RESEARCH

Deep learning-based lightweight radar target detection method

Siyuan Liang[1] · Rongrong Chen[1] · Guodong Duan[2] · Jianbo Du[1]

Abstract

For target detection tasks in complicated backgrounds, a deep learning-based radar target detection method is suggested to address the problems of a high false alarm rate and the difficulties of achieving high-performance detection by conventional methods. Considering the issues of large parameter count and memory occupation of the deep learning-based target detection models, a lightweight target detection method based on improved YOLOv4-tiny is proposed. The technique applies depthwise separable convolution (DSC) and bottleneck architecture (BA) to the YOLOv4-tiny network. Moreover, it introduces the convolutional block attention module (CBAM) in the improved feature fusion network. It allows the network to be lightweight while ensuring detection accuracy. We choose a certain number of pulses from the pulse-compressed radar data for clutter suppression and Doppler processing to obtain range–Doppler (R–D) images. Experiments are run on the R–D two-dimensional echo images, and the results demonstrate that the proposed method can quickly and accurately detect dim radar targets against complicated backgrounds. Compared with other algorithms, our approach is more balanced regarding detection accuracy, model size, and detection speed.

Keywords Deep learning · Target detection · YOLOv4-tiny · Lightweight · Radar signal processing

1 Introduction

In the information society, a significant amount of visual information data are produced daily. Computer vision is a technology that teaches computers how to "see" these data. Further, it is a technology that uses imaging devices like cameras in place of human vision and analyzes the data by computers to correctly perceive the real world. Computer vision includes several subfields, such as target detection, image classification, image segmentation, etc. Among them, target detection is one of the most important research areas in the field of computer vision, and serves as a crucial building block for comprehending the high-level semantic information contained in images.

Radar has all-day, all-weather, and other operating characteristics and is extensively utilized in both military and civilian fields. Target detection is a crucial direction for radar research. Traditional radar target detection methods are usually based on constant false alarm rate (CFAR) detectors. So far, scholars have proposed many detection methods based on improved CFAR. These methods are mainly broken down into mean level CFAR and ordered statistics CFAR [1], where mean level CFAR can be further divided into cell-averaging CFAR [2], smallest of CFAR [3], and greatest of CFAR [4]. Compared to CFAR, these modified methods effectively enhance radar detection performance. However, with the increasing complexity of the cluttered environment and the number of targets, the conventional method for detecting radar targets is unsatisfactory.

In recent years, with the quick advancement of artificial intelligence technology and deep learning, it has become a trend to apply deep learning in the field of image processing. The development of deep learning has fueled the progress of image target detection technology, and a series of deep learning-based target detection algorithms have been proposed. These algorithms can be broadly divided into two categories: the two-stage target detection algorithm, represented by R-CNN, Faster R-CNN, and others [5–7], which splits the classification and localization of the target, resulting in high detection accuracy but slow detection speed;

✉ Rongrong Chen
  taronion@163.com

1 School of Communications and Information Engineering,
  Key Laboratory of Information Communication
  Network and Security, Xi'an University of Posts
  and Telecommunications, Xi'an 710121, China

2 Hunan Vanguard Group Co., Ltd., Changsha 410100, China

and the one-stage target detection algorithm, represented by SSD, YOLO, and others [8–12], which can complete the classification and localization of the target simultaneously, due to high detection speed but low detection accuracy.

The combination of deep learning and target detection tasks has been widely used in many practical applications, such as self-driving [13], face detection [14], video surveillance [15], and other systems. As a result, numerous researchers have proposed applying deep learning techniques in the field of radar target detection. To precisely locate radar targets like UAVs, Wang et al. [16] suggested a CNN-based target detection approach. Guo et al. [17] applied an enhanced YOLOX network model to the target detection task of synthetic aperture radar (SAR) images, effectively solving the challenges faced by SAR image target detection. Tm et al. [18] employed CNN for single-shot target detection to detect radar targets from R–D map data of airborne radar. To improve the detection performance of small-sized vehicles in SAR images, Zou et al. [19] proposed a CNN-based vehicle detector for the task of small target detection (e.g., vehicles) in complicated SAR image scenarios.

The majority of the time, we think about converting radar echo data into images like SAR images [20], R–D spectrograms [21], pulse distance maps [22], time–frequency images [23], etc. Most of these studies in the field of radar target detection are based on SAR images, while relatively few studies are based on R–D spectrograms. Target detection in R–D maps is challenging, since there is little information available about the features of the targets. In this study, we propose a radar target detection method for R–D maps, which successfully overcomes the above difficulties.

Although deep learning has achieved great advancements in the field of target detection, deep learning-based target detection networks are frequently complex in structure and difficult to operate directly on small devices. In response to this issue, several lightweight target detection models have been developed. These lightweight network models are often employed in various domains, because they have the characteristics of less computation, few parameters, and short running time, while can ensure the accuracy of target detection. Ye et al. [24] developed a lightweight target detection network based on YOLOv3 to meet the demands of railroad obstacle detection algorithms for real-time performance. Khoshboresh-Masouleh and Shah-Hosseini [25] employed lightweight deep learning models for the detection of target/anomaly in UAV images. To achieve real-time detection of vehicle targets in aerial photography scenes, Shen et al. [26] established a lightweight target detection approach that effectively lowers the computational cost of the model.

The tiny version of YOLO is the common lightweight target detection model. In 2020, YOLOv4-tiny was proposed as a streamlined version of YOLOv4, with only 6 million parameters equivalent to one-tenth of the original algorithm,

and the detection speed is significantly faster. To enhance the detection effectiveness of YOLOv4-tiny for small targets and further reduce its number of parameters, this paper suggests a lightweight radar target detection method based on modified YOLOv4-tiny. This way obtains the light network by compact network design, then improves the feature fusion network, and applies attention mechanisms at multiple locations to increase the model detection accuracy.

The rest of this paper is as follows. Section 2 details the improved network structure. The radar dataset used for the experiment and the data processing method are given in Sect. 3. Section 4 shows the experimental results and analysis. In Sect. 5, the conclusion is drawn.

2 Network structure design

2.1 Network structure

This study is based on the YOLOv4-tiny network for improvement, and the structure of the improved target detection network is shown in Fig. 1. The complete network consists of three main parts: the backbone feature extraction network CSPDarkNet53-Tiny, the neck feature enhancement extraction network FPN, and the head detection network YOLO Head. First, this paper replaces a portion of the ordinary convolution in the network with DSC, significantly reducing the model size with less accuracy loss. Next, we add BA to the residual blocks to increase the model detection accuracy while reducing the number of parameters. Then, to improve the extraction of small target feature information, a 52×52 feature layer is added to the FPN. Additionally, lightweight attention modules called CBAM are embedded in the network to make it entirely focused on crucial information and enhance the model detection effect.

2.2 DSC

Compact network design is a typical method for lightweight neural networks. This paper uses DSC to achieve a compact and efficient network structure to lower the number of parameters required for convolutional computation. As shown in Fig. 2, the convolution calculation of DSC is divided into two parts: depthwise convolution (DC) and pointwise convolution (PC) [27]. First, DC is performed for all channels, one convolution kernel corresponds to one channel, and the channel number of the feature map remains constant throughout this process. However, DC performs a separate convolution operation for each channel, making it unreasonable to use the feature information from different channels at the same location. As a result, the output from several channels is combined using PC to
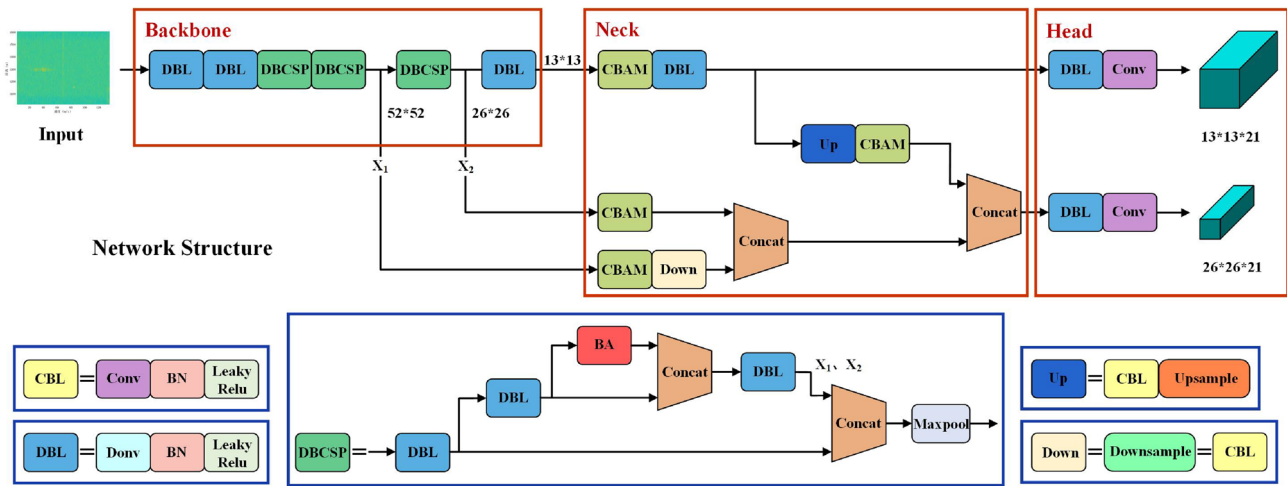
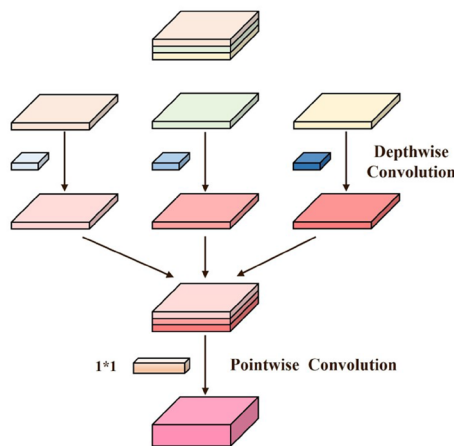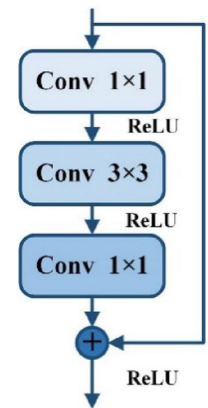**Fig. 1** The improved network structure of YOLOv4-tiny

Fig. 2 DSC structure diagram

Fig. 3 BA schematic diagram

generate a new feature map, and PC, namely, an ordinary convolution of 1×1.

Compared to regular convolution, DSC sacrifices a minor accuracy loss in exchange for fewer convolutional operation parameters, allowing the network to perform more computational tasks simultaneously and improving the detection speed. CBL is an ordinary convolutional block, including convolution, batch normalization, and activation function, and DSC replaces the convolution in CBL to obtain DBL. This algorithm dramatically reduces the number of parameters in the network by replacing all CBLs in the backbone network with DBL, certain CBLs in the neck network with DBL, and the 3×3 convolutions in the head structure with DSC.

2.3 BA

BA is a residual structure with a decreasing and subsequently increasing number of channels, so named because it resembles the neck of a bottle [28]. This structure achieves jump connections by summation to connect the various layers of the network, thus facilitating the backpropagation of the gradient and speeding up the convergence of the network. Figure 3 depicts the BA schematic. The number of channels in the feature map is first compressed using a 1×1 convolution layer. Then, the feature is extracted using a 3×3 convolution network. Finally, the feature is expanded using a 1×1 convolution layer, so that the number of output channels equals the original. These two 1×1 convolution layers are used to downscale and upscale the input feature dimension, respectively, to decrease the number of network parameters and deepen the network layers. Increasing the network depth can generally boost detection accuracy, but doing so increases the number of network parameters. However, BA can increase

the detection accuracy of the network while saving the number of parameters.

Assume that the input feature map has D channels and that BA's default channel expansion rate is 0.5. Then, the number of parameters for the convolutional network directly using a 3×3 and the number of filters D is as follows:

$$D \times 3 \times 3 \times D = 9D^2. \tag{1}$$

The number of parameters that use BA is as follows:

$$D \times 1 \times 1 \times \frac{1}{2}D + \frac{1}{2}D \times 3 \times 3 \times \frac{1}{2}D + \frac{1}{2}D \times 1 \times 1 \times D = \frac{13}{4}D^2. \tag{2}$$

The ratio of parameter quantities between the two is approximately 3. It is clear that, in terms of the number of parameters, the use of BA has a substantial advantage over using a 3×3 convolution directly. In this paper, due to the addition of BA and the substitution of DBL for CBL in the residual structure, the generated DBCSP block has a much smaller number of parameters than the original residual structure.

2.4 Improved FPN

In the YOLO series target detection algorithm, feature maps of various scales detect targets of different sizes. Large-scale feature maps are used to detect small targets, and small-scale feature maps are used to detect large targets. Unlike YOLOv4-tiny, which only uses two feature layers of various scales for the multiscale feature fusion portion, this study selects additional scale feature layers for information fusion to improve the network detection accuracy for small targets. That is, one branch is drawn forth at the 52×52 feature layer of the backbone network, then which the high-level features and the low-level features extracted from the three branches are combined with raising the network detection precision.

In neural networks, the pooling operation is frequently used to reduce the size of feature maps. However, the pooling operation usually causes some information loss. To perform downsampling and preserve as much of the feature map information as feasible, this approach uses convolution in the multiscale feature fusion network rather than a pooling operation.

2.5 CBAM

Target detection can be made more accurate and effective using the attention mechanism to make the network adaptively focus on the relevant information related to the target while ignoring the irrelevant information and allocating the limited computing resources to this relevant information. Attention mechanisms can usually be classified as channel attention mechanisms, spatial attention mechanisms, and hybrid attention mechanisms.

The attention mechanism can be flexibly applied to all parts of the network. This paper uses the attention mechanism for the upsampling results and the three effective feature layers extracted from the backbone network. The CBAM, a hybrid attention mechanism, is employed in this study. The channel attention module (CAM) and the spatial attention module (SAM), which make up the CBAM, are joined by a tandem connection [29]. CAM enables the network to pay more attention to the target class information in the feature map, while SAM is used to pay more attention to the target location information. The CBAM structure diagram is shown in Fig. 4. The input feature map first passes through CAM to obtain the channel attention weight coefficients, which are multiplied by the input feature map to produce the output feature map. The output feature map then goes through SAM to achieve the spatial attention weight coefficients, which are then multiplied with the input feature map to generate the final feature map.

2.6 Loss function

During the training phase of the model, the input data are propagated forward to obtain the predicted value. The difference between the predicted value and the true value is the loss value, and the loss function is essentially a function of calculating the loss value. We wish to minimize the loss value during target detection, so it is crucial to select the appropriate loss function.

The loss of YOLO consists of three parts: regression loss, confidence loss, and classification loss. In this paper, the complete intersection over union (CIoU) loss [30] is used for regression loss, while the cross-entropy loss is employed for

Fig. 4 CBAM structure diagram
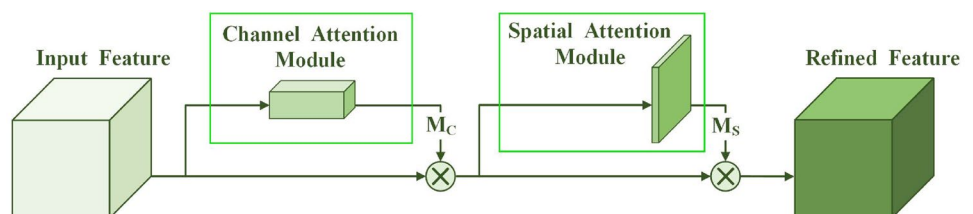
confidence loss and classification loss. CIoU adds the calculation of center point distance and aspect ratio on the basis of the calculation of overlapping area, so that the prediction box will be closer to the ground truth box. Cross-entropy loss is used to measure the degree of difference between the predicted probability and the true probability of the same random variable. The lower the cross-entropy value, the closer the two probability distributions are, and the higher the prediction accuracy of the model.

3 Dataset and data processing method

3.1 Dataset

To address the gap between standard public datasets in radar target detection and identification, Song et al. release a standard dataset for radar target detection and tracking with fixed-wing UAVs in clutter [31]. The dataset includes radar echo data, radar wave gate data, and labeled truth data. The labeled truth data provide the labeling results at a data rate of 50 ms, i.e., one labeled data output every 1600 pulses. Table 1 displays the labeled data results at the 50th ms moment in data2.

3.2 Data processing method

The radar echo data in the dataset are a pulse sequence formed after pulse compression in the fast time dimension. By selecting a certain number of pulses for FFT in the slow time dimension, the energy distribution of the target in the R–D domain can be obtained, that is, the R–D spectrogram. Figure 5 shows the R–D plot after the first 1600 pulses of data2 are subjected to FFT, where X, Y, and Z denote the velocity, distance, and signal amplitude, respectively.

Data2 is measured under the medium signal-to-noise ratio condition, and it is evident from Table 1 and Fig. 5 that object2 has become completely obscured by the clutter, making it extremely challenging to identify the target later on.

To reduce the impact of clutter, we consider performing clutter suppression in the radar echo sequence prior to FFT. The dataset is primarily collected in the ground background, and the clutter in the data is mainly static. Common static clutter filtering methods include the moving target indicator (MTI) and the mean phase elimination algorithm. Similar to the effect of high-pass filters, MTI employs filters to suppress the clutter components in radar echo data. This work applies MTI to suppress clutter and raise the signal-to-noise ratio of the radar.

Table 1 Labeled truth data

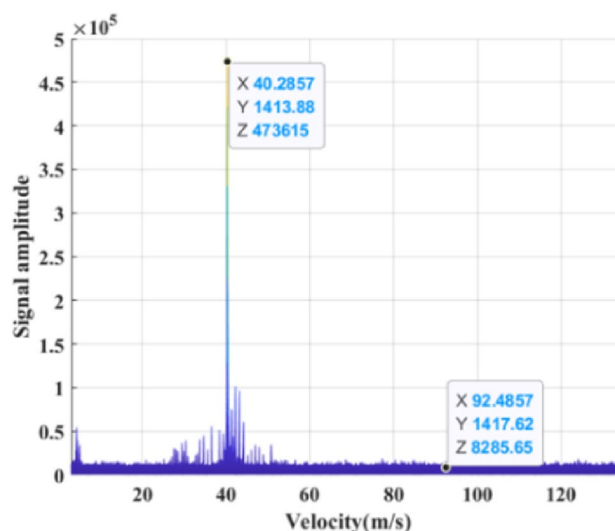| Target	Speed (m/s)	Distance (m)
object1	40.31	1411.63
object2	92.36	1414.76

Fig. 5 R–D plot after FFT

Figure 6 shows the R–D diagram after MTI processing, and Figs. 5 and 6 are created using the same data segment to make it easier to see the effects of MTI processing.

Figure 6 shows that, after clutter suppression, object2 can be distinguished, making it possible to label the target on the R–D diagram. By comparing Figs. 5 and 6, it can be observed that the clutter suppression method is effective and has essential significance in improving the radar target detection performance. In addition, a small portion of the data in the dataset is collected in the ground-sea background, and the experiment proves that the clutter suppression method is also applicable.
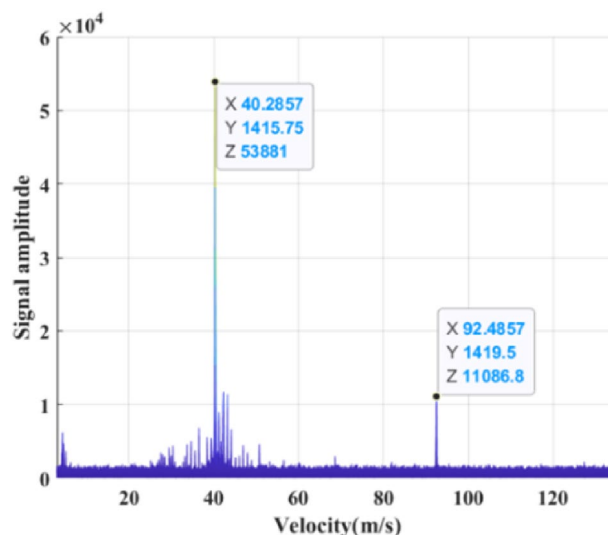

Fig. 6 R–D plot after MTI + FFT

The labeling results are presented every 50 ms for the actual value data, so 1600 pulses are selected each time for MTI and FFT processing to produce R–D plots for comparison with the actual value data when the target is labeled.

In this paper, we conduct experiments on a dataset composed of R–D graphs, including two types of targets with a total of 1200 images. A sample R–D image dataset is displayed in Fig. 7. The dataset is calibrated with the LabelImg image calibration tool, using object1 and object2 to represent the two different types of targets. Labeling results are recorded in PascalVOC format for input into the neural network during training.

4 Experimental results and analysis

The experiment uses the mean average precision (mAP), model size, and detection speed FPS at the intersection over union (IoU) equal to 0.5 as indicators to evaluate the model performance.

The operating system used for the experiments is Windows 11, the GPU is NVIDIA GeForce RTX 3050, the CPU is AMD Ryzen 7 5800H with Radeon Graphics, and the deep learning framework used for the algorithms is PyTorch.

4.1 Optimizer comparison experiment

By adjusting the model parameters, optimizers can find the model's optimal solution, which minimizes the loss function. The outcomes of using several optimizers in the same model can vary greatly, so it is crucial to select the appropriate optimizer for your model. Optimizers can be broadly classified into three categories: gradient descent, momentum optimization, and adaptive learning rate optimization algorithms.
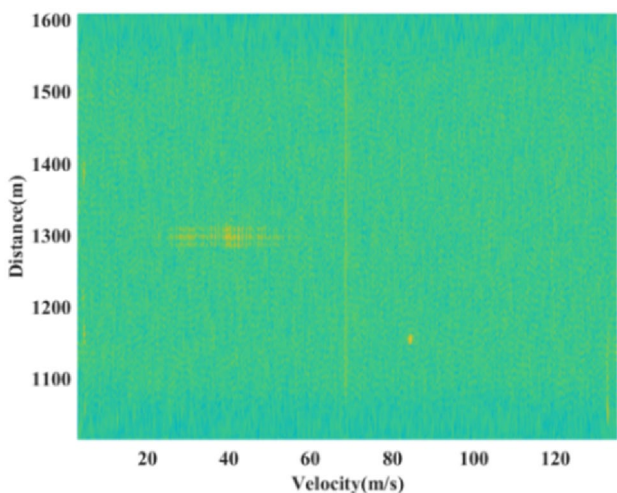

Fig. 7 Dataset sample

Among them, stochastic gradient descent (SGD) and adaptive moment estimation (Adam) are the more commonly utilized optimizers, and Table 2 compares their effects on the YOLOv4-tiny model.

When using the Adam algorithm, the model is trained for 200 epochs. Since SGD converges slower than Adam, the training epochs are set to 300. According to Table 2, we can obtain that the Adam algorithm outperforms the SGD algorithm in terms of mAP and detection speed with fewer training epochs. In light of fact that the Adam algorithm is more suitable for our model, all optimization methods used in the YOLOv4-tiny model mentioned in the following experiments are Adam.

4.2 Lightweight experiments

The YOLOv4-tiny network is significantly reduced when the improved algorithm adds DSC. Table 3 compares the network before and after the DSC application.

According to Table 3, by incorporating DSC into the YOLOv4-tiny network, the mAP is decreased; however, the model size is reduced by 84.75%, making it easier to deploy the model on devices with constrained computational capacity. Furthermore, because the computational volume of the model is lessened, the model detection speed is increased.

4.3 Ablation experiments

To observe the effects of various improvements on the network detection effect, we conduct ablation experiments based on YOLOv4-tiny + DSC. The results of the ablation experiments with different enhancements are shown in Table 4.

Among these, Dv4-tiny is a YOLOv4-tiny network with DSC added, way 1 represents the addition of BA in the residual structure, way 2 denotes the introduction of three branches from the backbone network for multiscale feature fusion, and way 3 means the addition of the CBAM.

Table 2 Performance comparison of different optimizers

Optimizer	mAP (%)	Epoch	FPS
SGD	89.58	300	105.34
Adam	92.91	200	114.36

Table 3 Network comparison results before and after adding DSC

Network	mAP (%)	Model size (M)	FPS
YOLOv4-tiny	92.91	23.60	114.36
YOLOv4-tiny + DSC	90.26	3.60	124.21

Table 4 Results of ablation experiments

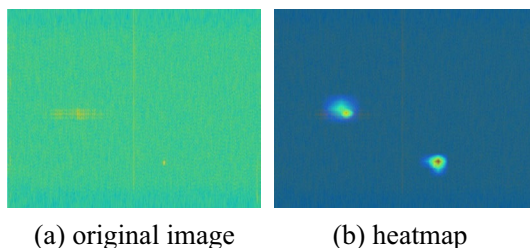| Dv4-tiny	Way 1	Way 2	Way 3	mAP (%)	Model size (M)	FPS
√				90.26	3.60	124.21
√	√			91.62	3.46	113.14
√	√	√		92.33	4.68	106.96
√	√	√	√	95.69	5.04	92.63

(a) original image   (b) heatmap

Fig. 8 Visualization result

As shown in Table 4, adding BA reduces the model size and improves mAP by 1.36% but also lowers FPS by 8.91% because BA deepens the network layers, thereby enhancing network detection accuracy while decreasing network detection speed. The upgraded FPN raises the mAP by 0.71% by merging feature information at diverse scales. Nevertheless, the network model grows and detection speed drops due to introducing one more branch. In this experiment, the CBAM is placed at four locations to strengthen the feature extraction ability of the network. As a result, the algorithm boosts mAP by 3.36%, which is a significant effect. However, the involvement of the attention mechanism also increases the model size and makes the network detection speed slower.

The optimization model is obtained through ablation experiments, and this model is the final network model proposed in this study. To observe the performance of the optimized network, we got a heatmap of the network. The heatmap is a way of network visualization to check which area of the image contributes more to the final output of the model, and the visualization results of our model are displayed in Fig. 8. As shown in Fig. 8, the region on which the network focuses is where the two targets are located, so the network is more effective in detecting them.

4.4 Attention mechanism comparison experiment

The CBAM is chosen through comparison experiments with other attention mechanisms. At the same location, the SE attention mechanism [32], the CA attention mechanism [33], and the CBAM are led, respectively. The outcomes of the comparison experiment for several attention mechanisms are displayed in Table 5, where Dv4 stands for network Dv4-tiny + way 1 + way 2.

Table 5 Results of attentional mechanism comparison

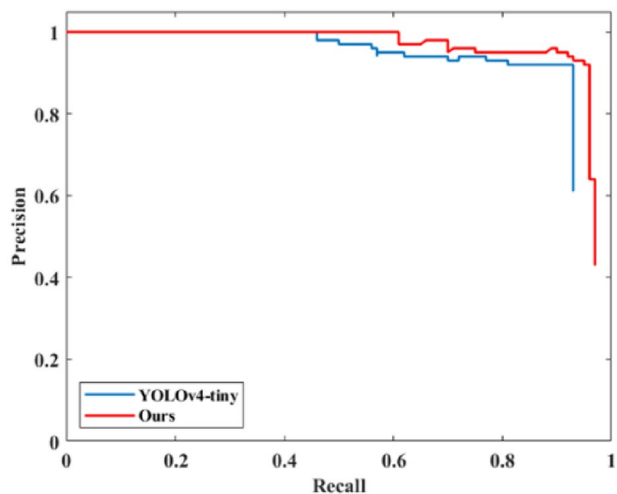| Network	mAP (%)	Model size (M)
Dv4	92.33	4.68
Dv4 + SE	94.52	4.86
Dv4 + CA	94.65	4.96
Dv4 + CBAM	95.69	5.04

Fig. 9 P–R curve of object1

Table 5 indicates that adding various attention mechanisms to Dv4 increases the mAP of the model to varying degrees. In comparison, CBAM offers the highest advancement in network detection accuracy.

4.5 Algorithm comparison experiment

To verify the detection performance of the improved algorithm, we compare it with the original network YOLOv4-tiny. The *P–R* curves for the two algorithms directed against object1 and object2 are depicted in Figs. 9 and 10.

Figures 9 and 10 suggest that, compared to the original algorithm, the AP values of both types of targets in the modified algorithm have enhanced, with the AP value of object1 improving more significantly. In addition, the recall rate of the proposed method is higher, indicating that the extracted feature information is more affluent than the
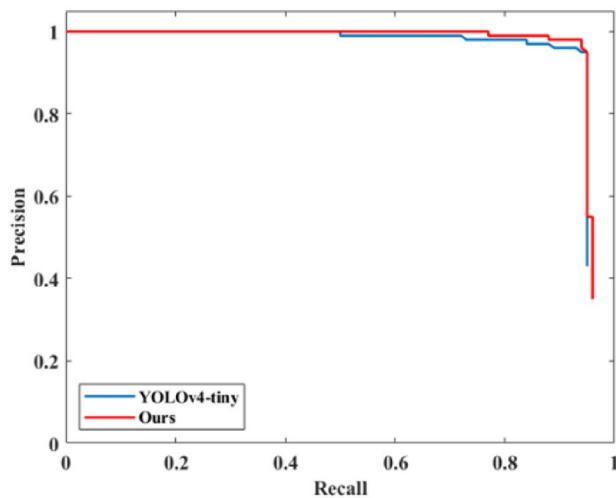
Fig. 10 P–R curve of object2

original network and that it is easier to recognize weak and small targets. However, in the case of a high recall rate, the accuracy of both algorithms declines for object1 and object2, because these two targets are smaller and hence contain fewer feature details.

IOU = 0.5 was set in the training phase, and the mAP values were recorded once every five training epochs. Figure 11 exhibits the graphs of the mAP values for the original approach and the modified method with training epochs.

In Fig. 11, we can see that the mAP values of the two algorithms stabilized at about 80 epochs, that the modified algorithm mAP values grow more quickly at the beginning of training, and that the enhanced algorithm mAP values are higher than the initial algorithm in all training epochs. The improved algorithm thus performs better in mAP.
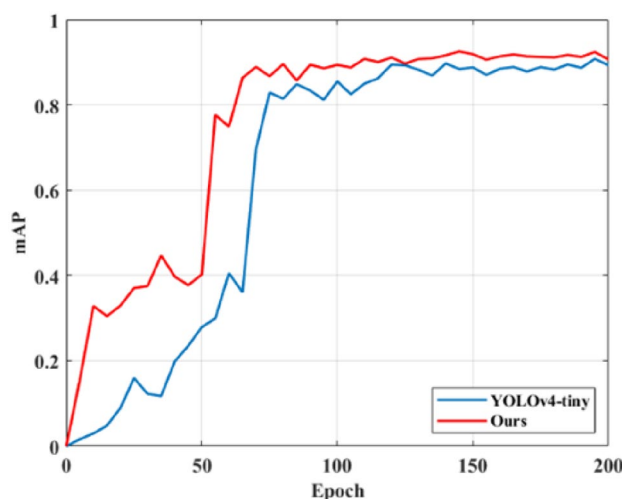

Fig. 11 mAP variable curve with training epochs

Table 6 Algorithm comparison results

Network	mAP (%)	Model size (M)	FPS
Faster R-CNN	97.71	546.84	6.28
SSD	91.95	94.98	48.21
YOLOv4-tiny	92.91	23.60	114.36
Ours	95.69	5.04	92.63

To further demonstrate the effectiveness of the proposed method, it is compared with the two-stage target detection algorithm Faster R-CNN, the single-stage target detection algorithm SSD, and YOLOv4-tiny in terms of mAP, model size, and FPS. Table 6 shows the comparison experiment results of several algorithms.

In Table 6, the backbone feature extraction networks employed in Faster R-CNN and SSD are ResNet and VGG, respectively. The Faster R-CNN method possesses the highest detection accuracy, but it has a larger model and a slower detection speed, as seen in Table 6. Compared to the SSD algorithm and the YOLOv4-tiny algorithm, the proposed approach performs better in the aspects of mAP and model size, but its detection speed is not as fast as the YOLOv4-tiny algorithm.

From the above experimental results, the suggested method is more well balanced and can ensure network detection accuracy and speed while being lightweight. The model size of this technique is 78.64% smaller than the previous YOLOv4-tiny algorithm, while the mAP is upped by 2.78%.

5 Conclusion

In this paper, we apply the currently popular deep learning models to radar data and propose a light radar target detection network based on improved YOLOv4-tiny, considering the demand for lightweight models from low computing power devices. The network uses DSC and BA to lower the number of model parameters efficiently; meanwhile, the model detection accuracy is increased by improving FPN and introducing the CBAM. The suggested algorithm is trained and validated on the R–D image. The experimental results show that the approach in this paper achieves superior results in terms of detection accuracy, model size, and detection speed. The lightweight model presented in this paper can also be extended to other application scenarios in future research.

Author contributions Siyuan Liang investigated the research background of the work, supervised the tests, and reviewed the manuscript. Rongrong Chen carried out the method design and software design and

Data availability The dataset can be found at https://www.scidb.cn/en/detail?dataSetId=720626420979597312.

Declarations

Conflict of interest All authors declare that they have no conflicts of interest affecting the work reported in this article.

References

1. Rohling, H.: Radar CFAR thresholding in clutter and multiple targets situation. IEEE Trans. Aerosp. Electron. Syst. **AES-19**(4), 608–621 (1983)

2. Finn, H.M., Johnson, R.S.: Adaptive detection mode with threshold control as a function of spatially sampled clutter-level estimates. RCA Rev. **29**(2), 414–464 (1968)

3. Trunk, G.V.: Range resolution of targets using automatic detectors. IEEE Trans. Aerosp. Electron. Syst. **AES-14**(5), 750–755 (1978)

4. Hansen, V.G., Sawyers, J.H.: Detectability loss due to "greatest of" selection in a cell-averaging CFAR. IEEE Trans. Aerosp. Electron. Syst. **AES-16**(1), 115–118 (1980)

5. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587. IEEE, Columbus, OH, USA (2014)

6. Girshick, R.: Fast R-CNN. In: 2015 IEEE International Conference on Computer Vision (ICCV), IEEE, Santiago, Chile, pp. 1440–1448 (2015)

7. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: toward real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017)

8. Liu, W., Anguelov, D., Erhan, D.: SSD: single shot multibox detector. European Conference on Computer Vision, pp. 21–37. Springer, Cham (2016)

9. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788. IEEE, Las Vegas, NV, USA (2016)

10. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6517–6525. IEEE, Honolulu, HI, USA (2017)

11. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement (2018). arXiv Preprint. arXiv:1804.02767

12. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: YOLOv4: optimal speed and accuracy of object detection (2020). arXiv Preprint. arXiv:2004.10934

13. Zhang, Y., Wang, J., Wang, X., Dolan, J.M.: Road-segmentation-based curb detection method for self-driving via a 3D-LiDAR sensor. IEEE Trans. Intell. Transp. Syst. **19**(12), 3981–3991 (2018)

14. Li, Z., Tang, X., Wu, X., Liu, J., He, R.: Progressively refined face detection through semantics-enriched representation learning. IEEE Trans. Inf. Forensics Secur. **15**, 1394–1406 (2019)

15. Huang, C., Wu, Z., Wen, J., Xu, Y., Jiang, Q., Wang, Y.: Abnormal event detection using deep contrastive learning for intelligent video surveillance system. IEEE Trans. Ind. Inform. **18**(8), 5171–5179 (2021)

16. Wang, C., Tian, J., Cao, J., Wang, X.: Deep learning-based UAV detection in pulse-Doppler radar. IEEE Trans. Geosci. Remote Sens. **60**, 1–12 (2022)

17. Guo, Q., Liu, J., Kaliuzhnyi, M.: YOLOX-SAR: high-precision object detection system based on visible and infrared sensors for SAR remote sensing. IEEE Sens. J. **22**(17), 17243–17253 (2022)

18. Tm, D., Verma, R., Rajesh, R., Varughese, S.: Single shot radar target detection and localization using deep neural network. In: 2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), IEEE, pp. 1–9 (2022)

19. Zou, B., Qin, J., Zhang, L.: Vehicle detection based on semantic-context enhancement for high-resolution SAR images in complex background. IEEE Geosci. Remote Sens. Lett. 19 (2021)

20. Gao, S., Liu, J.M., Miao, Y.H., He, Z.J.: A high-effective implementation of ship detector for SAR images. IEEE Geosci. Remote Sens. Lett. 19 (2021)

21. Rizvi, S.M.D., Ahmad, S., Khan, K., Hasan, A., Masood, A.: Deep learning approach for fixed and rotary-wing target detection and classification in radars. IEEE Aerosp. Electron. Syst. Mag. **37**(3), 32–42 (2022)

22. Pan, M., Chen, J., Wang, S., Dong, : Z.: A novel approach for marine small target detection based on deep learning. In: 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP), pp. 395–399. IEEE, Wuxi, China (2019)

23. Su, N., Chen, X., Guan, J., Li, Y.: Deep CNN-based radar detection for real maritime target under different sea states and polarizations. International Conference on Cognitive Systems and Signal Processing, pp. 321–331. Springer, Singapore (2019)

24. Ye, T., Zhao, Z., Wang, S., Zhou, F., Gao, X.: A stable lightweight and adaptive feature enhanced convolution neural network for efficient railway transit object detection. IEEE Trans. Intell. Transp. Syst. **23**(10), 17952–17965 (2022)

25. Khoshboresh-Masouleh, M., Shah-Hosseini, R.: 2D target/anomaly detection in time series drone images using deep few-shot learning in small training dataset. In: Houssein, E.H., Abd Elaziz, M., Oliva, D., Abualigah, L. (eds.) Integrating Meta-Heuristics and Machine Learning for Real-World Optimization Problems. Studies in Computational Intelligence, vol. 1038, pp. 257–271. Springer, Cham (2022)

26. Shen, J., Zhou, W., Liu, N., Sun, H., Li, D., Zhang, Y.: An anchor-free lightweight deep convolutional network for vehicle detection in aerial images. IEEE Trans. Intell. Transp. Syst. **23**(12), 24330–24342 (2022)

27. Howard, A.G., Zhu, M., Chen, B.: MobileNets: efficient convolutional neural networks for mobile vision applications (2017). arXiv Preprint. arXiv:1704.04861

28. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. IEEE, Las Vegas, NV, USA (2016)

29. Woo, S., Park, J., Lee, J.Y., Kweon, : I.S.: CBAM: convolutional block attention module. European Conference on Computer Vision, pp. 3–19. Springer, Cham (2018)

30. Zheng, Z., Wang, P., Ren, D.: Enhancing geometric factors in model learning and inference for object detection and instance segmentation. IEEE Trans. Cybern. **52**(8), 8574–8586 (2021)

31. Song, Z., Hui, B., Fan, H.: A dataset for detection and tracking of dim aircraft targets through radar echo sequences. China Sci. Data **5**(3), 272–285 (2020)

32. Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E.: Squeeze-and-excitation networks. IEEE Trans. Pattern Anal. Mach. Intell. **42**(8), 2011–2023 (2020)

33. Hou, Q., Zhou, D., Feng, J.: Coordinate attention for efficient mobile network design. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 13708–13717. IEEE, Nashville, TN, USA (2021)