Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

RD Mamba: A Lightweight Radar Range-Doppler Spectrum Object Detection Model

Fengde Jia
fdjia@dhu.edu.cn
Donghua University

Fan Zhang
swarms23@163.com
Donghua University

Ping Zhao
pingzhao2018ph@dhu.edu.cn
Donghua University

Guohao Sun
sgh2019@scu.edu.cn
Sichuan University

Research Article

Keywords: Radar, Object Detection, SSM, Lightweight, Autonomous Driving

DOI: https://doi.org/

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

RD Mamba: A Lightweight Radar Range-Doppler Spectrum Object Detection Model

Fengde Jia[1], Fan Zhang[1], Ping Zhao[1], Guohao Sun[2*]

[1]School of Information Science and Technology, Donghua University, Street, Shanghai, 201620, Shanghai, China.

[2*]School of Information Science and Technology, Sichuan University, Street, Chengdu, 610207, Sichuan, China.

*Corresponding author(s). E-mail(s): sgh2019@scu.edu.cn;
Contributing authors: fdjia@dhu.edu.cn; swarms23@163.com;
pingzhao2018ph@dhu.edu.cn;

Abstract

This paper proposes RD Mamba, a lightweight radar Range-Doppler (R-D) spectrum object detection model, designed to enhance the radar system's target detection capability in multi-target and noisy environments. We introduce the Dynamic Spatio-Temporal (DST) Module, which integrates the Dual-Axis Spatial Fusion (DF) Block, SS2D, and Triplet Attention (TA) Block to enhance local feature extraction and global context integration. The DF Block employs a dual-branch convolutional structure to capture robust local features, while the TA Block utilizes a triplet attention mechanism to efficiently aggregate multi-dimensional features. Experiments on the RADDet and CARRADA datasets demonstrate that RD Mamba achieves Mean Average Precision(mAP)50 scores of 68.82% and 62.35%, outperforming state-of-the-art models. With only 4.2M parameters and 9.8G Floating Point Operations Per Second(FLOPs), the model strikes a balance between accuracy and efficiency, making it suitable for real-time applications in resource-constrained environments, such as autonomous driving.

Keywords: Radar, Object Detection, SSM, Lightweight, Autonomous Driving

1 Introduction

Autonomous driving technology has rapidly advanced, significantly increasing the demand for vehicle safety and stability in complex environments. Multi-sensor fusion has become the mainstream approach in perception systems [1], with millimeter-wave radar emerging as a core sensor due to its low cost, strong anti-interference capabilities, and robustness in adverse weather conditions. While cameras provide rich semantic information and LiDAR offers high spatial resolution, their performance often deteriorates under extreme conditions, such as heavy rain, low visibility, or drastic lighting changes [2],[3]. In contrast, millimeter-wave radar maintains stable performance even in severe weather conditions[4], ensuring the reliability of autonomous driving systems and making it an indispensable component.

Millimeter-wave radar typically operates by transmitting frequency-modulated continuous wave (FMCW) signals. When these signals

encounter a target object, they are reflected, generating radar echoes. The receiving antenna uses Fast Fourier Transform and peak detection to extract information about the target's distance, angle, and Doppler velocity from the echo signals, producing radar point cloud data. However, compared to LiDAR, the point cloud resolution of millimeter-wave radar is lower, leading to a loss of detail, which makes it less suitable for object detection tasks. In contrast, R-D data not only includes the distance and velocity information of targets but also effectively differentiates between static and dynamic objects while reducing interference in complex scenarios. Additionally, R-D data has relatively low computational cost, making it highly suitable for real-time applications and improving system responsiveness and overall efficiency[5]. Therefore, this study focuses on target detection based on R-D data to enhance radar object recognition capabilities in autonomous driving.

Traditional radar target detection methods are primarily based on the CFAR algorithm[6] for statistical modeling, with the most common implementations being CA-CFAR and OS-CFAR[7]. These methods adaptively calculate detection thresholds and compare the reflected signal intensity against these thresholds to determine the presence of a target.(1)CFAR algorithms assume that the clutter background is known and static, but in real-world environments, clutter distributions are often dynamic and unpredictable, making accurate modeling difficult. (2)In multi-target scenarios, they are prone to interference from strong targets, leading to missed detections of weaker targets, especially in high-noise or highly interfered environments, significantly degrading detection performance[8]. (3)They rely solely on signal intensity for target detection, failing to fully utilize other discriminative features[9], which limits their ability to effectively classify targets.

To address these limitations, deep learning-based radar target detection methods have emerged as a research focus, leveraging advanced feature extraction and signal processing techniques to significantly improve accuracy and classification performance. Unlike traditional CFAR algorithms, which rely on predefined statistical models and struggle in dynamic environments, deep learning methods adaptively learn complex nonlinear features. Recent advancements have

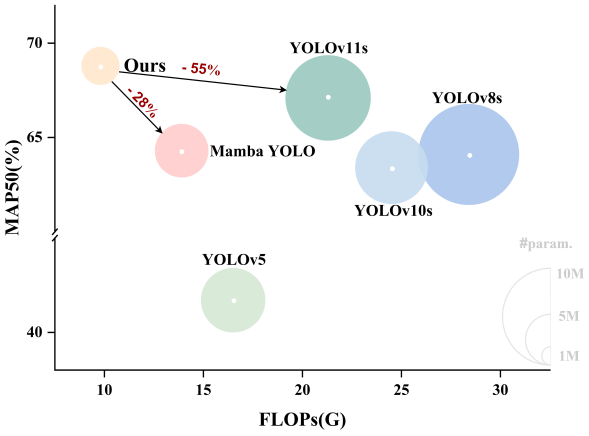


Fig. 1 This figure illustrates the relationship between MAP50, FLOPs, and model size

been fueled by millimeter-wave radar datasets such as CARRADA [10], CRUW [11], and RAD-Det [12]. For instance, [13] proposed a YOLO-like network for target detection, while [14] optimized Faster R-CNN to reduce model parameters. Other studies [15] integrated 3D CNNs and LSTM networks [16] to enhance accuracy. The RODNet model [11] demonstrated strong performance, and [17] further improved detection by introducing self-attention mechanisms and Swin Transformer modules in T-RODNet, advancing radar target detection technology.

Compared to Transformer based models, Selective State Space Sequence Models (SSMs) [18] demonstrate higher efficiency and effectiveness in long-sequence modeling, with computational complexity scaling linearly with sequence length, outperforming the quadratic complexity of Transformers. Building on SSMs, Albert Gu et al. [19] proposed the Mamba model, which achieves a global receptive field through its selective scanning mechanism, effectively capturing temporal correlations and global dependencies, making it particularly suitable for processing time-series data such as radar R-D spectra. In the field of object detection, Vmamba [20] introduced the Mamba module for the first time and further proposed the SS2D module, extending selective scanning to two-dimensional space to efficiently model local textures and global structures in images. Subsequently, researchers integrated SS2D with YOLO to propose the Mamba YOLO model [21], providing a novel and efficient solution for

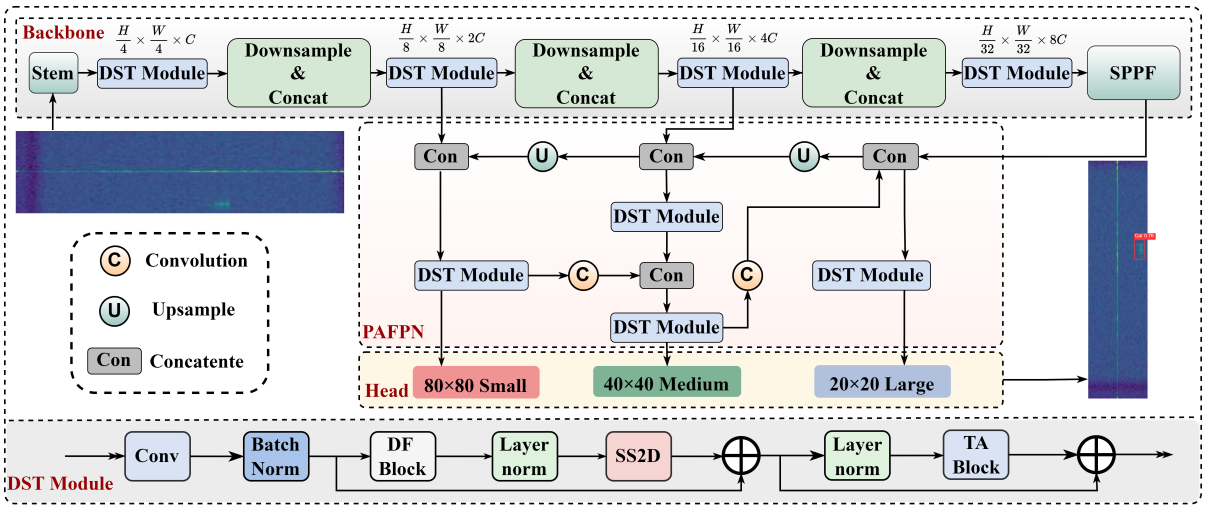


Fig. 2 The overall architectural design of RD Mamba and its core DST Module

object detection tasks. However, compared to the recently released YOLOv11 algorithm [22], the Mamba YOLO model still exhibits higher parameter counts and computational costs, failing to demonstrate a clear advantage in detection efficiency. Particularly in the field of autonomous driving, where real-time detection of radar R-D spectra is required, the model necessitates further reduction in parameter size and computational complexity to meet the stringent demands of efficient inference.

Building on this, this paper introduces a novel lightweight object detection model for radar R-D spectra, with the DST Module serving as the core of its backbone network. Drawing inspiration from the design principles of Mamba YOLO, the proposed network retains the advantages of a lightweight architecture while further optimizing for the unique characteristics of radar R-D spectra, thereby significantly improving model efficiency. Fig.1 illustrates the performance comparison between the proposed model and state-of-the-art methods.The key contributions of this paper are as follows:

- We propose a lightweight radar R-D spectrum object detection model called RD Mamba, which is based on SS2D for feature extraction. This model improves detection accuracy while reducing the number of parameters and computational costs.
- We designed the DST Module, integrating the SS2D, DF Block and TA Block. The DF Block extracts local spatial features, addressing the limitations of SS2D in local modeling, while the TA Block enhances multi-dimensional feature representation, thereby improving the robustness of the model.
- Extensive experimental results on the RADDet and CARRADA datasets demonstrate that the proposed model outperforms existing state-of-the-art methods in terms of performance.

2 Method

2.1 Preliminaries

SS2D is built upon the Selective State Space Model (SSM). The SSM system effectively captures key information in long sequences and dynamically selects what to remember and forget during sequence processing. It maps the input sequence $x(t) \in \mathbb{R}^N$ to the output sequence $y(t) \in \mathbb{R}^N$ through a hidden state $h(t) \in \mathbb{R}^N$. The continuous system is described by the following differential equations:

$$\begin{aligned} h'(t) &= Ah(t) + Bx(t) \\ y(t) &= Ch(t) \end{aligned} \tag{1}$$

where $A \in \mathbb{R}^{N\times N}$ is the state transition matrix, $B \in \mathbb{R}^{N\times 1}$ is the input weight matrix, and $C \in \mathbb{R}^{1\times N}$ is the output mapping matrix[23].

To enable computation in discrete time, SSM employs the Zero-Order Hold method to discretize the system, transforming the continuous parameters A and B into their discrete counterparts A_d and B_d. The discretization process is defined as:

$$\begin{aligned} A_d &= \exp(\Delta A) \\ B_d &= (\Delta A)^{-1}(\exp(\Delta A) - I)\Delta B \end{aligned} \tag{2}$$

where Δ is a time scale parameter that adjusts the model's temporal resolution, and I is the identity matrix. After discretization, SSM computes the hidden state at each time step using the following linear recurrence:

$$\begin{aligned} h_t &= A_d h_{t-1} + B_d x_t \\ y_t &= C h_t \end{aligned} \tag{3}$$

This discretization approach allows SSM to efficiently process discrete-time sequence data while preserving the dynamic properties of the continuous system. By leveraging the selective scanning mechanism, SSM dynamically adjusts state transition parameters, achieving global receptive fields and linear computational complexity in long-sequence modeling.

2.2 Overall Architecture

The overall architecture of RD Mamba is shown in Fig.2. The model consists of three main components: the backbone, the neck structure, and the detection head. The backbone includes the Simple Stem module, DST Module , and downsampling module, with the DST Module serving as the core component. It efficiently extracts features through a dynamic feature selection mechanism, significantly reducing model complexity. In the neck structure, we introduce a Path Aggregation Feature Pyramid Network framework and replace the original C2f module with the DST Module, enabling more efficient multi-scale feature fusion through cross-layer feature interaction. The detection head is designed with a lightweight architecture that simultaneously predicts object classes and bounding box information. This design allows the model to maintain excellent detection performance while significantly reducing the number of parameters, making it particularly suitable for radar object detection tasks that require real-time performance.

2.3 DST Module

The DST Module is the core module of RD Mamba, consisting primarily of the DF, SS2D, and TA Blocks, which handle spatial feature extraction, multi-dimensional scanning, and attention mechanisms, respectively, as shown in Fig. 2. Given the input feature F_{l-3}, it is first processed through a 1×1 convolutional layer followed by Batch Normalization (BN), and then a non-linear activation function $\hat{\Phi}$ (SiLU) is applied to generate the intermediate feature F_{l-2}. The design of Layer Normalization (LN) and Residual Linking in the DST Module ensures training stability and efficient feature flow, even under deep stacking. The computational steps are as follows: First, the local spatial transformation (DF Block, DF) and LN are applied to F_{l-2}, followed by processing through the SS2D module, which is then added to the original input F_{l-2} to obtain the intermediate feature F_{l-1}. Next, LN and attention processing (TA Block, TA) are applied to F_{l-1}, and the result is added to F_{l-1} to generate the final output feature F_l. The mathematical expressions are as follows:

$$F_{l-2} = \hat{\Phi}\left(\mathrm{BN}\left(\mathrm{Conv}_{1\times1}(F_{l-3})\right)\right) \tag{4}$$

$$F_{l-1} = \mathrm{SS2D}\left(\mathrm{LN}\left(\mathrm{DF}(F_{l-2})\right)\right) + F_{l-2} \tag{5}$$

$$F_l = \mathrm{TA}\left(\mathrm{LN}(F_{l-1})\right) + F_{l-1} \tag{6}$$

The SS2D is a key component of the DST Module, The specific structure is shown in Fig. 3 and its workflow consists of three key steps: Scan Expansion, S6 Block, and Scan Merge. During the Scan Expansion phase, the input image is decomposed into multiple sub-images, each corresponding to a specific scanning direction. From a diagonal perspective, the expansion operation proceeds along four symmetric directions (top-to-bottom, bottom-to-top, left-to-right, and right-to-left), ensuring full coverage of the input image and enhancing the multi-dimensional feature extraction capabilities. Subsequently, these sub-images undergo feature extraction through the S6 Block and are finally recombined into an output image of the same size as the original input through the Scan Merge operation. This design enables the SS2D to efficiently capture both global and

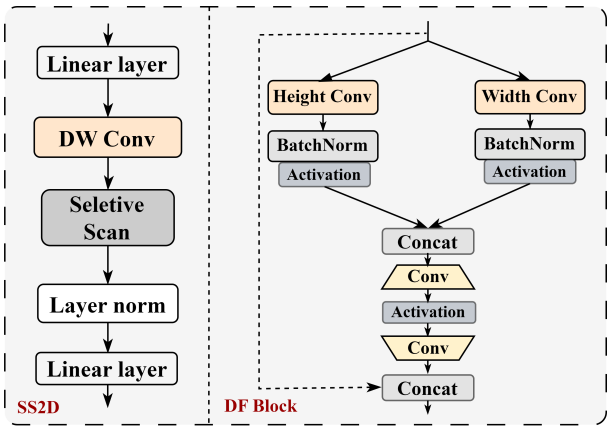


Fig. 3 The overall architecture of the SS2D and DF Block

local features while maintaining low computational complexity, further enhancing the feature extraction capability of the DST Module.

2.3.1 DF Block

Although the Mamba architecture performs exceptionally well in capturing long-range dependencies, it still faces certain limitations in extracting local features when dealing with complex scale variations. To address this issue, this paper proposes the DF Block, which enhances local spatial information extraction through a dual-branch convolutional structure and utilizes residual connections to maintain the integrity of the input features, thereby achieving more efficient feature representation.The specific structure is shown in Fig. 3

Given the input features $X_{l-1} \in \mathbb{R}^{C\times H\times W}$, DF Block employs two parallel convolutional paths to process local features in different directions. The first branch uses a 3×1 convolution kernel and group convolution (with 4 groups) to extract height-direction features, capturing vertical edges and texture information. The operation is expressed as:

$$X_{\text{height}} = \text{GELU}(\text{BN}(\text{Conv}_{3\times 1}^{\text{groups}=4}(X_{l-1}))) \quad (7)$$

The second branch uses a 1×3 convolution kernel and group convolution (with 4 groups) to extract width-direction features, capturing horizontal edges and lateral structures. The operation is expressed as:

$$X_{\text{width}} = \text{GELU}(\text{BN}(\text{Conv}_{1\times 3}^{\text{groups}=4}(X_{l-1}))) \quad (8)$$

The outputs from both branches are concatenated along the channel dimension to form the fused feature map X_{concat}, which is then passed through a 1×1 convolution to reduce the channel dimensionality while retaining multi-scale contextual information. After being processed by the GELU activation function, the non-linear representational power of the features is further enhanced. To ensure the integrity of the input information, DF Block uses a residual connection, fusing the original input X_{l-1} with the processed features. The final output features X_l are expressed as:

$$X_l = \text{Conv}_{1\times 1}(\Phi(\text{Conv}_{1\times 1}(X_{\text{concat}}))) \oplus X_{l-1} \quad (9)$$

Through this dual-branch structure and residual connection design, DF Block efficiently captures local features in the height and width directions with low computational cost, significantly enhancing the network's feature representation capability in complex tasks.

2.3.2 TA Block

The Triplet Attention[24] mechanism is an efficient attention mechanism capable of simultaneously modeling channel attention and spatial attention, thereby enhancing the feature extraction capabilities of the network.The specific structure is shown in Fig. 4 . Traditional methods such as Efficient Channel Attention (ECA)[25] and Convolutional Block Attention Module (CBAM)[26] suffer from limitations due to their inability to fully capture dependencies between channel and spatial dimensions. To address this issue, Triplet Attention introduces cross-dimensional interaction through three parallel branches, each modeling interactions between channel-height, channel-width, and height-width dimensions, thereby comprehensively integrating multi-dimensional features. The Z-Pool operation, a core component of this mechanism, reduces tensor dimensionality while preserving critical information by combining global max pooling and global average pooling, expressed mathematically

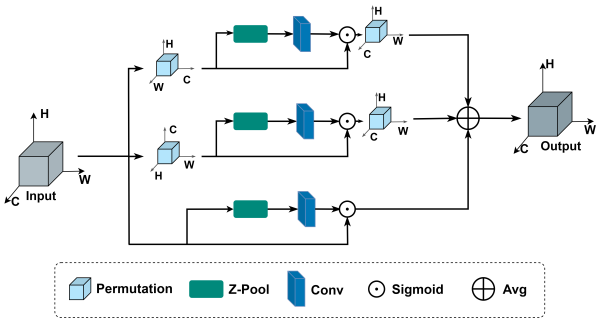


Fig. 4 Schematic Diagram of the Triplet Attention Mechanism

as:

$$Z_{\text{pool}}(x) = [\text{MaxPool}(x), \text{AvgPool}(x)] \tag{10}$$

The detailed structure is shown in Fig.4 .Given an input tensor $x \in \mathbb{R}^{C\times H\times W}$, the Triplet Attention module processes features through three branches. In the height-channel branch, the input tensor is rotated counterclockwise by 90° along the height axis to obtain $x_H \in \mathbb{R}^{W\times H\times C}$. Z-Pool is then applied to reduce its dimensionality to $\tilde{x}_H \in \mathbb{R}^{W\times H\times 2}$, followed by convolution, batch normalization, and a sigmoid activation function to compute the attention weights ω_H. Finally, the weights are multiplied with the rotated tensor x_H and rotated back to the original orientation to produce the output $x_{\text{out}}^{(1)}$. Similarly, in the width-channel branch, the tensor is rotated counterclockwise by 90° along the width axis to obtain $x_W \in \mathbb{R}^{H\times C\times W}$. After dimensionality reduction via Z-Pool, the attention weights ω_W are computed and multiplied with x_W to generate the output $x_{\text{out}}^{(2)}$. In the spatial branch, Z-Pool is directly applied to the original input x to obtain $\tilde{x}_C \in \mathbb{R}^{2\times H\times W}$. The attention weights ω_C are then computed and multiplied with x to produce $x_{\text{out}}^{(3)}$. The outputs from the three branches are averaged to generate the final enhanced feature map:

$$y = \frac{1}{3}\left(x_{\text{out}}^{(1)} + x_{\text{out}}^{(2)} + x_{\text{out}}^{(3)}\right) \tag{11}$$

By dynamically assigning weights across dimensions, this module efficiently captures dependencies between multi-dimensional features with minimal parameters, significantly improving the network's feature representation capabilities in complex tasks. Its design balances computational efficiency and model expressiveness, making it particularly suitable for scenarios requiring real-time processing of multi-dimensional data, such as radar R-D spectrum analysis.

3 Experiments

3.1 Experimental Settings

In this study, two representative radar datasets, RADDet and CARRADA, were selected for experimental validation. The RADDet dataset contains 10,158 frames of radar R-A-D tensor data, covering six target categories: cars, pedestrians, trucks, bicycles, motorcycles, and buses. The CARRADA dataset provides 12,666 frames of radar R-D, R-A, and R-A-D tensor data, including three target categories: pedestrians, trucks, and bicycles. By reducing the angular dimension of the RADDet dataset, R-D spectrum data can be obtained, while the R-D data from the CARRADA dataset can be used directly. In the visualization of the R-D spectrum, the intensity of pixel colors reflects the strength of the echo signals. Although these images appear visually abstract, their information complexity is relatively low, making them suitable for processing and analysis by SS2D.

The experimental data were randomly divided into training, validation, and test sets in an 8:1:1 ratio. During training, the Adam optimizer was used with a learning rate of 1e-4 and a batch size of 16. The training process lasted for 300 epochs, and the best model weights were selected based on the validation set. All experiments were conducted on an Nvidia RTX 4070Ti GPU using the PyTorch deep learning framework.

3.2 Experiments Result

3.2.1 Comparison with Other Models

To verify the effectiveness of the proposed method, this study compares several mainstream object detection algorithms and state-of-the-art technologies in the field of radar R-D spectrum detection, including Faster R-CNN, YOLOv5[27], YOLOv8[28], YOLOv10s[29], YOLOv11s[22], RADDet[12], and Mamba YOLO models. The experimental results show that the

Table 1 Different Radar Object Detection Models Performance Evaluated on RADDet and CARRADA Datasets

Dataset	Method	Precision(%)	Recall(%)	mAP50 (%)	Params (M)	FLOPs (G)
RADDet	Faster R-CNN	47.78	51.77	49.55	41.3	60.1
	YOLOv5s	66.21	30.21	41.71	7.1	16.5
	YOLOv8s	71.24	61.35	64.12	11.1	28.4
	YOLOv10s	72.89	56.35	63.44	8.0	24.5
	YOLOv11s	71.52	62.65	67.11	9.4	21.3
	RADDet	60.41	20.55	22.87	7.8	**5.0**
	Mamba Yolo	**76.02**	57.37	64.33	5.9	13.6
	Ours	72.12	**63.24**	**68.82**	**4.2**	9.8
CARRADA	Faster R-CNN	47.86	**67.21**	61.56	41.3	60.1
	YOLOv5s	65.68	29.76	40.16	7.1	16.5
	YOLOv8s	71.14	56.41	60.30	11.1	28.4
	YOLOv10s	68.76	54.68	57.64	8.0	24.5
	YOLOv11s	**73.19**	57.06	61.52	9.4	21.3
	RADDet	36.73	25.50	18.57	7.8	**5.0**
	Mamba Yolo	67.15	55.86	58.85	5.9	13.6
	Ours	71.04	56.27	**62.35**	**4.2**	9.8

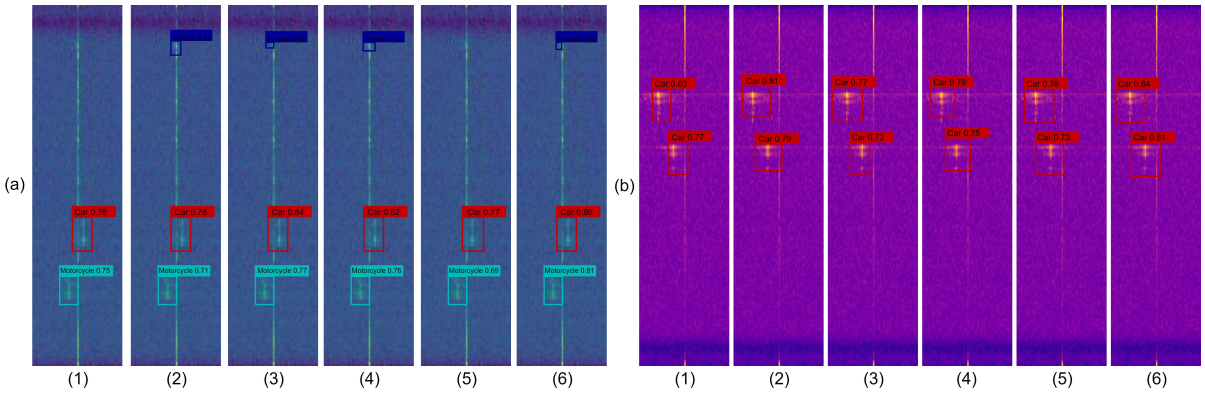


Fig. 5 Comparison of prediction results of various algorithms (a) RADDet (b) CARRADA (1) YOLOv5s (2) YOLOv8s (3) YOLOv10s (4) YOLOv11s (5) Mamba Yolo (6) Ours

proposed model demonstrates significant advantages in core metrics such as detection accuracy, mAP50, and lightweight performance.

The specific performance comparison is shown in Table 1, on the RADDet dataset, the mAP50 of the proposed model improves by 19.27%, 27.11%, 4.70%, 5.38%, and 1.71% compared to Faster R-CNN, YOLOv5s, YOLOv8s, YOLOv10s, and YOLOv11s, respectively, and by 4.49% compared to Mamba YOLO. On the CARRADA dataset, the mAP50 improves by 0.79%, 22.19%, 2.05%, 4.71%, and 0.83% compared to the aforementioned models, and by 3.50% compared to Mamba YOLO. Additionally, compared to the RADDet model, which is specifically designed for radar R-D spectrum detection, the proposed model achieves mAP improvements of 45.95% and 43.78% on the RADDet and CARRADA datasets, respectively. At the same time, our model achieves a good balance between precision and recall, demonstrating its adaptability to different data distributions and strong generalization ability. In terms of computational efficiency, the proposed model adopts a lightweight architecture design, reducing the parameter count to 4.2M and the computational cost to 9.8G FLOPs, significantly outperforming existing models. The experimental results show that, despite the significant reduction in model parameters and computational cost, the proposed method still maintains high accuracy. This result

Table 2 Results of the DST Module ablation experiments

Method		Param(M)	FLOPs(G)	RADDet mAP50 (%)	CARRADAR mAP50 (%)
Local Feature Extraction	**Global Feature Extraction**				
LocalSpatial Block	ResGated Block	5.98	13.6	64.33	58.85
	ECA	**4.16**	**9.3**	66.31	61.76
	CBAM	4.20	9.4	65.60	60.72
	TA Block	4.17	9.7	67.32	61.24
DF Block	ResGated Block	6.03	13.7	66.37	59.72
	ECA	4.20	9.4	67.61	61.37
	CBAM	4.24	9.5	64.10	59.35
	TA Block	4.21	9.8	**68.82**	**62.35**

verifies the effectiveness of the lightweight architecture design in radar object detection tasks and indicates its potential for deployment in real-time embedded systems. Fig.5 visualizes the detection results of various comparison methods on typical frames from the test set. Compared to other mainstream object detection methods, our model shows no missed detections and is able to accurately detect small targets, such as pedestrians, and distant targets with weak echo intensities.

3.2.2 Ablation experiment

The core backbone network of our model is the DST Module, which consists of the DF Block, SS2D module, and TA Block. The detailed design is shown in Fig. 2. Taking the Mamba YOLO model as a reference, which combines the advantages of selective state-space models and the YOLO algorithm, its backbone network includes the ResGated Block and LocalSpatial Block. These two modules are used to capture local features and global dependencies, similar to the functions of our DST Module and TA Block. To further validate the effectiveness of the DST Module and TA Block, we also included the commonly used ECA and CBAM modules in our experiments for comparison.

Table 2 presents the results of ablation experiments on different combinations of local feature extraction and global feature extraction modules. Our core model adopts the combination of DF Block and TA Block, achieving 68.82% and 62.35% mAP50 on the RADDet and CARRADA datasets, respectively, significantly outperforming other combinations. Specifically, when the global module is uniformly the TA Block, the mAP50 of DF Block on the RADDet dataset is 1.5% higher than that of LocalSpatial Block, indicating its stronger capability in local feature extraction. At the same time, with the same local module (DF Block), the TA Block improves mAP50 on the RADDet dataset by 1.21%, 4.72%, and 2.45% over ECA, CBAM, and ResGated Block, respectively, fully validating its effectiveness in global feature aggregation. Furthermore, the combination of DF Block and TA Block reduces the parameter count and computational cost by 30.2% and 28.5%, respectively, compared to ResGated Block, while improving accuracy by 2.45%, demonstrating the advantages of the lightweight design.

From the perspective of lightweight efficiency, the parameter count of DF Block + TA Block is close to that of ECA and CBAM, but mAP50 on the RADDet dataset improves by 1.21% to 4.72%, indicating that the spatiotemporal attention mechanism optimizes the allocation of computational resources. For example, on the CARRADA dataset, the TA Block achieves 62.35% mAP50, which is 0.98% higher than ECA, while the computational cost increases by only 0.4G. In contrast, the performance of the TA Block within the LocalSpatial Block framework is lower than that of the DF Block combination, further highlighting the importance of the collaborative design between local and global modules. Experimental results show that the DF Block enhances local detail capturing through multi-scale convolutions, while the TA Block integrates global context information using the spatiotemporal attention mechanism. The combination of the two achieves a balance between accuracy and efficiency. Despite reducing the parameter count by more than 30%, the model maintains optimal performance on both the RADDet and CARRADA radar datasets,

demonstrating its strong generalization ability. This combination provides an efficient solution for real-time radar target detection tasks with limited resources.

4 Conclusion

This paper presents a novel lightweight radar R-D spectrum object detection model, RD Mamba, aimed at balancing detection accuracy and computational efficiency. By integrating the DF Block and TA Block, the model performs excellently on the RADDet and CARRADA datasets, significantly outperforming existing methods. The lightweight design of the model makes it an ideal solution for real-time radar object detection in resource-constrained environments, particularly for embedded systems in autonomous driving. The experimental results validate that RD Mamba significantly reduces computational costs while improving detection accuracy.

Declarations

This work was supported in part by the Aeronautical Science Fund under Grant 2024M0170M2001, in part by the National Natural Science of China under Grant 62372100 and 62201371.

References

[1] Xiang, C., Feng, C., Xie, X., Shi, B., Lu, H., Lv, Y., Yang, M., Niu, Z.: Multi-sensor fusion and cooperative perception for autonomous driving: A review. IEEE Intelligent Transportation Systems Magazine (2023)

[2] Fernandes, D., Silva, A., Névoa, R., Simões, C., Gonzalez, D., Guevara, M., Novais, P., Monteiro, J., Melo-Pinto, P.: Point-cloud based 3d object detection and classification methods for self-driving applications: A survey and taxonomy. Information Fusion **68**, 161–191 (2021)

[3] Li, Y., Deng, J., Zhang, Y., Ji, J., Li, H., Zhang, Y.: A close look at the integration of lidar, millimeter-wave radar, and camera for accurate 3d object detection and tracking. IEEE Robotics and Automation Letters **7**(4), 11182–11189 (2022)

[4] Zang, S., Ding, M., Smith, D., Tyler, P., Rakotoarivelo, T., Kaafar, M.A.: The impact of adverse weather conditions on autonomous vehicles: How rain, snow, fog, and hail affect the performance of a self-driving car. IEEE vehicular technology magazine **14**(2), 103–111 (2019)

[5] Sun, Y., Fei, T., Pohl, N.: A high-resolution framework for range-doppler frequency estimation in automotive radar systems. IEEE Sensors Journal **19**(23), 11346–11358 (2019)

[6] Shnidman, D.A.: Radar detection probabilities and their calculation. IEEE Transactions on Aerospace and Electronic Systems **31**(3), 928–950 (1995)

[7] Rohling, H.: Ordered statistic cfar technique-an overview. In: 2011 12th International Radar Symposium (IRS), pp. 631–638 (2011). IEEE

[8] Liu, J., Massaro, D., Orlando, D., Farina, A.: Radar adaptive detection architectures for heterogeneous environments. IEEE Transactions on Signal Processing **68**, 4307–4319 (2020)

[9] Wang, S., Mei, L., Liu, R., Jiang, W., Yin, Z., Deng, X., He, T.: Multi-modal fusion sensing: A comprehensive review of millimeter-wave radar and its integration with other modalities. IEEE Communications Surveys & Tutorials (2024)

[10] Ouaknine, A., Newson, A., Pérez, P., Tupin, F., Rebut, J.: Multi-view radar semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 15671–15680 (2021)

[11] Wang, Y., Jiang, Z., Li, Y., Hwang, J.-N., Xing, G., Liu, H.: Rodnet: A real-time radar object detection network cross-supervised by camera-radar fused object 3d localization. IEEE Journal of Selected Topics in Signal Processing **15**(4), 954–967 (2021)

[12] Zhang, A., Nowruzi, F.E., Laganiere, R.: Raddet: Range-azimuth-doppler based radar object detection for dynamic road users. In:

2021 18th Conference on Robots and Vision (CRV), pp. 95–102 (2021). IEEE

[13] Pérez, R., Schubert, F., Rasshofer, R., Biebl, E.: Deep learning radar object detection and classification. In: Kleinheubacher Tagung 2019 (2019)

[14] Decourt, C., VanRullen, R., Salle, D., Oberlin, T.: Darod: A deep automotive radar object detector on range-doppler maps. In: 2022 IEEE Intelligent Vehicles Symposium (IV), pp. 112–118 (2022). IEEE

[15] Gao, X., Xing, G., Roy, S., Liu, H.: Ramp-cnn: A novel neural network for enhanced automotive radar object recognition. IEEE Sensors Journal **21**(4), 5119–5132 (2020)

[16] Decourt, C., VanRullen, R., Salle, D., Oberlin, T.: A recurrent cnn for online object detection on raw radar frames. IEEE Transactions on Intelligent Transportation Systems (2024)

[17] Jiang, T., Zhuang, L., An, Q., Wang, J., Xiao, K., Wang, A.: T-rodnet: Transformer for vehicular millimeter-wave radar object detection. IEEE Transactions on Instrumentation and Measurement **72**, 1–12 (2022)

[18] Gu, A., Goel, K., Ré, C.: Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396 (2021)

[19] Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)

[20] Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., Liu, Y.: Vmamba: Visual state space model. Advances in neural information processing systems **37**, 103031–103063 (2025)

[21] Wang, Z., Li, C., Xu, H., Zhu, X.: Mamba yolo: Ssms-based yolo for object detection. arXiv preprint arXiv:2406.05835 (2024)

[22] Khanam, R., Hussain, M.: Yolov11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725 (2024)

[23] Zhang, X., Xu, Q., Feng, F., Lu, X., Xu, L.: Fall-mamba: A multimodal fusion and masked mamba-based approach for fall detection. IEEE Internet of Things Journal, 1–1 (2025)

[24] Misra, D., Nalamada, T., Arasanipalai, A.U., Hou, Q.: Rotate to attend: Convolutional triplet attention module. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3139–3148 (2021)

[25] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q.: Eca-net: Efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11534–11542 (2020)

[26] Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19 (2018)

[27] Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Michael, K., Fang, J., Wong, C., Yifu, Z., Montes, D., et al.: ultralytics/yolov5: v6. 2-yolov5 classification models, apple m1, reproducibility, clearml and deci. ai integrations. Zenodo (2022)

[28] Varghese, R., Sambath, M.: Yolov8: A novel object detection algorithm with enhanced performance and robustness. In: 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), pp. 1–6 (2024). IEEE

[29] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., *et al.*: Yolov10: Real-time end-to-end object detection. Advances in Neural Information Processing Systems **37**, 107984–108011 (2024)

Figures

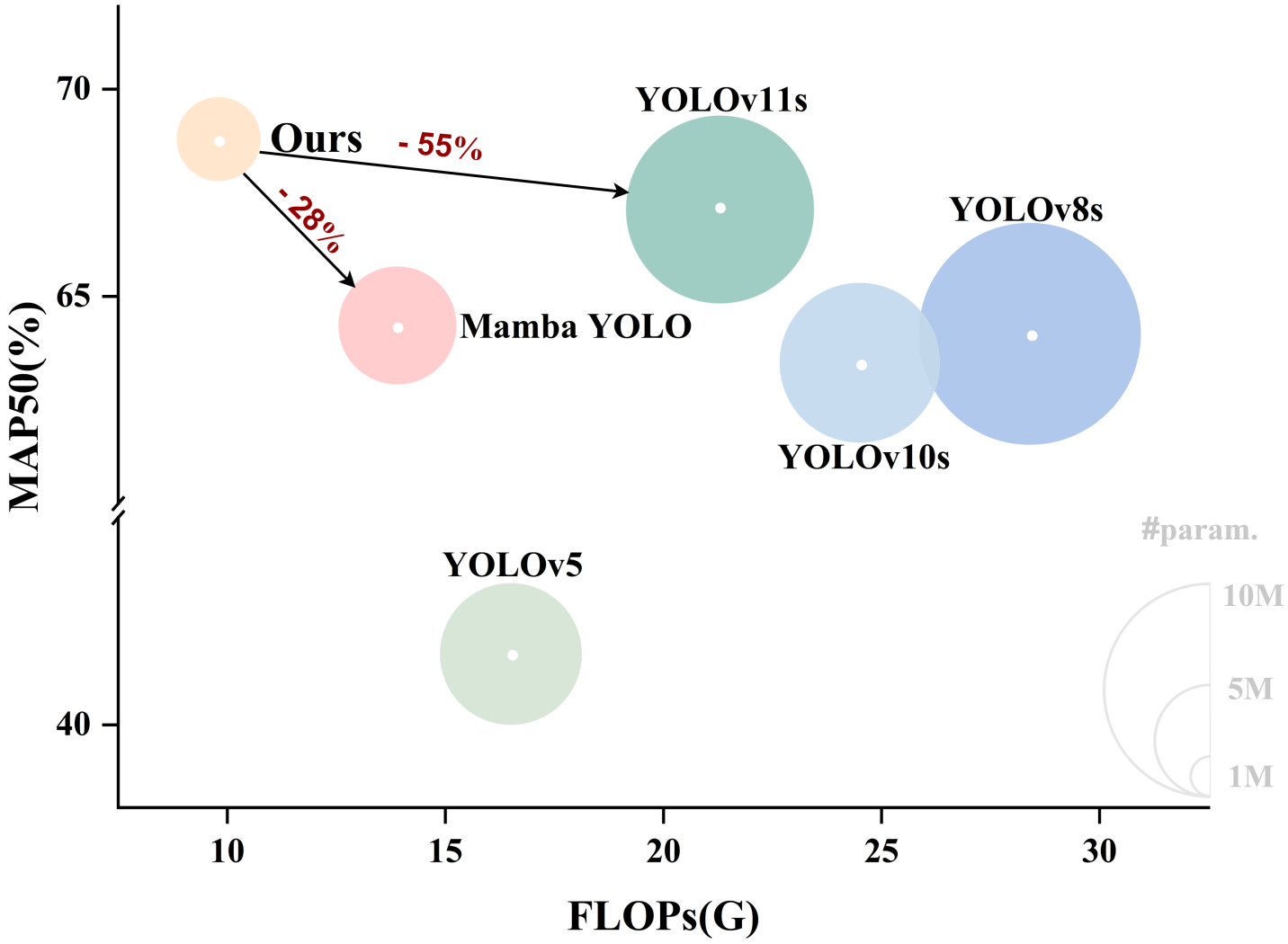


Figure 1

This figure illustrates the relationship between MAP50, FLOPs, and model size

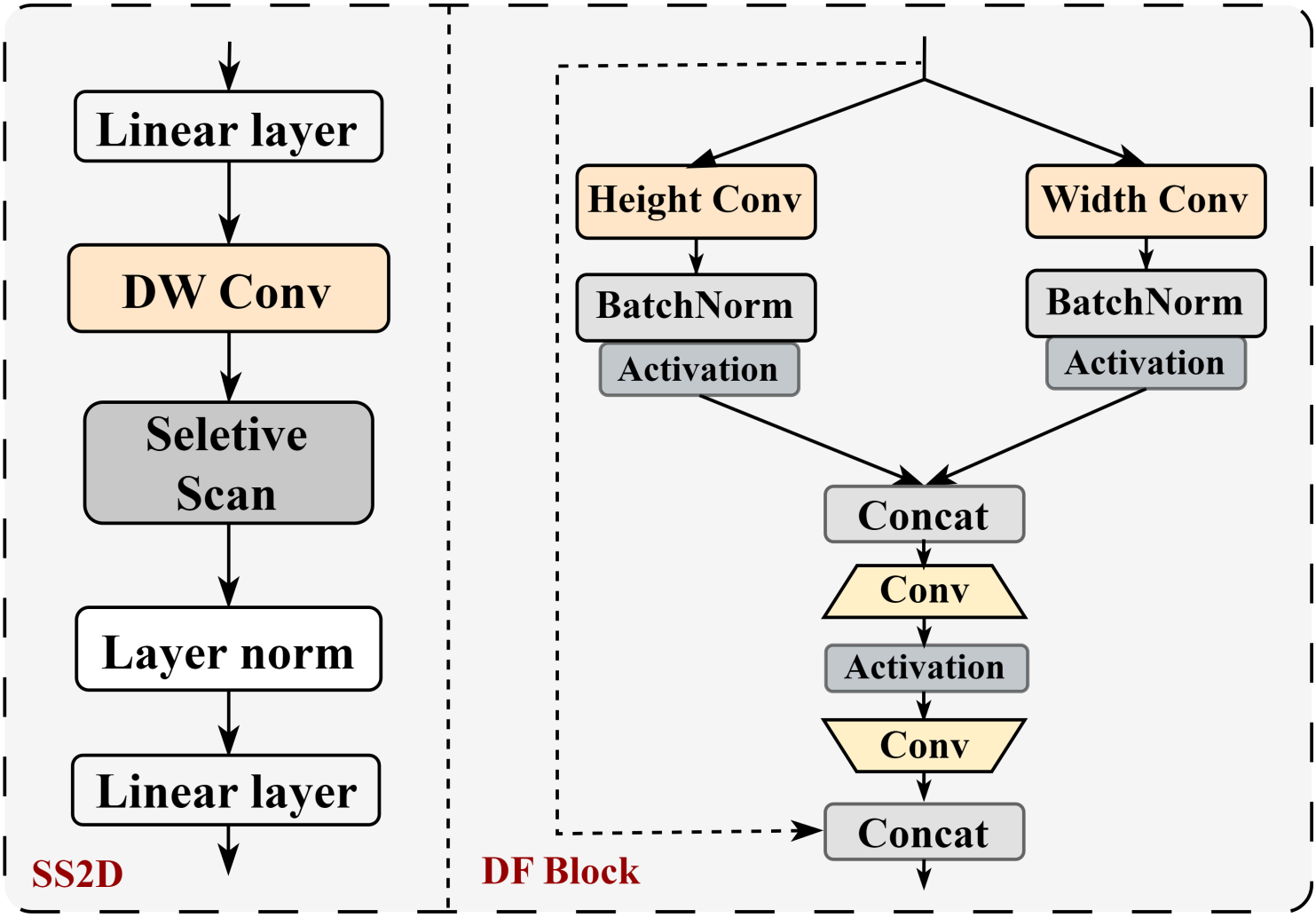


Figure 2

The overall architecture of the SS2D and DF Block

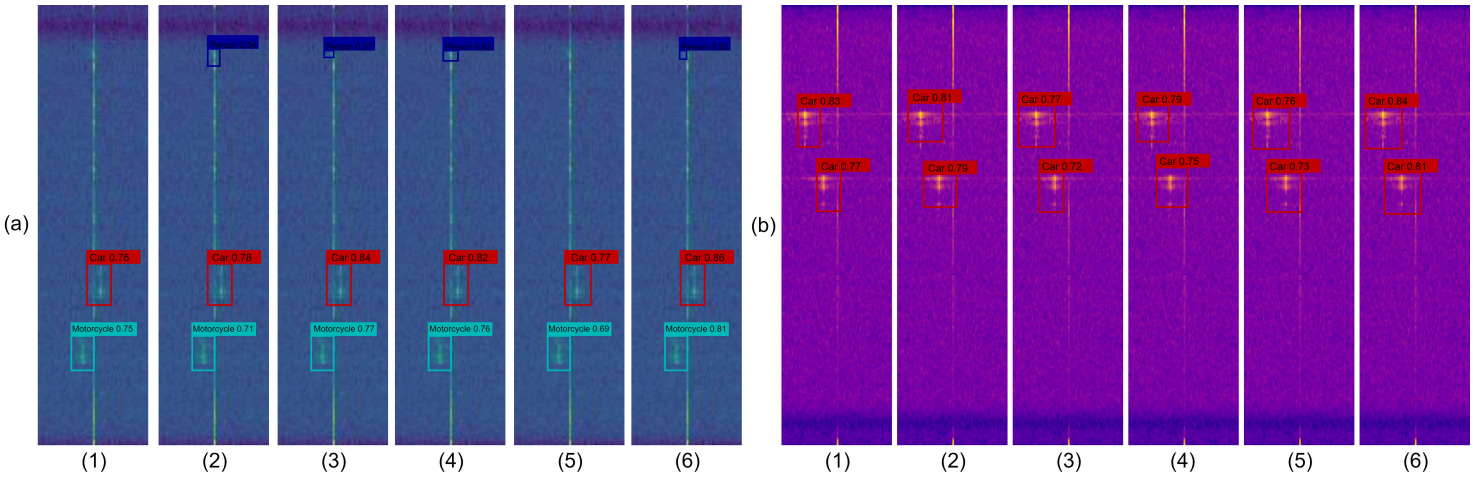


Figure 3

Comparison of prediction results of various algorithms (a) RADDet (b) CARRADA (1) YOLOv5s (2) YOLOv8s (3) YOLOv10s (4) YOLOv11s (5) Mamba Yolo (6) Ours