Transactions on Computational and Scientific Methods | Vo. 4, No. 8, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Improved Faster R-CNN Algorithm for Robust Detection of Tank Targets in Complex Environments

Zetao Chen
School of Electrical Engineering and Computer Science, Auburn University, USA
ztao.c8790@auburn.edu

Abstract: Target detection plays a vital role in various fields, including military applications, where the detection of tank targets is crucial for accurate battlefield assessments. However, the complexity of modern combat environments, including target occlusion and camouflage, presents significant challenges. Traditional target detection algorithms, once dominant, have been surpassed by deep learning methods, particularly convolutional neural networks (CNNs). This paper explores the application of the Faster R-CNN algorithm for tank target detection, focusing on enhancing detection accuracy and robustness in complex settings. To address the issue of target occlusion and limited data, the Faster R-CNN model is improved by integrating a ResNet50 backbone with a Feature Pyramid Network (FPN) for more refined feature extraction.

Keywords: Target Detection; Complex Environment; Convolutional Neural Network; Area of Candidate.

1. Introduction

Target extraction technology has been widely used in all aspects, in addition to the medical, monitoring and other fields of application, in the military field also plays a great role. Detection of military weapons is also critical, but the complexity of the modern battlefield environment is increasing, as is the need for accurate damage to enemy targets. Targets are easy to be affected by natural environment or perceived factors, so it is still very difficult to realize automatic detection of tank targets under these complex background, which is an urgent problem to be solved.

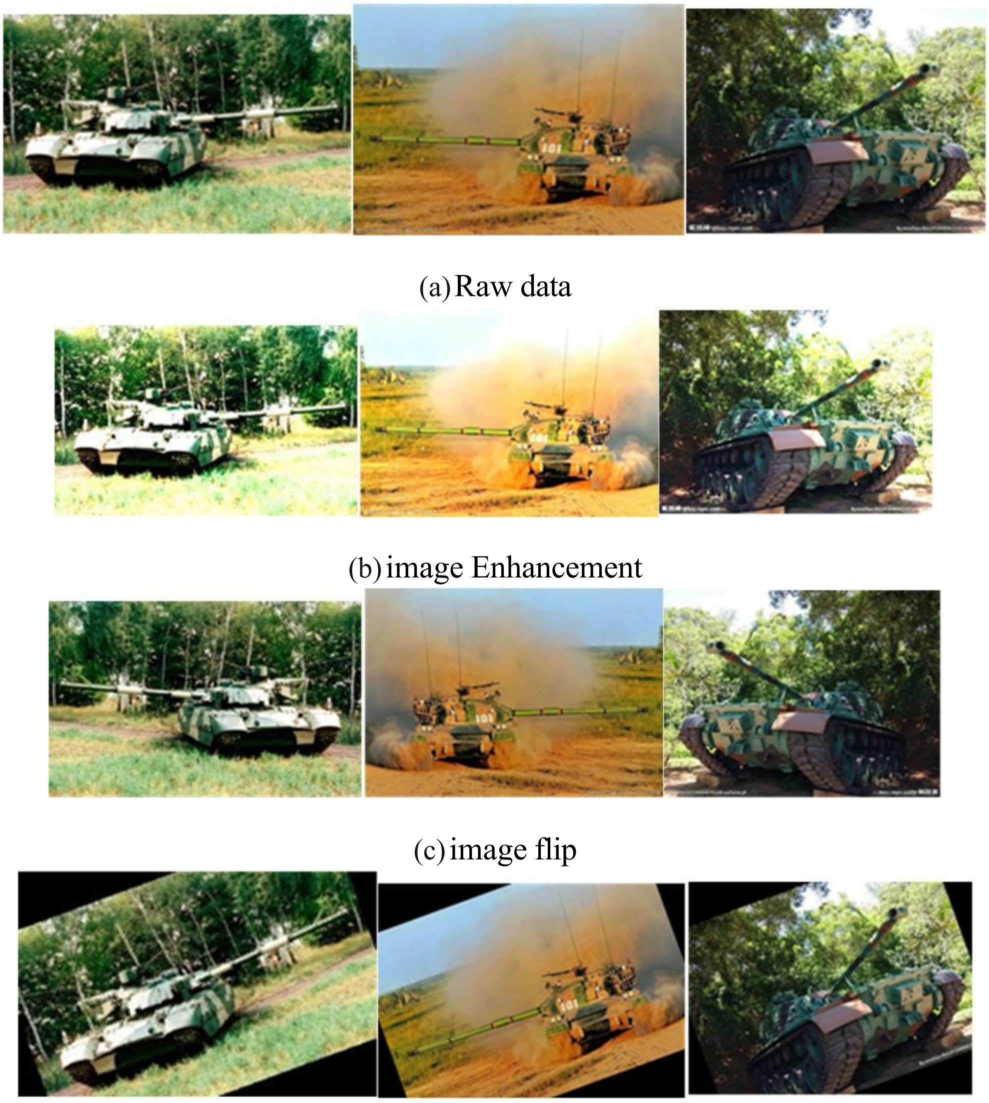
Before deep learning methods are widely used, traditional target detection algorithms are dominant, and classical feature operators include Hal feature [1], SIFT operator [2], HOG feature [3], LBP feature [4], etc. With the development and progress, Hinton [5][6] first discovered the deep learning method in 2006, and found that concolution neural network (CNN) has strong ability to learn and extract features. Compared with the traditional method, This method can gradually extract the abstract features of the target from the shallow to the deep, and can classify and recognize different data. In 2014 [7] Girshick et al proposed R-CNN (regions with CNN) algorithm for the first time to use CNN to detect the target. Compared with the traditional artificial model detection algorithm, the experimental results on the VOC2012 test set The mean average precision of R-CNN algorithm reaches more than 50%, and the detection accuracy of some improved algorithms Fast R-CNN[8] and [9]Faster R-CNN is also continuously improved. At present, the traditional detection model has been gradually replaced by the deep learning model, which provides a benchmark for the subsequent research.

In this paper, Faster R-CNN detection algorithm is chosen to study tank targets in complex environment, because the algorithm has better detection accuracy and speed. Firstly, the tank target data set is amplified to solve the problem of few target data sets. Secondly, in the complex background, the tank target is easy to be blocked, and the phenomenon of missing detection results in the decrease of detection accuracy. To solve this problem, this paper improved the current Faster R-CNN algorithm. First, the basic principle of aster R-CNN algorithm was introduced and resnet50 was used as the backbone network of the model. In order to extract more detailed features from the target, FPN network was added to the backbone network for feature fusion. Finally, the research of this paper is verified by experiment.

2. Database Construction

The section headings are in boldface capital and lowercase letters. Second level headings are typed as part of the succeeding paragraph (like the subsection heading of this paragraph). All manuscripts must be in English, also the table and figure texts, otherwise we cannot publish your paper. Please keep a second.

Traditional data enhancement methods mainly use preset data transformation rules to amplify data on the basis of existing data, and mainly do the original image to flip, rotate, shift, crop, deformation, scale and other operations. In this paper, the original 368 images were flipped, rotated, cropped and enhanced brightness respectively. Figure 2.1 shows part of the original data set and its generated images. The result is 1472 images. As shown in Figure 1.



(a) Raw data

(b) image Enhancement

(c) image flip

(d) image Rotation

Figure 1. Partial data sets

3. Tank Target Detection Algorithm

The Faster-RCNN algorithm generally uses VGG or ResNet as the backbone feature extraction network. Because this paper detects the target in a complex background, it increases the difficulty of detection and leads to an increased rate of missed detection.With the further deepening of the VGG network layer, the performance of the neural network will not continue to improve, and will even rise to the best state and then quickly fall down. In order to solve this problem, the residual network ResNet is used in this paper for feature extraction, and the extracted feature map is sent to the RPN and ROI Pooling layers to extract the suggestion box with uniform size. Finally, the tank target is etected through the full connection layer. Figure 2 shows the network model diagram of Faster-RCNN. As shown in Figure 2.

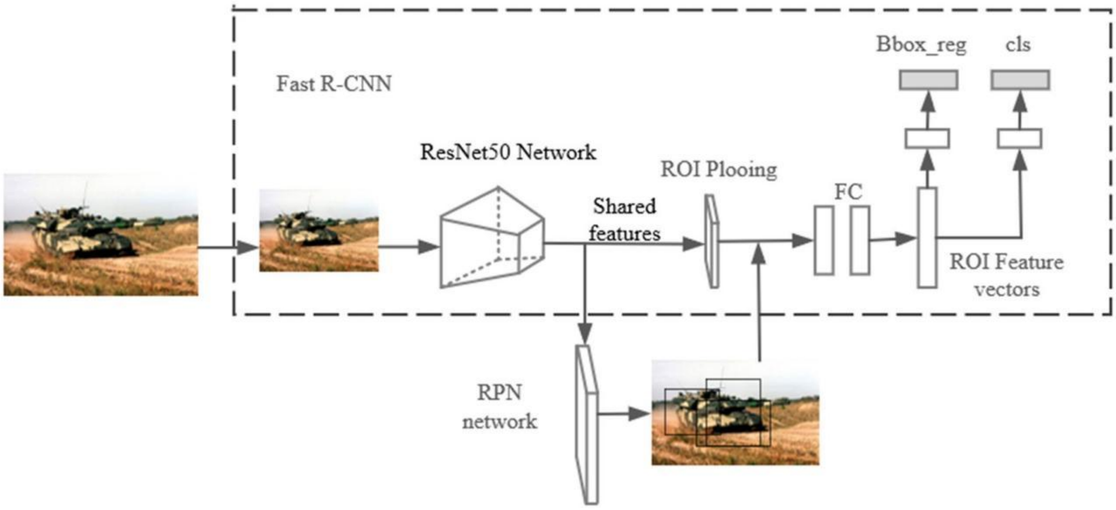


Figure 2. Schematic diagram of the Faster-RCNN network model

3.1 Feature Extraction Network

ResNet is a kind of residual network, which can also be understood as a seed network, which can form a deep network by stacking. The residual network has two structures, one is the shallow network as shown in FIG.3 (a), whose network layers are generally less than 34. The residual unit is composed of the convolution of two layers 3 and 3. Another kind of deep network is shown in Figure 3 (b). Due to its complex structure, the residual unit includes three layers of convolution, namely two layers of convolution of 1, 1 and one layer of convolution of 3, 3.

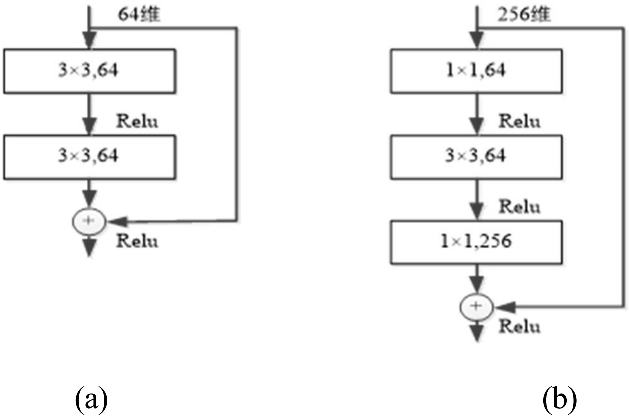


Figure 3. Residual unit

3.2 PN Network

In order to retain more detailed features of the target and reduce the rate of missed detection, this paper combines resnet50 network with FPN network, and then fuses higher-level feature maps with lower-level feature maps. Because lower-level features retain more edge information, while deep-level features retain more detailed features, the expression ability of image features can be improved through feature fusion. Its network structure is shown in Figure 4.

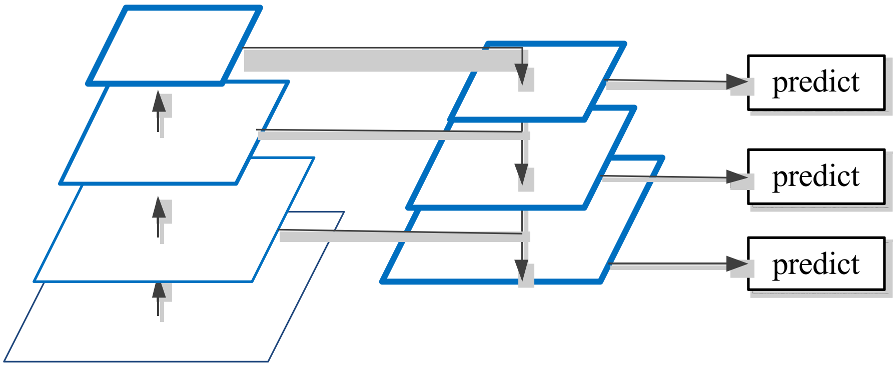


Figure 4. Characteristic pyramid structure

It can be seen from FIG. 4 that the input image generates feature maps of different levels through the convolutional neural network, and the size of feature maps will be continuously reduced. In FPN, a network segment is the first level of the pyramid, and all network segments form a complete feature pyramid. The output of the last layer of each network segment is taken as the feature diagram of the pyramid level. For ResNet50, there are 5 different network segments, so the output of the last residual unit of each network segment constitutes the first-level pyramid, and the fifth-level feature pyramid is formed after the introduction of FPN. The specific fusion process is shown in Figure 5.

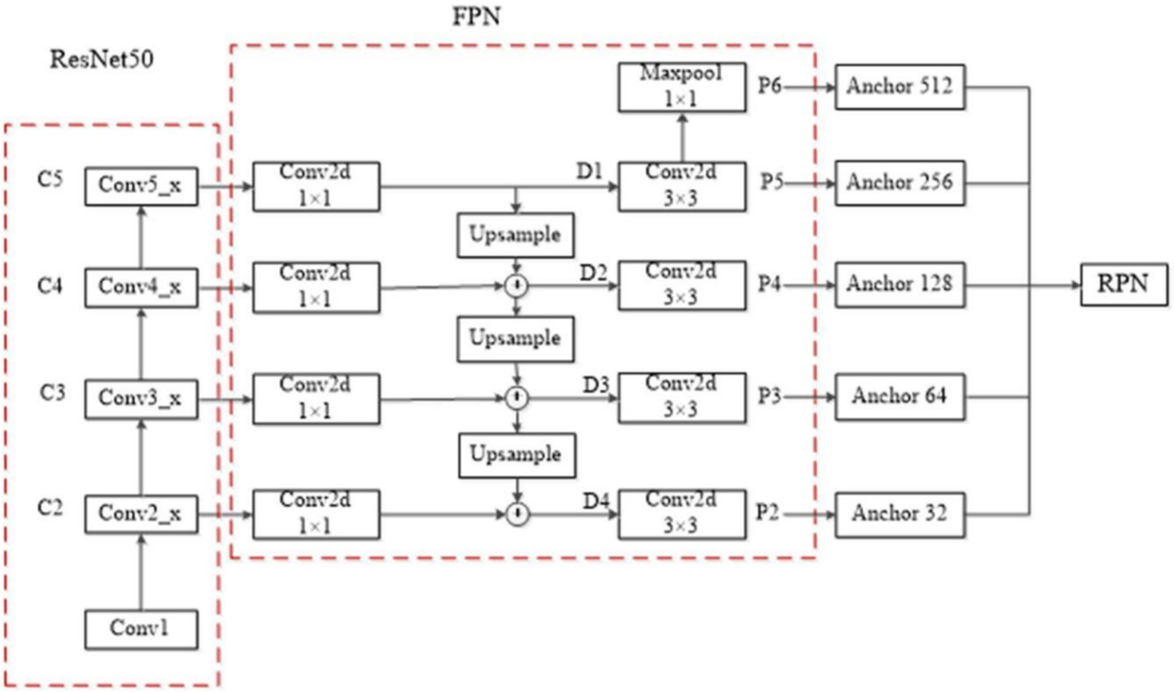
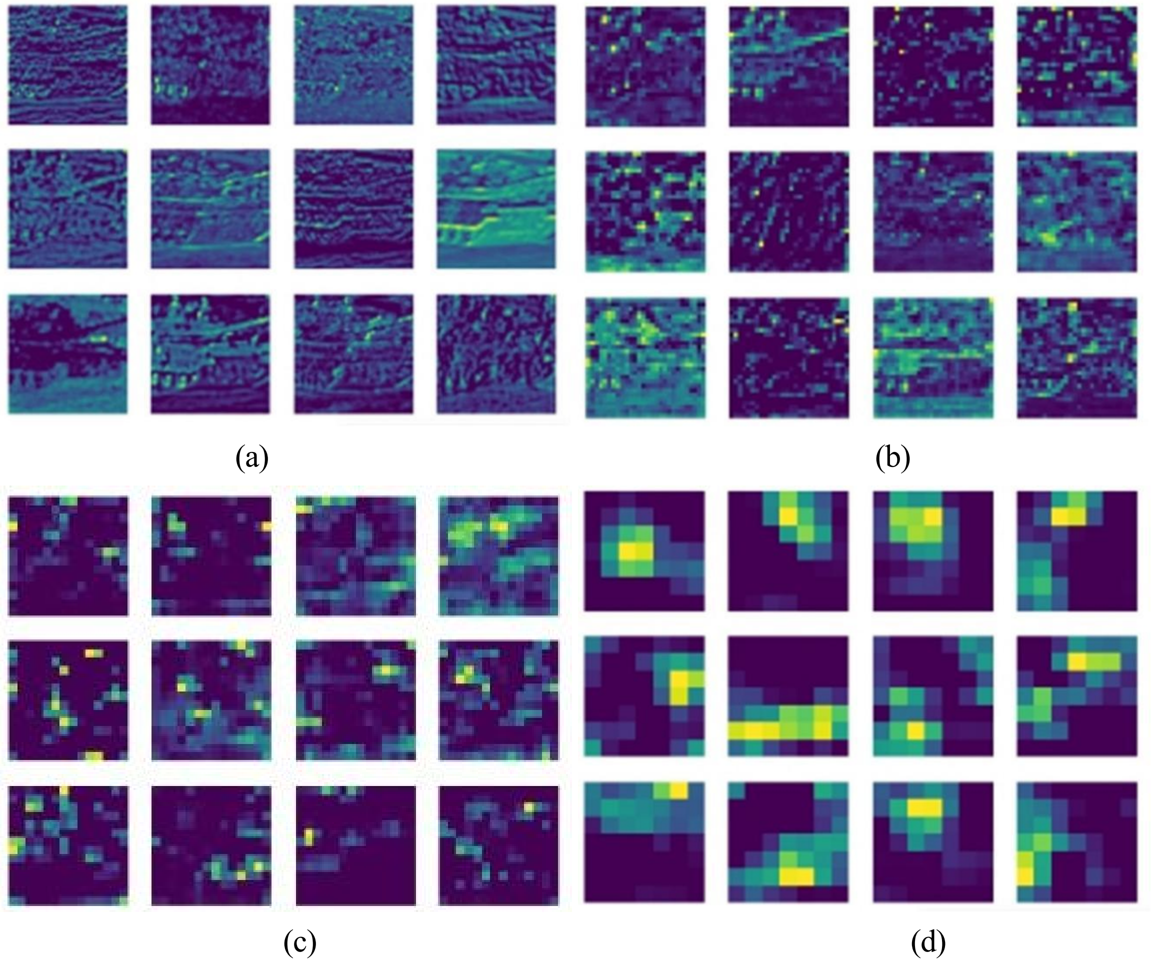


Figure 5. ResNet50+FPN structure diagram

After the input image passes through the ResNet50 network, four feature maps C2, C3, C4 and C5 will be obtained, and then the four feature maps will adjust the number of channels through a 1×1 convolution layer to prepare for the subsequent fusion. Then, C5 is up-sampled and fused with C4 to obtain the feature graph D2. Then, D2 is up-sampled and fused with C3 to obtain the feature graph D3. After up-sampling, D3 is fused with C2 to obtain the feature graph D4. Finally, feature maps D1, D2, D3 and D4 obtained by fusion are fused again through 3×3 convolution respectively to get feature maps P2, P3, P4 and P5, and P6 is obtained by downsampling of P5. Finally, P2, P3, P4, P5 and P6 are sent into RPN network for proposal prediction. The predicted proposal was mapped to the feature figures P2, P3, P4 and P5, and the final detection result was obtained through the Faster-RCNN detection network. The characteristics of C2, C3, C4 and C5 are shown in FIG. 6 (a), (b), (c) and (d).



(a)

(b)

(c)

(d)

Figure 6. Convolution feature diagram

As can be seen from FIG. 6, the feature maps of C2, C3, C4 and C5 contain different feature information. C2 retains more edge information and can clearly see the overall features of the tank. Different feature layers retain different feature information. Therefore, in order to retain more complete feature information, it is necessary to consider the low-level features as well as the deep-level features.

3.3 Region Proposals

In this paper, FPN is used to forecast multi-scale feature maps, and different anchors are set for feature maps of different scales. In FIG. 3.13, P2, P3, P4, P5 and P6 respectively correspond to the sizes of anchors 322, 642, 1282, 2562 and 5122, with 1:1 values for each anchor. In a ratio of 1:2 and 2:1, there are 15 candidate boxes. However, the statistical results showed that the aspect ratio of 99.94% of real frames was 1:2 and 2:1. Therefore, based on the relationship between anchor and real frame, the size of real frame and aspect ratio were calculated to set anchor specifications. In this paper, the ratio of anchor was set as 1:2 and 2:1, and there were 10 different candidate frames. Compared with the 9 candidate boxes set in RPN extraction of single-layer feature map, there is only one more candidate box. In this way, the detection rate can be improved without reducing the detection speed.

4. Experimental Results and Analysis

The performance of the tank target detection algorithm in this paper can be obtained by the performance test of the tank target on the test data set. The software simulation experiment on the computer can get the experimental data and compare the performance.

4.1 Performance Index

The trained model weight file is loaded into the tank target detection model, and the test set and verification set constructed in this paper are used to test and verify the two models.

In target detection tasks, Precision, Recall and Average Precision are commonly used evaluation indexes when evaluating the performance of a certain target detection algorithm. The meanings of these evaluation indexes are briefly introduced below. TP is True Positive, indicating the number of real tanks detected. FP is False Positive, that is, the number of tank targets that were mistakenly detected. FN is False Negative, that is, the number of real tanks that are incorrectly checked.

1 Precision

precision represents the proportion of real tank targets detected. The calculation formula is shown in Equation (1).

$$\text{precision}=TP/(TP+FP) \tag{1}$$

2 Recall rate

recall refers to the proportion of detected tank targets in the total tank targets. The recall rate represents the strength of the model's recognition ability. The calculation formula is shown in Equation (2).

$$\text{Recall}=TP/(TP+FN) \tag{2}$$

3 Average recognition rate (AP)

The average recognition rate AP is used to finally measure the trained model, the advantages and disadvantages of each category detection, and represents the area under the curve of confidence and recall rate in mathematical coordinates. This is an indicator that can be used to measure the accuracy of the category and position of the model prediction box. The calculation formula is shown in Equation (3).

$$AP=(1/c)\sum(TP/(TP+FP)) \tag{3}$$

4.2 Quantitative Analysis

In order to reduce the influence on target occlusion in complex environment and improve the accuracy of target detection, Resnet50 network and FPN network are fused in this paper, and the rejection loss function is introduced. The amplified data set is used. The average detection recognition rate of Faster R-CNN algorithm compared with the improved Faster R-CNN algorithm is shown in the table 1.

Table 1. Test results

Algorithm	AP(%)
Faster R-CNN	89.39
Improved Faster R-CNN	95.24

As can be seen from the table 1, the average recognition rate of the original algorithm is 89.39%, and that of the improved algorithm is 95.24%, with the average recognition rate increased by 5.85%. This result shows the effectiveness of the improved algorithm.

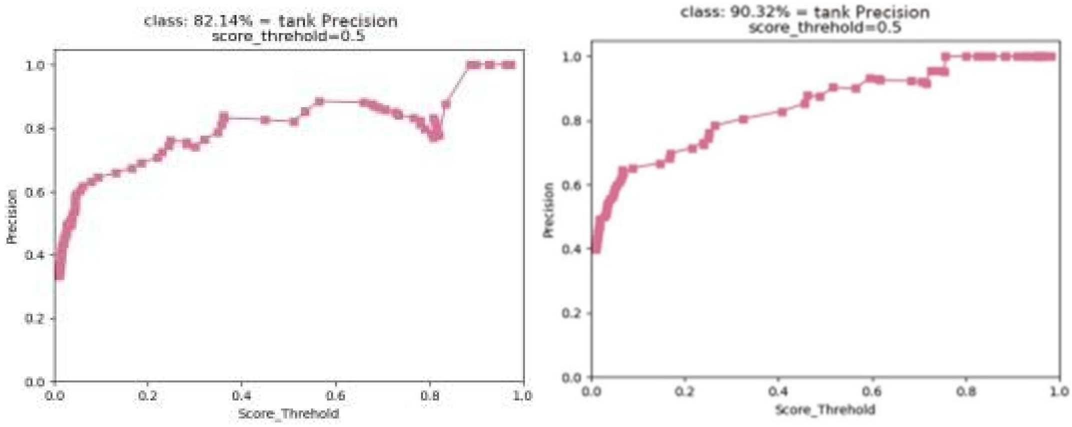
In order to improve detection accuracy and ensure detection speed, feature maps of different scales were allocated anchors of different sizes and proportions in this paper, and tests were carried out on the amplified data set. The time and accuracy of extracting candidate frames were shown in the table.

Table 2. Extract the candidate box time comparison

Algorithm	Andidate box	Time(/ms)
RPN	2000	245
Improved RPN	2000	251

As can be seen from Table 2, when the same number of candidate boxes are extracted, the proposed algorithm takes about the same time to extract candidate regions on multiple layers as it does only on the last layer of convolution. Therefore, the improved RPN algorithm adopted in this paper takes less time. Therefore, the improved RPN algorithm not only improves the detection accuracy, but also saves a lot of time for the overall test system.

Figure 7(a) shows the accuracy rate of the original Faster-RCNN network model, and Figure 7(b) shows the accuracy rate of the optimized Faster-RCNN network model.



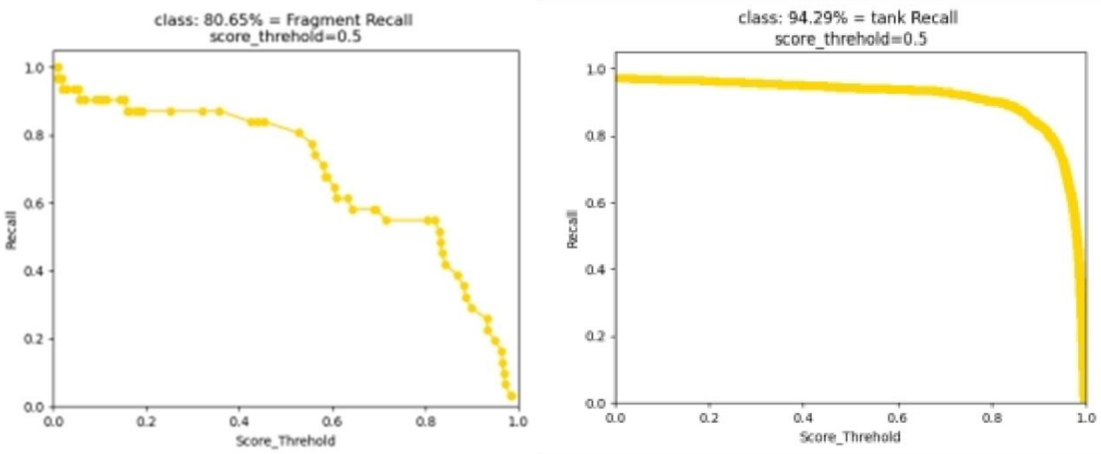
(a)

(b)

Figure 7. Accuracy rate of network model

As can be seen from Figure 7, the accuracy rate of the original farier-RCNN network model is 82.14%, and that of the optimized farier-RCNN network model is 90.32%. Compared with the original network model, the accuracy rate of the optimized network model is improved by 8.18%.

Figure 8(a) shows the recall rate of the original Faster-RCNN network model, and Figure 8(b) shows the recall rate of the optimized Faster-RCNN network model.



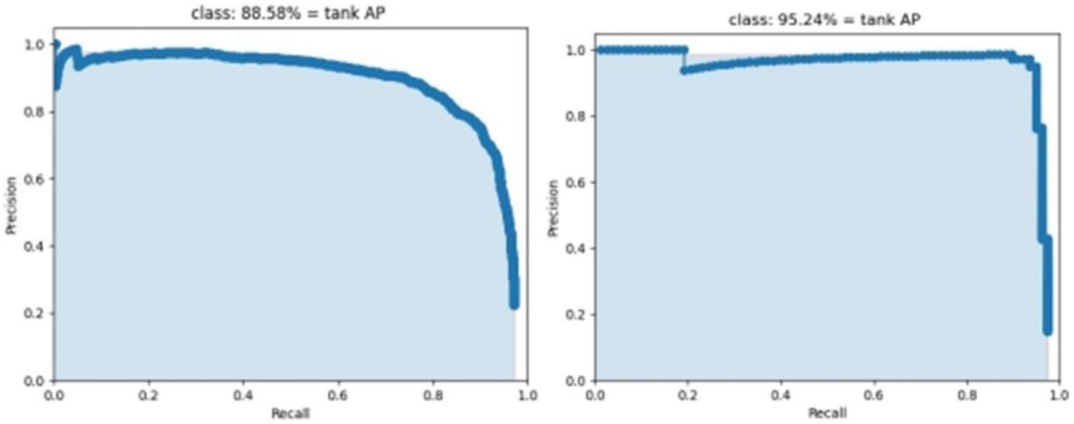
(a)

(b)

Figure 8. Network model recall rate

As can be seen from Figure 8, the recall rate of the original Faster-RCNN network model is 80.65%, and the recall rate of the optimized Faster-RCNN network model is 94.29%. Compared with the original network model, the accuracy rate of the optimized network model is improved by 13.64%.

Figure 9(a) shows the average recognition rate of the original Faster-RCNN network model, and Figure 9(b) shows the average recognition rate of the optimized Faster-RCNN network model.



(a) (b)

Figure 9. AP rate of network model

As shown in Figure 9, the average recognition rate of the original Faster-RCNN network model is 88.58%, and the average recognition rate of the optimized Faster-RCNN network model is 95.24%. Compared with the original network model, the average recognition rate of the optimized network model is improved by 6.66%.

The original network model is compared with the optimized network model from three indexes of target detection accuracy rate, recall rate and average recognition rate. The detection results are shown in Table 3

Table 3. Tank target detection performance comparison table

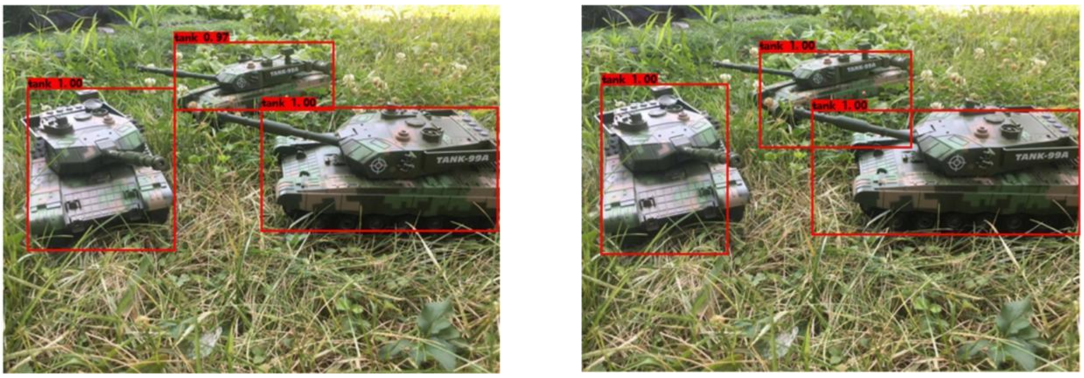
Algorithm	AP	recall	precision
Original network model	88.58%	80.65%	82.14%
Improved network model	95.24%	94.29%	90.32%
Ascension	6.66%	13.64%	8.18%

As can be seen from Table 3, in tank target detection, the average recognition rate increases by 6.66%, recall rate by 13.64%, and accuracy rate by 8.18%. Compared with the original model, the optimized model has improved to some extent, indicating that the optimized network model has better detection effect and higher accuracy than the original network model.

4.3 Qualitative Analysis

In order to compare the tank detection results under different models, the tank model images taken by mobile phones are used as the test set. The detection effects of each algorithm can be seen through the position and detection accuracy of the tank marker box in the images, so as to compare the one.

First, the tank target is detected in an unobstructed environment. The detection results are shown in Figure 10. Figure (a) is the detection result of the original network model, and Figure (b) is the detection result of the optimized network model.

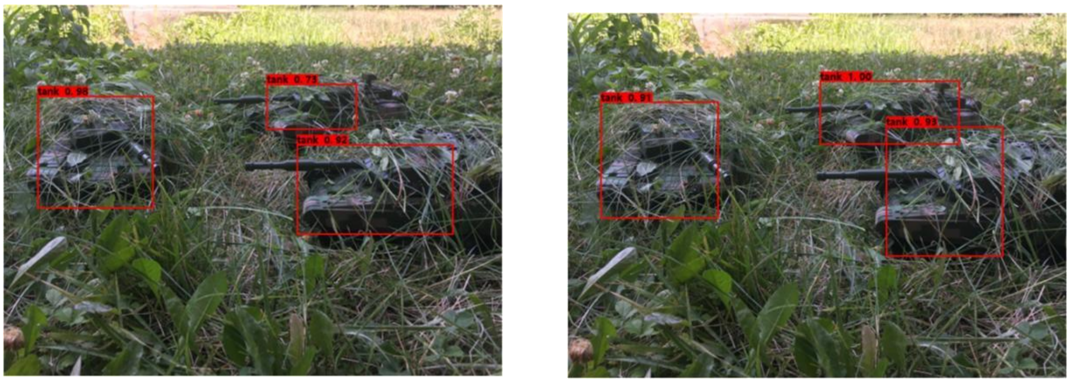


(a) (b)

Figure 10. Unsheltered

It can be seen from FIG. 10 that in the detection of tank targets in an unobstructed environment, both the original network model and the optimized network model have good detection effects, with confidence of more than 95% and accurate target positioning, but the overall detection confidence of the optimized network model is higher and the detection effect is better.

The tank target was detected in the environment of 30%~60% occlusion, as shown in Figure 10. Figure (a) is the detection result of the original network model, and Figure (b) is the detection result of the optimized network model.

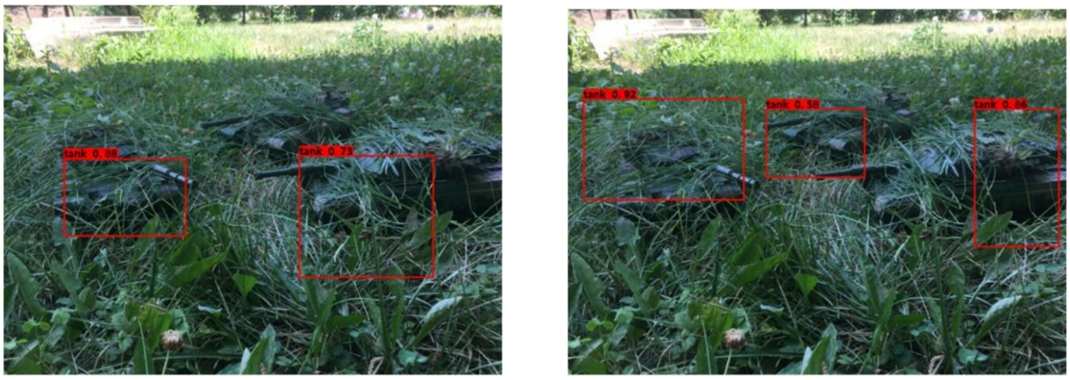


(a) (b)

Figure 11. Shelter 30%~60%

As can be seen from FIG. 11, in the environment of 30%~60% occlusion, with the increase of occlusion, the confidence of some detection targets in the original network model decreases a lot, while the confidence of detection targets in the optimized model network remains above 90%. Therefore, overall, the optimized network model has a better detection effect.

In the environment of 70%~100% occlusion, the tank target is detected, as shown in FIG. 12. FIG. (a) is the detection result of the original network model, and (b) is the detection result of the optimized network model.



(a) (b)

Figure 12. Shelter 70%~100%

As can be seen from FIG. 12, when the tank target is detected in the environment of 70%~100% occlusion, it is difficult for human eyes to distinguish the tank target in this environment. Compared with the detection results in the previous two environments, the detection accuracy in this environment is decreased, and the phenomenon of missing detection appears in the original network model. However, there is no missing detection phenomenon in the optimized network model, and the position of the marker box is accurate. Therefore, overall, the optimized network model has a better detection effect, and to some extent, it solves the problem of missing detection of tank targets in the complex background.

5. Conclusion

In order to solve the problem of tank target occlusion and camouflage under complex background, the Faster R-CNN algorithm is studied and improved in this paper. In order to improve the target detection accuracy and reduce the rate of missed detection, the backbone network and FPN network were fused to optimize the candidate region of the feature map. The experimental results show that the improved method still has high effectiveness and robustness in the case of severe occlusion, and the average recognition rate is up to 95.24%. Compared with the original algorithm, the detection accuracy has been greatly improved.

References

[1] H, Yu. Robust characteristics based on the improved fast algorithm of face detection research [J]. Journal of Nanjing university of science and technology, 2017, 9 (6): 714-719. The DOI: 10.14177 / j. carol carroll nki. 32-1397 n. 2017.41.06.008.C. Li, W.Q. Yin, X.B. Feng, et al. Brushless DC motor stepless speed regulation system based on fuzzy adaptive PI controller, Journal of Mechanical & Electrical Engineering, vol. 29 (2012), 49-52.

[2] Stefan G. Stanciu, Radu Hristu, Radu Boriga, George A. Stanciu. On the Suitability of SIFT Technique to Deal with Image Modifications Specific to Confocal Scanning Laser Microscopy[J]. Microscopy and Microanalysis,2010,16(5).

[3] Erazo-Aux Jorge, Loaiza-Correa H, Restrepo-Giron A D. Histograms of oriented gradients for automatic detection of defective regions in thermograms. [J]. Applied optics,2019,58(13).

[4] Timo Ojala, Matti Pietikäinen, Topi Mäenpää. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. [J]. IEEE Trans. Pattern Anal. Mach. Intell.2002,24(7).

[5] Hinton G E, Osindero S, Teh Y W. A fast-learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18: 1527-1554

[6] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.

[7] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2014: 580-587.

[8] Girshick R. Fast R-CNN[C] //Proceedings of the IEEE Interna-tional Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 1440-1448.

[9] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelli-gence, 2017, 39(6): 1137-1149.

[10] GIRSHICK R. DONAHUE J, DARRELL T, et al. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(1):142-158.

[11] GIRSHICK R. Fast R-CNN [C] //IEEE International Conference on Computer Vision. USA: IEEE Computer Society,2015:14440-14448.

[12] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEE. Transactions on Pattern Analysis & Machine Intelligence,2015,39:1137-1151.

[13] EVERINGHAM M, COOL L V, WILLIAMS C K I, et al. The Pascal Visual Object Classes(V0C) Challenge [J]. International Journal of Computer Vision,2010,88(2):303-338.

[14] JIA, Yangqing. Caffe: Concolutional Architecture for Fas Feature Embedding [J]. Eprint Arxiv, 2014 (6): 675-678.

[15] Ziming Liu, Guangyu Gao, Lin Sun, Zhiyuan Fang. HRDNet: High-resolution Detection Network for Small Objects[C]// 2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society,2020.