution operation, $MaxPool\,(\cdot)$ and $AvgPool\,(\cdot)$ denote the max pooling and average pooling operation respectively. Next, according to $\boldsymbol{M}$, the redefined feature map $\boldsymbol{F}'$ can be obtained:

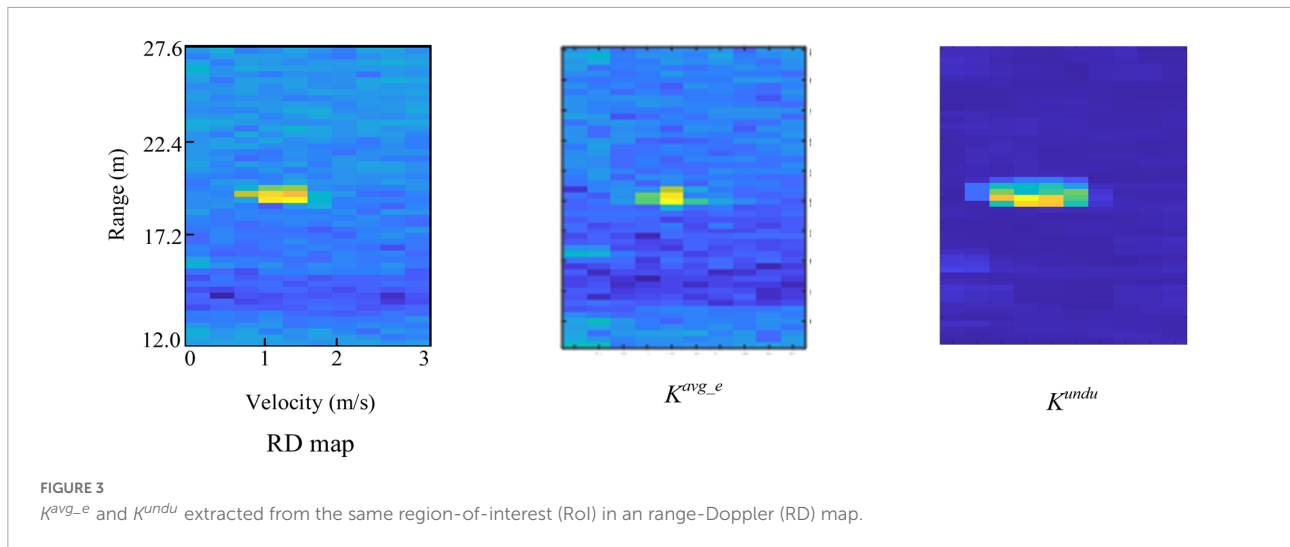$$\boldsymbol{F}' = \boldsymbol{M} \otimes \boldsymbol{F}, \quad (4)$$

where $\otimes$ denotes Hadamard product.

Compared with most existing attention mechanisms, KAM improves the physical perception ability of the network and can explore more accurate attention distribution by embedding

image knowledge which is obtained from human expertise and contains precise spatial information of samples.

Knowledge injection module

Since the RoI from a whole RD map only represents a portion of the radar field-of-view, the network trained with the data will lack the radial velocity, range, and other information of objects in the real world. However, for the same object, the shape or extension in the RD map may vary with its velocity, range, and azimuth relative to the radar sensor. Missing this information can lead to poor classification performance of the network. Therefore, to generate more discriminative features for classification, in KIM, object knowledge is injected into the network by combining with the deep features. Object knowledge includes the velocity, range, azimuth, range profile (P_r), and velocity profile (P_v) of the object in each RoI. These five kinds of information not only offer real-world information about the objects, but also have the capability of classification (Prophet et al., 2018). In this way, the network can improve the overall perception of samples, which is more similar to the cognition of the human brain and can enhance the performance of the semantic classification task.

FIGURE 3
K^{avg_e} and K^{undu} extracted from the same region-of-interest (RoI) in an range-Doppler (RD) map.

The velocity, range, and azimuth can be obtained by 3D-FFT. From the RD map obtained by 2D-FFT, the radial range and relative velocity can be captured. Then, FFT is performed on the range-velocity bins to estimate the azimuth. P_r and P_v describe the target extensions in range and velocity dimensions, as shown in **Figure 4**. P_r and P_v are the maximum length of detected points in range and velocity dimensions of the object, respectively, and can be calculated by:

$$P_r = (r_e - r_s + 1) \cdot \Delta R, \qquad (5)$$

$$P_v = (v_e - v_s + 1) \cdot \Delta v, \qquad (6)$$

where r_s and r_e denote the starting and ending points detected, ΔR represents the range resolution. In (6), v_s, v_e, and Δv denote the similar meanings in velocity dimension.

The structure of KIM is given in **Figure 1**, a ConvLSTM (Shi et al., 2015) layer is employed to extract the deep

FIGURE 4
The schematic diagram of P_r and P_v.
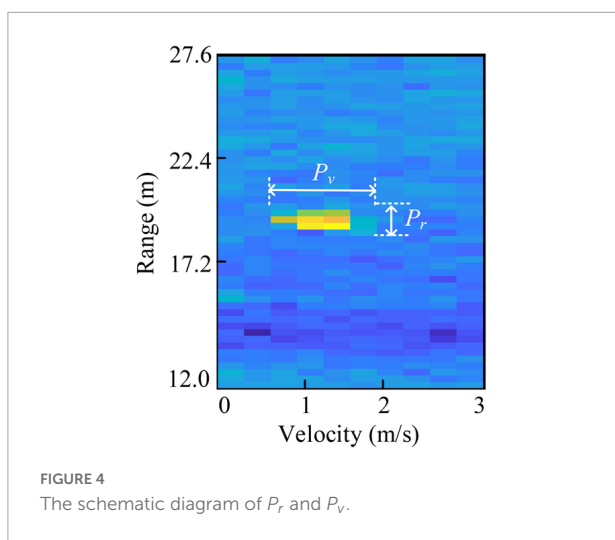
features containing spatiotemporal information. ConvLSTM network is a recurrent structure owing good capability of modeling sequential data and extracting temporal information. Meanwhile, it can learn the spatial information of each individual time step due to the convolution operation. Therefore, considering that the input is a sequence, ConvLSTM network is suitable for extracting deep features from both temporal and spatial dimensions simultaneously.

Then, object knowledge is combined with deep features. Considering that there is a gap between the two features and the same feature may have different contributions in different classification tasks, inspired by squeeze-and-excitation networks (Hu et al., 2018), we adopt an attention-based method to combine object knowledge and deep features in an adaptive way, as shown in **Figure 5**. Specifically, the deep features and object knowledge are first mapped to the same dimension through an fully connected (FC) layer and scaled to 0~1 by sigmoid activation, respectively, making them similar and conducive for combination. Then, the mapped features, F_d and F_{OBK}, are connected in channel dimension and F_c can be acquired. After that, the global pooling operation is utilized to squeeze F_c, and two FC layers are adopted to learn the attention weight vector $\boldsymbol{a}$ containing two elements:

$$\boldsymbol{a} = [a_{OBK}, a_{deep}] = \delta \left(\boldsymbol{W}_2 \cdot \sigma \left(\boldsymbol{W}_1 \cdot AvgPool \left(\boldsymbol{F}_c \right) \right) \right), \qquad (7)$$

where a_{deep} and a_{OBK} are the weights of the deep features and object knowledge, respectively, δ and σ are "sigmoid" and "relu" activation functions, $\boldsymbol{W}_2$ and $\boldsymbol{W}_1$ are parameter matrices, $AvgPool (\cdot)$ denotes the average pooling operation. Subsequently, object knowledge and the deep features are redefined by multiplying with the corresponding weights. Next, $\boldsymbol{F}_d$ and $\boldsymbol{F}_{OBK}$ are added to their redefined results to preserve original information from different sources. Finally, the
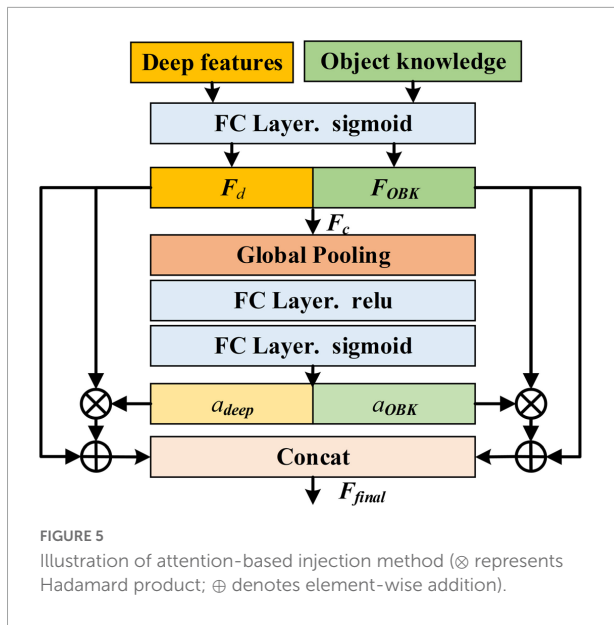
FIGURE 5
Illustration of attention-based injection method (⊗ represents Hadamard product; ⊕ denotes element-wise addition).

concatenated features are used for classification:

$$F_{final} = concat\left(F_d + a_{deep}F_d, \quad F_{OBK} + a_{OBK}F_{OBK}\right), \quad (8)$$

In this module, by injecting object knowledge, the network trains with sufficient information about samples, which improves its learning capability and classification performance.

Experiment

In this section, to evaluate the performance of KANN, we conduct a variety of experiments based on a measured dataset. The dataset is first introduced in detail. Then, the classification performance of KANN is assessed by comparative experiments. Additionally, we analyze the influence of the knowledge assistance and network structure on the performance of KANN through experiments.

Dataset preparation and implementation details

The measured dataset is collected by an automotive MMW multiple-input multiple-output (MIMO) radar with 4 Tx and 8 Rx producing a total of 32 virtual antennas. It uses the Frequency Modulated Continuous Waveform (FMCW) which is widely used in automotive radar (Hu et al., 2019). The specific configurations of radar are provided in Table 1.

The radar sensor is assembled and mounted on the front of the car as shown in Figure 6. The data is collected under different lighting conditions in different scenarios, such as city streets, elevated roads, and tunnels. Some sample scenarios are shown in Figure 7. Four kinds of objects are

TABLE 1   The specific configurations of radar.

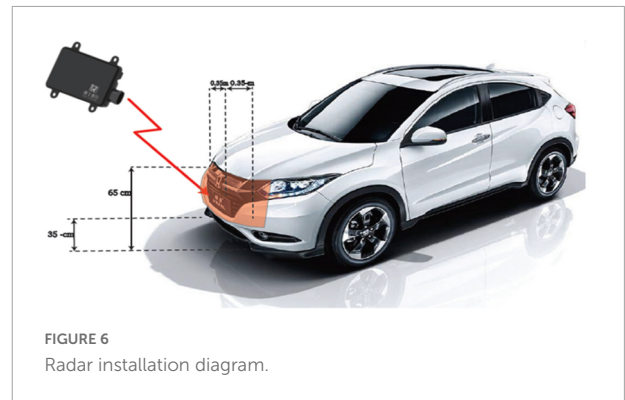| Parameter	Value
Center frequency and bandwidth	77 GHz and 600 MHz
Maximum range, resolution	80 m, 0.3 m
Maximum radial velocity, resolution	40 m/s, 0.3 m/s
Field of view	$-45° \sim 45°$
Number of chirps per frame	64
Number of samples per chirp	256

FIGURE 6
Radar installation diagram.

considered, including pedestrian, runner, vehicle, and cyclist, with overlapping speed ranges.

After collecting the original radar echoes, we perform the sequence generation method and knowledge extraction algorithms to obtain the RD map sequences and two kinds of knowledge. It should be noted that in the experiments we stack the RoIs of the same object in the RD maps of five continuous frames to construct an RD map sequence, and the RoIs in different RD map sequences are completely different. There are some samples given in Figure 8. Then, the samples are randomly divided into training and testing datasets. The detailed settings are listed in Table 2.

Moreover, the implementation details are shown. The experiments are conducted on a server cluster with a 64-bit Linux operating system. In the training phase, the batch size is set to 64, the learning rate is 0.01, and the network is optimized with adaptive moment estimation (Adam) algorithm.

Evaluation metrics

In order to evaluate the performance of different methods, the average accuracy (AA) of all classes is applied. This metric takes into account the imbalance of the data and can provide a more objective assessment of performance. It can be calculated by:

$$AA = \frac{1}{C}\sum_{c=1}^{C}\frac{N_{TP}}{N_c}, \quad (9)$$

FIGURE 7
Samples of different data collection scenarios.
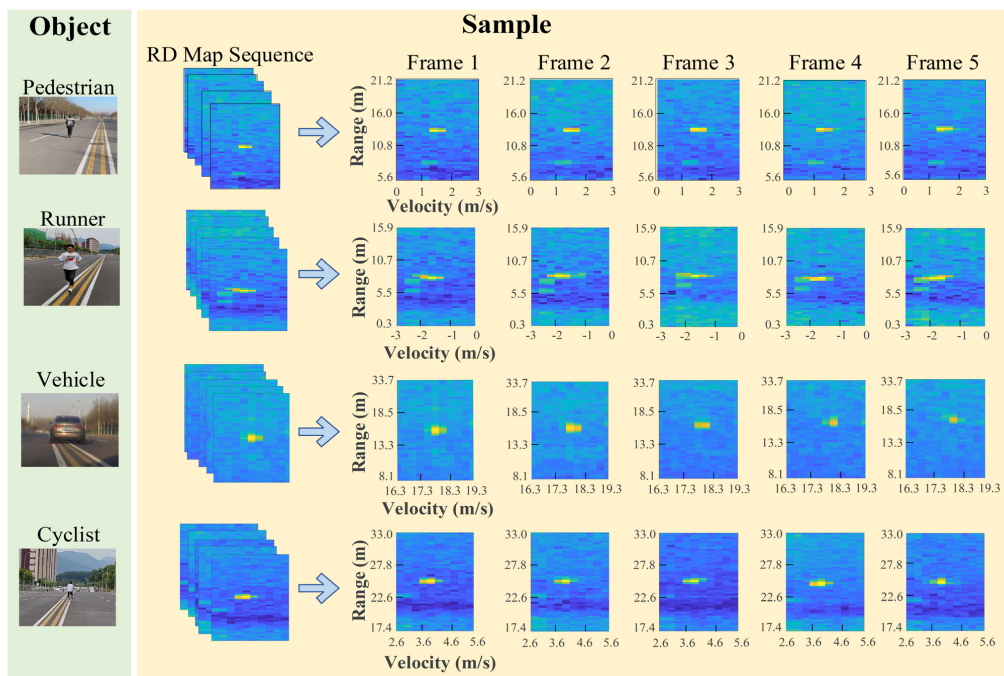
FIGURE 8
Samples of four kinds of objects.

where C is the number of classes, N_{TP} is the number of samples classified correctly in class c, and N_c is the total number of samples in class c.

Quantitative analysis

Object classification

In this part, we assess the performance of KANN based on the measured dataset. Additionally, six methods that have been studied in this field are served as comparisons. Two of them are feature-based methods that extract the features contained in object knowledge and then utilize SVM (Heuel and Rohling, 2012) and random forest (RF) classifiers to predict the object labels, respectively. The remaining four comparison methods establish different neural networks to accomplish the task, containing Three Layer-CNN (TL-CNN) (Patel et al., 2019), VGG-16 (Gao et al., 2019), FCN (Ouaknine et al., 2021), and RadarResNet (Zhang A. et al., 2021).

The per-class accuracy and AA of different methods are shown in Table 3. The values in bold in the table are the highest accuracy among the methods. We can observe that KANN

can achieve advanced performance, and its AA and per-class accuracy is all above 90%. These results demonstrate that KANN is effective in the MMW radar object classification task. Since the two kinds of knowledge are obtained from the human brain's wisdom and logic, integrating them can inspire the network to learn the samples in a way more similar to the human brain and extract more discriminative and comprehensive features.

In order to assess the performance of KANN in practical application, the runtime of different methods is analyzed, and the results are listed in Table 4. It can be observed that the feature-based methods cost the shortest time because their input is a set of prepared artificial feature vectors but not raw data. Among the deep learning methods, the runtime of KANN is the longest, which is 2.053 s. It can be inferred that the structure

TABLE 2 The detailed settings of the dataset.

Object	Number of training data	Number of testing data
Pedestrian	840	279
Runner	873	350
Vehicle	1,085	368
Cyclist	714	246
Total	3,512	1,243

TABLE 3 Experimental results on different methods.

Method	Accuracy (%)				
	Pedestrian	Runner	Vehicle	Cyclist	AA
Features+SVM (Villeval et al., 2014)	60.93	76.13	91.70	91.60	80.09
Features+RF	78.49	67.90	91.70	88.55	81.66
TL-CNN (Patel et al., 2019)	83.52	71.19	91.27	90.08	84.02
VGG-16 (Gao et al., 2019)	88.53	79.42	93.01	93.13	88.52
FCN (Ouaknine et al., 2021)	87.81	**92.59**	94.32	94.66	92.35
RadarResNet (Zhang A. et al., 2021)	**94.98**	85.60	**96.51**	90.84	91.98
KANN	93.19	91.35	**96.51**	95.42	**93.39**

TABLE 4 Computational costs on different methods.

Model	Runtime (s)
Features+SVM (Villeval et al., 2014)	0.013
Features+RF	0.007
TL-CNN (Patel et al., 2019)	0.081
VGG-16 (Gao et al., 2019)	0.864
FCN (Ouaknine et al., 2021)	0.079
RadarResNet (Zhang A. et al., 2021)	0.293
KANN	2.053

of ConvLSTM contained in KANN costs more time compared with convolution operation due to its serial units. As for the application, in general, it is an acceptable efficiency for the proposed method and its accuracy is the highest.

Ablation study

To investigate the advantage of knowledge assistance, we conduct the ablation experiment. The basic network of KANN without knowledge assistance is regarded as the baseline. Then, image knowledge and object knowledge, are separately added to the baseline for assistance. KANN is compared with these three models. Moreover, to assess the structure of KANN, we exchange the positions of KAM and KIM to conduct the experiments. Besides, considering that the spatial information can also be obtained by convolutional operation, in KANN, we remove the image knowledge and apply two convolution kernels with randomly initialized parameters to extract spatial information to compare the effect of artificial image knowledge and the automatically obtained spatial information. The results are listed in Table 5. The values in bold in the table are the highest accuracy among the methods.

It is shown that the baseline performs worst. When one kind of knowledge is added, the accuracy improves. The integration of image knowledge increases AA by 3.63%, and object knowledge makes the network achieve an 8.06% increase in AA. KANN with image knowledge and object knowledge achieves the best performance, whose AA is more than 10% higher than the baseline. Additionally, we can observe that when KIM lies in the front, AA drops by about 3%. As for the comparison of artificial image knowledge and spatial information from convolution operation, it can be seen that the AA of the model with artificial image knowledge is approximately 5% higher than the model with learnable spatial information. It can be inferred that though the network can automatically extract the information, the artificial features can supplement

TABLE 5 Experimental results of the ablation study.

Model	Accuracy (%)				
	Pedestrian	Runner	Vehicle	Cyclist	AA
Baseline	78.92	68.91	87.77	83.97	79.89
Baseline + image knowledge	79.93	73.25	90.83	90.08	83.52
Baseline + object knowledge	81.83	86.83	90.39	92.76	87.95
Baseline + image knowledge + object knowledge (KANN)	93.19	**91.35**	**96.51**	**95.42**	**93.39**
KIM + KAM	83.87	81.84	90.83	92.37	90.53
KANN with convolution operation	**98.57**	84.77	87.77	82.44	88.38

the network with physical and discriminative information which the network lacks.

From the results, we can conclude that with the knowledge assistance, the network is inspired to learn sample information no longer solely by optimizing data. It can explore information in a way more like the human brain. Although an attention mechanism is built to help the network focus on the object regions first, it is trained by the network learning mechanism. Image knowledge contains sample spatial information, which is acquired based on human brain wisdom and logic. By introducing image knowledge into the attention mechanism, the network can assign attention not only based on the network learning results, but also according to the visual cognition of the human brain. As a result, the network concentrates more precisely on the object region and achieves more accurate classification performance. As for object knowledge, it provides semantic information about objects in the real world, which is in line with the human brain perception when classifying objects. Adding this information supplements more information for the network and can improve the accuracy. Besides, by contrast, object knowledge plays a more significant role in KANN. It can be inferred that object knowledge offers the semantic information which the network lacks, while image knowledge modifies the attention weight matrix. On the other hand, from the module location experiment, it can be inferred that the network can achieve better performance by focusing on the object regions in the early stage and delicately combining object knowledge and deep features through the attention-based method just before classification.

Comparison of combination methods

In this part, to evaluate the performance of the attention-based combination method in KIM, different combination methods are applied to the fusion of object knowledge and deep features, and the results are discussed. Two common approaches, including concatenation and element-wise addition are served as the comparison methods. The results are listed in Table 6, we can observe that three combination methods can all achieve good results with AA all above 90%. The values in bold in the table are the highest accuracy among the methods. The experiment proves that the object knowledge definitely

supplements physical and discriminative information for the network and improves the classification performance. Among them, the attention-based method performs best because the network can adaptively assign weights of different features, and the features generated are more suitable for this classification task. As for the other two, they are just the simple combination of the two features, and their accuracy is lower than our adaptive attention-based combination method.

Conclusion

In this paper, we propose a knowledge-assisted network KANN based on RD map sequence for automotive MMW radar object classification. We introduce two kinds of prior knowledge to help the network learn information from samples in a way more similar to the human brain. In this way, the neural network can generate more discriminative features for semantic classification tasks. Specifically, image knowledge helps the network more accurately focus on the object regions. Object knowledge is fused with the deep feature from the network to provide more comprehensive information for classification. To effectively combine the two aspects of information, an attention-based injection method is employed to achieve the adaptive combination. Experiments based on measured data of four classes of objects verify the effectiveness of KANN and demonstrate that knowledge assistance can improve the performance of the network. Our research is continuing, and the data in more complex traffic scenarios, e.g., the crowded situation and strong interference conditions, is still being collected and processed. Since some researches show that introducing knowledge into the network can mitigate the network data size dependence (Zhang L. et al., 2021; Huang et al., 2022), in future research, based on our expanded dataset, we will conduct further experiments to assess the effect of knowledge injection with the training data size as the main topic. Simultaneously, the practical application value will be further evaluated with the data in more complex traffic scenarios.

Data availability statement

The datasets presented in this article are not readily available because the dataset is part of ongoing work. Requests to access the datasets should be directed to LZ, zhangliang@bit.edu.cn.

Author contributions

YW first proposed the idea that introducing prior knowledge into neural networks for assistance, participated in the construction of KANN, analyzed

TABLE 6 Experimental results of different combination methods.

Combination method	Accuracy (%)				
	Pedestrian	Runner	Vehicle	Cyclist	AA
Concatenation	90.32	90.53	92.14	93.13	91.53
Element-wise addition	89.96	86.83	94.76	90.84	90.59
Attention-based method	**93.19**	**91.35**	**96.51**	**95.42**	**93.39**

Funding

Conflict of interest

JL and QA were employed by Beijing Rxbit Electronic Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alsubai, S., Khan, H. U., Alqahtani, A., Sha, M., Abbas, S., and Mohammad, U. G. (2022). Ensemble deep learning for brain tumor detection. *Front. Comput. Neurosci.* 16:1005617. doi: 10.3389/fncom.2022.1005617

Angelov, A., Robertson, A., Murray-Smith, R., and Fioranelli, F. (2018). Practical classification of different moving targets using automotive radar and deep neural networks. *IET Radar Sonar Navig.* 12, 1082–1089. doi: 10.1049/iet-rsn.2018.0103

Bijelic, M., Gruber, T., and Ritter, W. (2018). "A benchmark for lidar sensors in fog: Is detection breaking down?," in *Proceedings of the 2018 IEEE intelligent vehicles symposium (IV)* (Piscataway, NJ: IEEE), 760–767. doi: 10.1109/IVS.2018.8500543

Chen, Y., and Zhang, D. (2022). Integration of knowledge and data in machine learning. *arXiv* [preprint] arXiv: 2202.10337v1

Cornelio, P., Haggard, P., Hornbaek, K., Georgiou, O., Bergström, J., Subramanian, S., et al. (2022). The sense of agency in emerging technologies for human–computer integration: A review. *Front. Neurosci.* 16:949138. doi: 10.3389/fnins.2022.949138

Danelljan, M., Khan, F. S., Felsberg, M., and Van De Weijer, J. (2014). "Adaptive color attributes for real-time visual tracking," in *Proceedings of the 2014 IEEE conference on computer vision and pattern recognition (CVPR)* (Piscataway, NJ: IEEE), 1090–1097. doi: 10.1109/CVPR.2014.143

Deng, C., Jing, D., Han, Y., Wang, S., and Wang, H. (2022). FAR-Net: Fast anchor refining for arbitrary-oriented object detection. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi: 10.1109/LGRS.2022.3144513

Felguera-Martin, D., Gonzalez-Partida, J.-T., Almorox-Gonzalez, P., and Burgos-García, M. (2012). Vehicular traffic surveillance and road lane detection using radar interferometry. *IEEE Trans. Veh. Technol.* 61, 959–970. doi: 10.1109/TVT.2012.2186323

Feng, Z., Zhang, S., Kunert, M., GmbH, R. B., and Wiesbeck, W. (2019). "Point cloud segmentation with a high-resolution automotive radar," in *Proceedings of the AmE 2019–automotive meets electronics; 10th GMM-Symposium* (Piscataway, NJ: IEEE), 1–5.

Gao, X., Xing, G., Roy, S., and Liu, H. (2019). "Experiments with mmWave automotive radar test-bed," in *Proceedings of the 53rd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA* (Piscataway, NJ: IEEE), 1–6. doi: 10.1109/IEEECONF44664.2019.9048939

Han, Y., Deng, C., Zhao, B., and Zhao, B. (2019b). Spatial-temporal context-aware tracking. *IEEE Signal Process. Lett.* 26, 500–504. doi: 10.1109/LSP.2019.2895962

Han, Y., Deng, C., Zhao, B., and Tao, D. (2019a). State-Aware anti-drift object tracking. *IEEE Trans. Image Process.* 28, 4075–4086. doi: 10.1109/TIP.2019.2905984

Held, P., Steinhauser, D., Kamann, A., Koch, A., Brandmeier, T., and Schwarz, U. T. (2019). "Normalization of micro-doppler spectra for cyclists using high-resolution projection technique," in *Proceedings of the 2019 IEEE international conference on vehicular electronics and safety (ICVES)* (Piscataway, NJ: IEEE), 1–6. doi: 10.1109/ICVES.2019.8906495

Heuel, S., and Rohling, H. (2011). "Two-Stage pedestrian classification in automotive radar systems," in *Proceedings of the 2011 12th international radar symposium (IRS)* (Piscataway: IEEE), 8.

Heuel, S., and Rohling, H. (2012). "Pedestrian classification in automotive radar systems," in *Proceedings of the 2012 13th international radar symposium (IRS)* (Piscataway: IEEE), 39–44. doi: 10.1109/IRS.2012.6233285

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *Proceedings of the 2018 IEEE/CVF conference on computer vision and pattern recognition* (Piscataway, NJ: IEEE), 7132–7141. doi: 10.1109/CVPR.2018.00745

Hu, X., Li, Y., Lu, M., Wang, Y., and Yang, X. (2019). A multi-carrier-frequency random-transmission chirp sequence for TDM MIMO automotive radar. *IEEE Trans. Veh. Technol.* 68, 3672–3685. doi: 10.1109/TVT.2019.2900357

Huang, Z., Yao, X., Liu, Y., Dumitru, C. O., Datcu, M., and Han, J. (2022). Physically explainable CNN for SAR image classification. *arXiv* [preprint] arXiv: 2110.14144v1

Kim, J. U., and Ro, Y. M. (2019). "Attentive layer separation for object classification and object localization in object detection," in *Proceedings of the 2019 IEEE international conference on image processing (ICIP)* (Piscataway, NJ: IEEE), 3995–3999. doi: 10.1109/ICIP.2019.8803439

Kuroda, N., Ikeda, K., and Teramoto, W. (2022). Visual self-motion information contributes to passable width perception during a bike riding situation. *Front. Neurosci.* 16:938446. doi: 10.3389/fnins.2022.938446

Lindsay, G. W. (2020). Attention in psychology, neuroscience, and machine learning. *Front. Comput. Neurosci.* 14:29. doi: 10.3389/fncom.2020.00029

Liu, A., Wang, F., Xu, H., and Li, L. (2018). N-SAR: A new multichannel multimode polarimetric airborne SAR. *IEEE J. Selected Top. Appl. Earth Observ. Remote Sens.* 11, 3155–3166. doi: 10.1109/JSTARS.2018.2848945

Liu, Q., Zhang, X., Liu, Y., Huo, K., Jiang, W., and Li, X. (2021). Multi-polarization fusion few-shot HRRP target recognition based on meta-learning framework. *IEEE Sens. J.* 21, 18085–18100. doi: 10.1109/JSEN.2021.3085671

Major, B., Fontijne, D., Ansari, A., Sukhavasi, R. T., Gowaikar, R., Hamilton, M., et al. (2019). "Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors," in *Proceedings of the 2019 IEEE/CVF international conference on computer vision workshop (ICCVW), Seoul, Korea (South)* (Piscataway, NJ: IEEE), 924–932. doi: 10.1109/ICCVW.2019.00121

Marblestone, A. H., Wayne, G., and Kording, K. P. (2016). Toward an integration of deep learning and neuroscience. *Front. Comput. Neurosci.* 10:94. doi: 10.3389/fncom.2016.00094

Munoz-Ferreras, J. M., Perez-Martinez, F., Calvo-Gallego, J., Asensio-Lopez, A., Dorta-Naranjo, B. P., and Blanco-del-Campo, A. (2008). Traffic surveillance system based on a high-resolution radar. *IEEE Trans. Geosci. Remote Sens.* 46, 1624–1633. doi: 10.1109/TGRS.2008.916465

Nam, H., and Han, B. (2016). "Learning multi-domain convolutional neural networks for visual tracking," in *Proceedings of the 2016 IEEE conference on computer vision and pattern recognition (CVPR)* (Piscataway, NJ: IEEE), 4293–4302. doi: 10.1109/CVPR.2016.465

Ouaknine, A., Newson, A., Rebut, J., Tupin, F., and Pérez, P. (2021). "CARRADA dataset: Camera and automotive radar with range-angle-doppler annotations," in *Proceedings of the 25th international conference on pattern recognition (ICPR)* (Milan), 5068–5075. doi: 10.1109/ICPR48806.2021.9413181

Palffy, A., Dong, J., Kooij, J. F. P., and Gavrila, D. M. (2020). CNN based road user detection using the 3D radar cube. *IEEE Robot. Autom. Lett.* 5, 1263–1270. doi: 10.1109/LRA.2020.2967272

Patel, K., Rambach, K., Visentin, T., Rusev, D., Pfeiffer, M., and Yang, B. (2019). "Deep learning-based object classification on automotive radar spectra," in *Proceedings of the 2019 IEEE radar conference, Boston, MA, USA* (Piscataway, NJ: IEEE), 1–6. doi: 10.1109/RADAR.2019.8835775

Prophet, R., Hoffmann, M., Vossiek, M., Sturm, C., Ossowska, A., Malik, W., et al. (2018). "Pedestrian classification with a 79 GHz automotive radar sensor," in *Proceedings of the 2018 19th international radar symposium (IRS)* (Piscataway, NJ: IEEE), 1–6. doi: 10.23919/IRS.2018.8448161

Prophet, R., Li, G., Sturm, C., and Vossiek, M. (2019). "Semantic segmentation on automotive radar maps," in *Proceedings of the 2019 IEEE intelligent vehicles symposium (IV)* (Piscataway, NJ: IEEE), 756–763. doi: 10.1109/IVS.2019.8813808

Qi, C. R., Liu, W., Wu, C., Su, H., and Guibas, L. J. (2018). "Frustum PointNets for 3D object detection from RGB-D data," in *Proceedings of the 2018 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (Piscataway, NJ: IEEE), 918–927. doi: 10.1109/CVPR.2018.00102

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *Proceedings of the 2016 IEEE conference on computer vision and pattern recognition (CVPR)* (Piscataway, NJ: IEEE), 779–788. doi: 10.1109/CVPR.2016.91

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Rohling, H., Heuel, S., and Ritter, H. (2010). "Pedestrian detection procedure integrated into an 24 GHz automotive radar," in *Proceedings of the 2010 IEEE radar conference* (Piscataway, NJ: IEEE), 1229–1232. doi: 10.1109/RADAR.2010.5494432

Shi, S., Wang, X., and Li, H. (2019). "PointRCNN: 3D object proposal generation and detection from point cloud," in *Proceedings of the 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (Piscataway, NJ: IEEE), 770–779. doi: 10.1109/CVPR.2019.00086

Shi, S., Wang, Z., Shi, J., Wang, X., and Li, H. (2021). From points to parts: 3D object detection from point cloud with part-aware and part-aggregation network. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 2647–2664. doi: 10.1109/TPAMI.2020.2977026

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., and Woo, W. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *arXiv* [preprint] arXiv: 1506.04214v1

Shirakata, N., Iwasa, K., Yui, T., Yomo, H., Murata, T., and Sato, J. (2019). "Object and direction classification based on range-doppler map of 79 GHz MIMO radar using a convolutional neural network," in *Proceedings of the 2019 12th global symposium on millimeter waves (GSMM)* (Piscataway, NJ: IEEE), 1–3. doi: 10.1109/GSMM.2019.8797649

Simonyan, K., and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv* [preprint] arXiv:1409.1556

Smeulders, A. W. M., Chu, D. M., Cucchiara, R., Calderara, S., Dehghan, A., and Shah, M. (2014). Visual tracking: An experimental survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 1442–1468. doi: 10.1109/TPAMI.2013.230

Soenksen, L. R., Ma, Y., Zeng, C., Boussioux, L., Villalobos Carballo, K., Na, L., et al. (2022). Integrated multimodal artificial intelligence framework for healthcare applications. *NPJ Digit. Med.* 5, 149. doi: 10.1038/s41746-022-00689-4

Tang, L., Tang, W., Qu, X., Han, Y., Wang, W., and Zhao, B. (2022). A scale-aware pyramid network for multi-scale object detection in SAR images. *Remote Sens.* 14, 973. doi: 10.3390/rs14040973

van Dyck, L. E., Denzler, S. J., and Gruber, W. R. (2022). Guiding visual attention in deep convolutional neural networks based on human eye movements. *Front. Neurosci.* 16:975639. doi: 10.3389/fnins.2022.975639

Villeval, S., Bilik, I., and Gürbuz, S. Z. (2014). "Application of a 24 GHz FMCW automotive radar for urban target classification," in *Proceedings of the 2014 IEEE radar conference* (Piscataway, NJ: IEEE), 1237–1240. doi: 10.1109/RADAR.2014.6875787

Wang, J., Zheng, T., Lei, P., and Bai, X. (2018). Ground target classification in noisy SAR images using convolutional neural networks. *IEEE J. Selected Top. Appl. Earth Observ. Remote Sens.* 11, 4180–4192. doi: 10.1109/JSTARS.2018.2871556

Wang, Y., Jiang, Z., Li, Y., Hwang, J.-N., Xing, G., and Liu, H. (2021). RODNet: A real-time radar object detection network cross-supervised by camera-radar fused object 3D localization. *IEEE J. Sel. Top. Signal Process.* 15, 954–967. doi: 10.1109/JSTSP.2021.3058895

Zhang, A., Nowruzi, F. E., and Laganiere, R. (2021). "RADDet: Range-azimuth-doppler based radar object detection for dynamic road users," in *Proceedings of the 18th conference on robots and vision (CRV)* (Piscataway, NJ: IEEE), 95–102. doi: 10.1109/CRV52889.2021.00021

Zhang, L., Han, C., Wang, Y., Li, Y., and Long, T. (2021). Polarimetric HRRP recognition based on feature-guided Transformer model. *Electron. Lett.* 57, 705–707. doi: 10.1049/ell2.12225

Zhang, L., Leng, X., Feng, S., Ma, X., Ji, K., Kuang, G., et al. (2022). Domain knowledge powered two-stream deep network for few-shot SAR vehicle recognition. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15. doi: 10.1109/TGRS.2021.3116349

Zhao, Z., Han, Y., Xu, T., Li, X., Song, H., and Luo, J. (2017). A reliable and real-time tracking method with color distribution. *Sensors* 17:2303. doi: 10.3390/s17102303

Zhao, Z., Song, Y., Cui, F., Zhu, J., Song, C., Xu, Z., et al. (2020). Point cloud features-based kernel SVM for human-vehicle classification in millimeter wave radar. *IEEE Access* 8, 26012–26021. doi: 10.1109/ACCESS.2020.2970533

Zhu, F., Lv, Y., Chen, Y., Wang, X., Xiong, G., and Wang, F.-Y. (2020). Parallel transportation systems: Toward IoT-enabled smart urban traffic control and management. *IEEE Trans. Intell. Transport. Syst.* 21, 4063–4071. doi: 10.1109/TITS.2019.2934991

Zhu, J., Su, H., and Zhang, B. (2020). Toward the third generation of artificial intelligence. *Sci. Sin. Inf.* 50:1281. doi: 10.1360/SSI-2020-0204

Zhu, L., Yu, F. R., Wang, Y., Ning, B., and Tang, T. (2019). Big data analytics in intelligent transportation systems: A survey. *IEEE Trans. Intell. Transport. Syst.* 20, 383–398. doi: 10.1109/TITS.2018.2815678