Sensing and Imaging (2021) 22:24
https://doi.org/10.1007/s11220-021-00348-0

ORIGINAL PAPER

A Deep Lightweight Convolutional Neural Network Method for Real-Time Small Object Detection in Optical Remote Sensing Images

Yanyong Han[1] · **Yandong Han**[2]

Received: 25 May 2020 / Revised: 13 April 2021 / Accepted: 13 May 2021

Abstract

The existing object detection algorithms mainly detect large objects. There are few kinds of researches on small objects. And the small object detection accuracy is low, which is difficult to satisfy the real-time requirement. Therefore, this paper proposes a deep lightweight convolutional neural network method for real-time small object detection in optical remote sensing images. Firstly, we build a data set with small objects. The object occupies a very small proportion in the image. There are some interferences such as truncation and occlusion in this data set, which can better evaluate the advantages and disadvantages of the detection method of small objects. Secondly, combining with the region proposal network, we propose a method for generating high-quality candidate boxes of small objects to improve the detection accuracy and speed. Meanwhile, two new learning rate strategies are proposed to improve the performance of the model and further improve the detection accuracy. Experimental results show that the proposed method is an effective small object detection algorithm and achieves real-time detection effect compared with other state-of-the-art object detection methods.

Keywords Deep lightweight convolutional neural network · Small object detection · Optical remote sensing · RPN

✉ Yanyong Han
yslinhit@163.com

Yandong Han
hanydvip@163.com

[1] School of Mechanical Engineering, Zhengzhou University of Science and Technology, Zhengzhou, China

[2] School of Electrical Engineering, Zhengzhou University of Science and Technology, Zhengzhou, China

Published online: 25 May 2021

1 Introduction

In practical applications such as resource exploration, earthquake and fire rescue based on aerial photography, the imaging of objects is relatively small due to the distance of shooting, while complex background information will interfere with the detection. How to detect such small objects in real-time has become a difficult and hot issue in current research.

In recent years, various object detection algorithms such as Faster RCNN, SSD, YOLOv2, etc. [1–4]. have made significant achievements in the field of computer vision, which is manifested in the continuous improvement of detection performance on PASCAL VOC and other common data sets [5]. The objects contained in the images of these common data sets usually occupy a relatively large proportion in the whole image. However, according to the evaluation of literature [6], it is found that the object detection algorithm mentioned above has poor accuracy in testing small objects in the image and cannot meet the needs of small object detection applications.

Some scholars have done some researches on the problem of small object detection. Chen et al. [7] combined context information with the RCNN algorithm for small object detection, which improved the test accuracy compared with the traditional object detection algorithm, but it still had the problems of low efficiency and large storage space. Subsequently, some researchers applied the improved RCNN algorithm (Fast RCNN and Faster RCNN algorithm) to small object detection to improve the accuracy and the detection speed. References [8–10] used the context information of Fast RCNN to detect small objects to improve the detection performance. Zhang [11] used Faster RCNN to detect pedestrians and analyzed that the error of pedestrian detection mainly came from low-resolution feature maps and background interference. Then it improved the detection accuracy by modifying the region proposal network (RPN). Chen [12] used Faster RCNN to detect the small object of the company logo, and further analyzed the influence of object size and feature graphs in different levels on the detection effect. In addition, some scholars also designed a new network structure for small object detection. In reference [13], an end-to-end convolutional neural network was proposed to detect small traffic signs, which was better than the Fast RCNN algorithm in accuracy and speed. Ren [14] investigated on how to modify Faster R-CNN for the task of small object detection in optical remote sensing images. Rabbi [15] applied a new edge-enhanced super-resolution GAN (EESRGAN) to improve the quality of remote sensing images and used different detector networks in an end-to-end manner where detector loss was back-propagated into the EESRGAN to improve the detection performance. Although these studies have made a lot of achievements in the detection of small objects and provided a lot of novel ideas, the small objects still occupy a relatively large proportion in the image, and the real-time processing is not up to the requirements.

Therefore, this paper mainly studies the real-time small object detection problem. For small objects, this paper does not limit them to smaller objects in the

real world, but refers to small objects in a broad sense, that is, those objects occupy a small proportion in an image.

The following difficulties should be considered in the detection of small objects. First of all, compared with the whole image, the object to be detected accounts for a small proportion, and background information will cause a great interference to the detection, which will greatly increase the difficulty of accurately locating the small object. Secondly, compared with the larger object, the smaller object has fewer pixels, so less effective feature information can be extracted. In addition, the number of small and medium-sized objects in an image is often large and overlap with each other in practical applications, which further increases the difficulty of detection.

PVANet is a deep and lightweight convolutional neural network that can be used for real-time object detection [16]. It adopts the feature extraction network to generate feature vector graphs and then generates high-quality region proposals based on the RPN for subsequent object detection and localization. The test results on common data sets such as PASCAL VOC show that the performance of the PVANet algorithm is better than that of Faster RCNN, SSD, YOLOv2, and so on. In particular, the convolution kernel in the feature extraction layer of PVANet is small, so we can keep as many low-level features as possible, which is beneficial for small object detection. For example, Fig. 1 is an example for the detected small occlusion object with the improved PVANet. The original PVANet cannot detect them.

To sum up, this paper improves the detection performance of small objects by improving the PVANet algorithm, and the main contributions are as follows.

Building a benchmark data set dedicated to small object detection. Compared with the data set used in other small object detection studies, the object in this data set occupies a smaller proportion in the image, and incomplete object information such as truncation and occlusion will increase the detection difficulty, which can train the small object detection model with more stable performance.
Aiming at the problem of poor localization of small objects with the original PVANet algorithm, a new method to generate high-quality candidate boxes of small objects is proposed, which improves the accuracy and speed of detection.
Besides, according to the characteristics of the training model, two new learning rate strategies are selected to further improve the performance of the model.

(a) **(b)** **(c)**

Fig. 1 Some small objects detected by improved PVANet. **a** Detected small ship of about 10×10 pixels. **b** Detected small airplane of about 20×20 pixels. **c** Detected occluded airplane

2 Building a Small Object Data Set

Data set is the key factor to analyze and evaluate the network model based on deep learning. Neovision2 Tower data set [17] is built by the Defense Advanced Research Projects Agency (DARPA) for the object detection and real-time tracking of video image data sets. Due to the shooting distance, the size of the data set is small. The multiple and chaotic objects in the shooting scene with variable lighting and blocking interference make it become a challenging object detection data set. Therefore, the Neovision2 Tower data set is selected in this paper to build a small object dataset. Neovision2 Tower data set contains 100 video clips, and each video clip has been captured into 900 PNG photo sets with high resolution 1920×1080pixel, which can retain as much small object information as possible. This paper mainly constructs small object data sets from the following aspects. In order to improve universality, the format refers to PASCAL VOC.

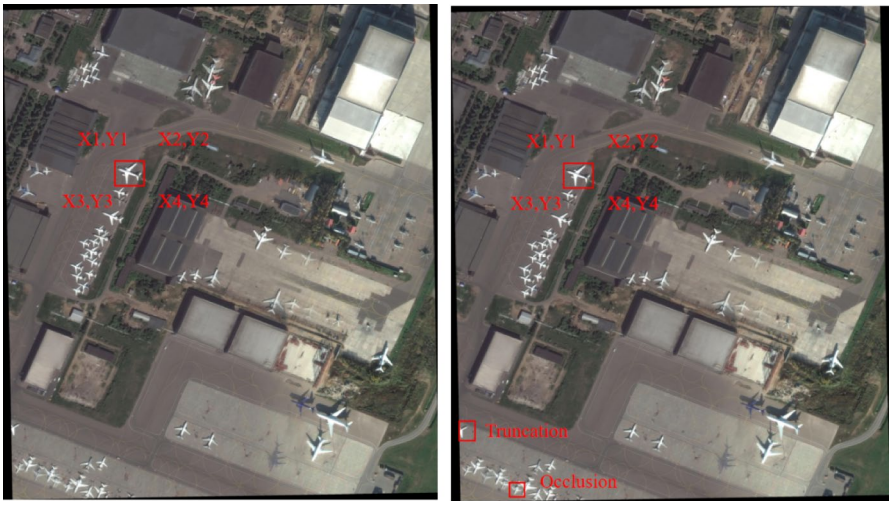
a. Image dimension reduction. It changes the size as 960×544 pixel and compresses it into.jpg format to make the data format the same as PASCAL VOC, which is necessary to speed up the training process of the network model.
b. Modifying the bounding box of the object. The labeling information of the object in the original data set is composed of four boundary coordinates (X1,Y1), (X2,Y2), (X3,Y3), (X4,Y4). Although the four coordinates form a boundary box as close to the object as possible, it is not rectangular and does not conform to the generic data set specification. For this reason, the labeling box is modified as a rectangular box, and it is determined by the upper-left coordinate (Xmin,Ymin) and the lower-right coordinate (Xmax, Ymax), as shown in Fig. 2. The new coordinate can be expressed as:

$$\begin{aligned} X\min &= \min(X1, X2, X3, X4) \\ Y\min &= \min(Y1, Y2, Y3, Y4) \\ X\max &= \max(X1, X2, X3, X4) \\ Y\max &= \max(Y1, Y2, Y3, Y4) \end{aligned} \tag{1}$$

In Fig. 2b, elliptic labeling is also used to illustrate the presence of interference such as truncation and occlusion of the object in the constructed dataset.

Finally, the processed data set is divided into two subsets: the training set and the testing set, each contains 50 images. Then 10 images are randomly selected from the training set as the verification set, and the remaining 40 images are used for training. This paper selects small aircraft objects as research objects.

The small object data set is from Google earth and then rotated. This method can increase the diversity of training data and make the test model have better generalization performance. In the small object data set constructed in this paper, the average size of the aircraft is 18×25 pixels, accounting for about 0.055% of the entire image with 970×540 pixels. The average object proportion in the small object data set proposed in this paper is 0.176%. Compared with the PASCAL VOC data set and the small object data set proposed in reference [18], the object proportion is smaller.



(a) Original labeling data **(b)** The proposed labeling data

Fig. 2 Diagram of the bounding boxes of the object in tow datasets

The small object data set proposed in reference [18] is an excellent example for constructing data sets, and has been adopted by many researchers for small object detection. Compared with reference [18], the small object data set designed in this paper has at least the following two advantages. Firstly, the object in the small object data set constructed in this paper is smaller, and there is also passive occlusion of the background and active occlusion between the objects. In addition, the object will be truncated at the boundary of the image, which increases the difficulty of small object detection and better evaluates the advantages and disadvantages of the model in small object detection. Secondly, the data set designed in this paper can be used to continuously sample the shape of objects, which is conducive to training a more robust detection model because of the temporal information and correlation between images [19].

3 Improved PVANet for Small Object Detection

3.1 PVANet

PVANet is a lightweight object detection algorithm, which is implemented in two phases. Firstly, the feature extraction network outputs feature graph to RPN and generate the object candidate box. Secondly, the object candidate box and feature map generated in the previous stage are sent to the classification layer after the pooling layer and the full connection layer to determine the type of object and the boundary box regression layer, which further adjusts the position of the object border. The overall framework structure of PVANet is shown in Fig. 3.

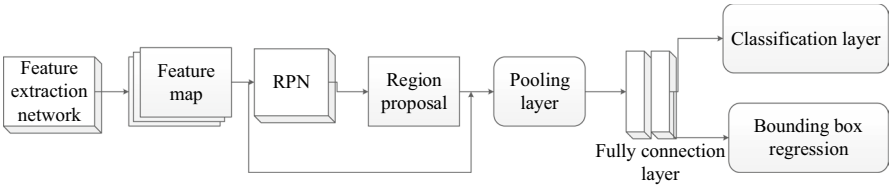


Fig. 3 PVANet architecture

The main contribution of PVANet is to propose an efficient feature extraction network. Based on the design principle of more layers but few channels, the feature maps are generated by using the ReLU, Incepionl, HyperNet, and residual connections. It achieves the goal of accelerating model performance without compromising detection accuracy. The feature extraction network of PVANet is shown in Fig. 4.

The first few layers of PVANet feature extraction network are composed of ReLU modules. It is found that there is a negative correlation between the convolution kernel of the first few layers in the convolutional neural network (CNN). ReLU simply connects the feature of the output value of each convolution kernels and its negative value. Then it is zoomed or shifted and conducts ReLU calculation, which makes the slope of each channel and the activation threshold be different from its opposite channel. It also reduces the number of output channels by half, eliminating the need to store the parameters of their opposite channels without losing accuracy. The adoption of ReLU module is an important reason why PVANet can achieve lightweight effect. In addition, the Inception module is used for the remaining feature extraction network. As one of the most cost-effective artifacts that can capture both small and large objects in an image, Inception modules can generate activation values for different sizes. In particular, the 1×1 convolution kernel in the Inception module helps locate small object candidate boxes and capture small objects more accurately.

In conclusion, for real-time small object detection, PVANet is superior to other algorithms in the following three aspects.

a. First, it uses the ReLU module to reduce computation and improve detection speed.
b. In addition, it adopts RPN networks to generate high-quality object candidate boxes.

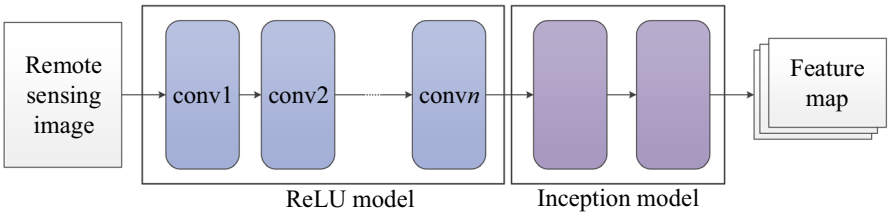


Fig. 4 PVANet feature extraction network

c. In particular, the using of the Inception module enables it to store as much information as possible about the underlying network, which is beneficial for small object detection.

4 Improved PVANet

The previous work pointed out that the challenge of small object detection mainly came from the generation of object candidate boxes. Therefore, this paper focuses on generating high-quality small object candidate boxes, mainly by setting appropriate anchor boxes in the RPN network. In addition, compared with other super parameters, learning rate is one of the most important parameters that affects the object detection performance and controls the effective capacity of the model in a more complex way. When the learning rate is optimal, the effective capacity of the model is maximum. Based on this, this paper will compare different learning rate strategies and choose the optimal strategy fine-tuning model to improve the performance of small object detection. Figure 5 is the structure of improved PVANet. It mainly consists of feature extraction network and candidate object extraction network.

5 Generate Small Object Candidate Boxes

Selective search and edge box are commonly used to generate object candidate boxes in object detection and have achieved better results on common data sets such as PASCAL VOC. However, Selective search is slow to generate object candidate boxes taking about two seconds per image on the CPU. Although Edge Box strikes a good balance between the quality of generating object candidate boxes and the processing speed, it still requires 0.2 s. Compared with the whole object detection, the above two methods consume too much time to generate the object candidate box, so they cannot meet the real-time requirements.

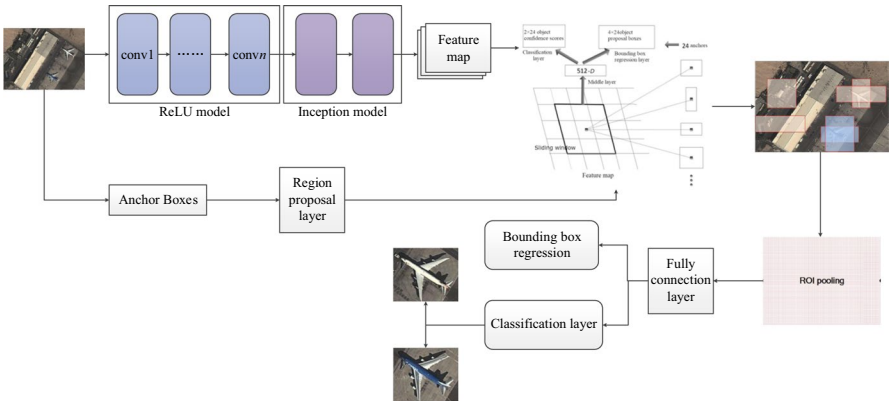


Fig. 5 The structure of improved PVANet

Additionally, Selective Search and Edge Box perform well when generating candidate boxes for large objects, but it has poor effect when generating candidate boxes for small objects. Because these two methods are sensitive to the important features of the object such as the outline and unique color. While the small object usually contains little information. Therefore, these two methods cannot generate high-quality candidate boxes for the small object.

RPN has been proved to be the optimal method to generate object candidate boxes, which greatly reduces the generation time of object candidate boxes [20]. It outputs 512-dimensional features by applying the 3×3 sliding window and anchor box to the feature graph generated by the feature extraction network, and then inputs them to the following two sub-full connection layers, the classification layer and the boundary box regression layer. The classification layer predicts the probability of the object candidate box and the probability of the background. The boundary box regression layer outputs 4 position coordinates of the object candidate box. In the original PVANet, 25 anchor frames are generated at each position of the sliding window, determined by five different scales (95,195,300,512,750) and five different aspect ratios (0.4,0.656,1.0,1.4,1.9). In the small object data set constructed in this paper, the average size of the aircraft is 18×25 pixel and 40×38 pixel respectively. Obviously, the original scale of RPN is too large for the small object in this paper, and the accuracy is poor when it is directly used to detect the small object. So the anchor frame needs to be reduced to fit the size of the small object. The modified PVANet adopts 6 scales (32,48,80,144,256,512) and 7 aspect ratios (0.332.0.4,0.658,1.0,1.4,2.0,2.9) to form 42 anchor frames. The new PVANet increases the number of anchor frames to expand the scope of object detection and improves the average accuracy when testing on PASCAL VOC by nearly 3% compared to the original PVANet. However, the comparison shows that the size range of these anchor frames is too large and its smallest size is larger than the average size of the small object constructed in this paper.

The size of the object in the small object data set constructed in this paper does not change significantly, especially the size difference of the small object aircraft is no more than 20 pixel. Therefore, the number and size of the anchor frame are reduced in this paper. However, since the shape of the aircraft's boundary frame is mainly rectangular, the anchor frame is kept as vertical and horizontal as possible to locate small objects more accurately.

Therefore, this paper selects 24 anchor frames for each position of the sliding window, including 4 sizes (16,24,32,64) and 6 aspect ratios (0.332,0.4,0.658,1.4,2.0). The RPN structure is shown in Fig. 6.

The detailed processes of the proposed method are as followings:

1. Using the ReLU model to extract features of the first three layers;
2. Using Inception module for the remaining feature extraction network;
3. Proposed RPN is used to generate small object candidate boxes;
4. Through classification and regression of the improved RPN network, a series of region proposals that may contain objects are output, and use the Soft-NMS to eliminate some overlapping region proposals in the Proposal layer.

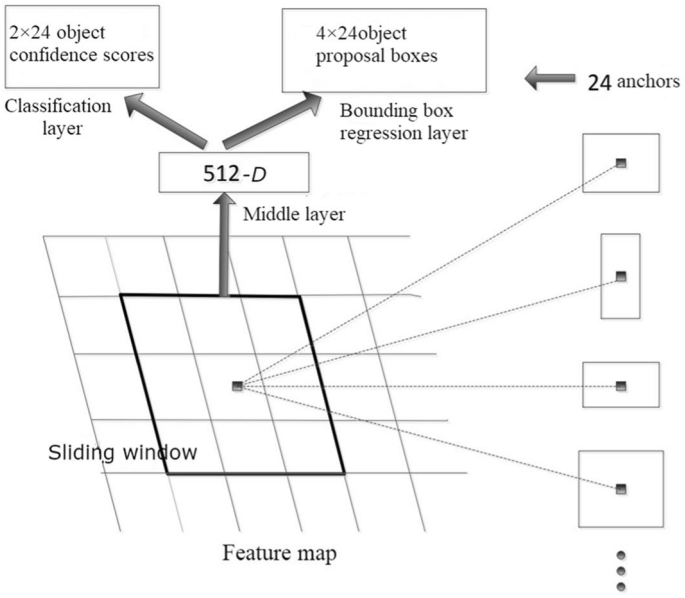


Fig. 6 RPN structure

5. Finally, through the full connection layer into the subsequent bounding box regression, the border trimming and specific category classification by softmax are performed.

6 Selecting a New Learning Rate Strategy

Learning rate is a very important super parameter in deep learning, which can guide researchers to adjust the weight of network through the gradient of loss function. Generally speaking, a better learning rate strategy means that a better network model can be trained in a shorter time. So adjusting the learning rate is one of the important means to improve the performance of the model through the training process.

PVANet dynamically controls the learning rate based on the "plateau" strategy. This policy monitors the average value of changes in the loss function. If it is found that in a certain iteration period, the improvement is lower than a certain threshold, then the change of the loss function is determined to be at a "plateau" and the learning rate is reduced by a constant factor. However, this paper first adopts the "plateau" learning rate strategy to train the model, and sets the number of iterations as 10^5. We find that the learning rate remains unchanged at the initial value of 0.001. The main reason for the constant learning rate is that the object area is very small compared with the background area in the training process, which results in a large negative sample space and slow convergence of the model itself. Therefore, the "plateau" strategy training model is directly adopted to change the learning rate by evaluating the dynamic mean value of the loss function. It is hard to change the learning rate. Therefore, other learning

rate strategies should be adopted to change the learning rate to accelerate the model convergence. After observing the curve of the loss function, it is found that it tends to flatten after 50,000 iterations, and the test finds that the detection accuracy improves very slowly after 50,000 iterations. Therefore, when training with 50,000 iterations, the gradient of the loss function is close to the "plateau" state. It is difficult to improve the training loss. In order to avoid loss function with the "plateau" state, the "step" learning rate strategy is adopted. The calculation formula is defined as:

$$LR = base\_lr \times \delta^{(floor(iter|stepsize))} \tag{2}$$

where LR denotes the learning rate. *base_lr* refers to the initial learning rate. δ and stepsize are parameters, and *iter* is the number of iterations. When the number of iterations reaches an integer multiple of stepsize, the learning rate begins to decrease. In this paper, the initial learning rate is set to 0.001, which is reduced to 0.0001 after 50,000 iterations. After 10,000 iterations, the detection accuracy is improved by about 0.45% compared with the "plateau" learning rate strategy.

Although a low learning rate ensures that one does not miss any minimum value, it also means that more time has to be spent converging the model, especially if the loss function falls into the "plateau state". Smith [21] indicated that the difficulty to reduce the loss mainly came from the saddle point rather than the local minimum point on the error surface. With this factor in mind, this paper tries another learning rate strategy—inv, which dynamically changes the learning rate to accelerate the convergence of the loss function, rather than uniformly changing it like the "step" strategy. The initial learning rate is set to 0.001, the parameter δ and power value are set to 0.0001 and 0.75 respectively. The "inv" learning rate formula is defined as:

$$LR = base\_lr \times (1 + \delta \times iter)^{(-power)} \tag{3}$$

Experimental results show that using the "inv" learning rate strategy can obtain better performance network model than "plateau" and "step" strategy, which is more suitable for the small object detection in this paper.

7 Experimental Results and Analysis

In this paper, GEFORCE GTX 1060 GPU is used for related experiments. First, for the methods generating object candidate boxes, we conduct comparison with proposed RPN and ERPN [22], GPRPN [23] and RASRPN [24]. The evolution index is IoU. Bigger IoU denotes better candidate box. The results are shown in Table 1, which illustrates that the subsequent processing with proposed RPN is better than that of other methods.

Table 1 IoU comparison with different RPN-based methods

Method	ERPN	GPRPN	RASRPN	Proposed RPN
IoU/%	52.6	58.4	62.5	71.6

In order to evaluate the effectiveness of the algorithm in small object detection, average precision (AP) and mean average precision (mAP) of all categories are used as evaluation indexes to measure the performance of the model. AP is the most intuitive index to evaluate the detection accuracy of a single category, while mAP is the mean value of all categories of AP, which can evaluate the comprehensive performance of the model. Table 2 displays the comparison between the proposed method and the state-of-the-art object detection algorithm including Faster RCNN, SSD, YOLOv2 and PVANet in terms of the accuracy and running time. Also MCAA [25], GAN [26], AAB [27] are compared. Figure 7 is the detection result with the proposed method.

1. Detection effect analysis with different methods

Existing object detection algorithms based on deep learning can be roughly divided into two categories. One category generates object candidate regions first, and then performs object classification and object boundary box prediction, represented by Faster RCNN, PVANet, etc. This kind of algorithms can locate the object well, but the detection speed is slow. The second is the end-to-end object detection framework that can directly predict object confidence score and object boundary box, represented by the YOLO and SSD algorithm, which has the advantages of simple network structure and fast test speed. But it cannot well determine object location, especially the accuracy is poor for an adjacent or similar object.

As can be seen from Table 2, the detection accuracy of PVANet and Faster RCNN algorithm on small objects is higher than that of YOLOv2 and SSD. SSD combines YOLO and Faster RCNN algorithm to predict the object boundary box with multi-scale idea, which improves the test accuracy. The proposed generating candidate box method for small object in this paper can greatly improve the performance, because the features of small object detection are fully considered. By comparing the various algorithms in Table 2, it can be found that the proposed method in this paper is also

Table 2 Comparison of test accuracy and running time of several algorithms

Method(Learning rate strategy)	AP(aircraft)/%	AP(vehicle)/%	mAP%	Time/s
Faster R-CNN(step)	36.28	83.92	58.99	1.5
SSD300 (step)	33.21	77.48	55.35	16
SSD500(step)	35.91	79.33	57.67	7
YOLOv2 416×416(step)	29.34	75.45	52.81	32
YOLOv2 544×544(step)	32.47	76.97	54.72	21
PVANet(plateau)	39.31	84.75	62.53	8
MCAA	41.28	85.66	64.58	10
GAN	45.67	86.21	68.57	9
AAB	49.85	88.39	69.92	11
Proposed (plateau)	54.75	89.37	71.21	8
Proposed (step)	55.29	89.72	72.51	7
Proposed (inv)	56.57	89.83	73.46	7

Table 3 mAP comparison with different methods

Class	MCAA	GAN	AAB	Proposed
airplane	76.9%	79.5%	73.8%	85.9%
oil tank	91.3%	91.8%	92.6%	93.4%
overpass	85.6%	86.7%	91.1%	92.5%
playground	97.5%	97.3%	99.4%	99.6%
mAP	85.9%	87.2%	91.6%	93.8%

Fig. 7 Detection result with proposed method

relatively fast in detection speed. However, under the condition of ensuring the basic realization of real-time detection, the test accuracy is significantly better than other methods. In general, the proposed method in this paper is an effective small-target detection algorithm.

2. Detection effect analysis of using different strategies to train network

The learning rate has a great influence on the model convergence to the local minimum, that is, reaching the highest accuracy. The proposed method firstly follows the "plateau" learning rate strategy in PVANet to train the network. As can be seen from Table 2, the average test accuracy reaches to 71.21%, and compared with the original PVANet algorithm, the detection accuracy improves by 9.53%.

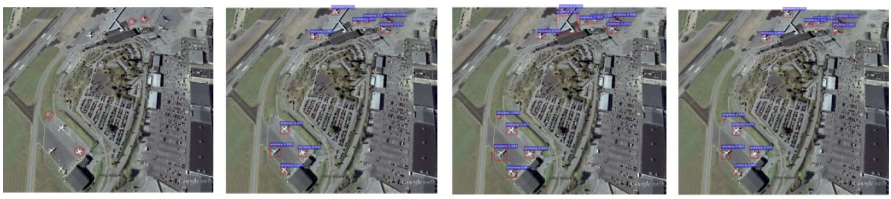
However, in the small target detection problem, the target area occupies a small proportion, the negative sample space is large, and the model convergence rate is slow. In the PVANet model, the convolutional layer and the full connection layer have 94 layers. The network is deep and narrow, which determines that the loss function of the model is prone to oscillation during the training process. The "plateau" learning rate strategy changes the learning rate by monitoring the change value of the loss function, and reduces the learning rate when the change value is less than the set threshold value in a certain period of time. However, when the loss function begins to oscillate and the change value of the oscillation in a period of time is greater than the monitoring threshold, the learning rate will not change and the loss function will not converge. In order to further improve the detection accuracy, this paper adopts the strategy "step" with uniform change of learning rate and the strategy "inv" with dynamic change of learning rate to train the network. When these two learning rate strategies are adopted, even when the loss function falls into an oscillation period, the learning rate will still decrease with the increase of iteration times, and the model can be further converged. It can be seen from Table 2 that the detection accuracy of these two learning rate strategies is improved by 0.45% and 1.14% respectively compared with that of "plateau".

In addition, the "step "learning rate strategy requires to manually set the iteration interval to reduce the learning rate, while the "inv" learning rate strategy makes the learning rate decrease with a small number, eliminating the possible improper setting of the iteration interval to change the learning rate manually. Table 2 shows that using "inv" dynamic learning rate strategy can train the optimal network model, and the detection accuracy is improved by 0.69% compared with using "step" learning rate strategy. The method in this paper uses the "inv" learning rate strategy to achieve an average test accuracy of 72.09% on the constructed small target data set, which improves the detection performance by 10.67% compared with the original PVANet algorithm. It can be seen that the dynamic change of learning rate plays an important role in faster crossing the saddle points of the error surface in the training process and improving the detection accuracy.

We also make comparison experiments on “DOTA” dataset. “DOTA” is an open data set of remote sensing images established by Long [28]. The features of the DOTA data set are: including four classes (4993 airplanes, 1586 oil tanks, 180 overpasses, and 191 playgrounds); containing a total of 2326 remote sensing images obtained from Google Earth and Tianditu. Those images have spatial resolutions between 0.3 and 3.0, manually annotated with BBs [29]. Table 3 shows the results with different methods.

Table 3 indicates that proposed method achieves 9.0%, 6.4%, 12.1% better than other methods in the airplane class, respectively. The airplane class contains more small targets, which means that our method can perform better in detecting small targets. The obvious superiority of proposed method in the airplane class puts its mAP on top of those of the four methods, which is 2.2% higher than that of the AAB method.

We select a raw image from the DOTA data set, which pictures 11 airplanes, including 4 small airplanes marked in red circles as shown in Fig. 8a, and test three methods on it. The GAN method failed to detect the smallest airplane, as shown in



(a) (b) (c) (d)

Fig. 8 Comparison on DOTA dataset with different methods. **a** Raw image. **b** Detection results of GAN. **c** Detection results of ABB. **d** Detection results of proposed method

Fig. 8b; the AAB method detected all the eleven targets but also wrongly identified an unrelated object as an airplane, as shown in Fig. 8c, while our proposed method detected exactly 11 airplanes, as shown in Fig. 8d.

8 Conclusion

In this paper, we propose a new PVANet algorithm for small object detection in real-time. By combining the method of generating high-quality candidate box of small object with RPN network and selecting the appropriate learning rate strategy, the difficulty of locating small object with blocking interference is effectively solved. Experimental results show that the proposed method has better robustness in the detection of small objects, the detection time is greatly improved, and the real-time detection effect is achieved.

References

1. Yin, S., Zhang, Ye., & Karim, S. (2018). Large scale remote sensing image segmentation based on fuzzy region competition and gaussian mixture model. *IEEE Access., 6*, 26069–26080.
2. Yin, S., Zhang, Y., & Karim, S. (2019). Region search based on hybrid convolutional neural network in optical remote sensing images. *International Journal of Distributed Sensor Networks*, *15*(5). https://doi.org/10.1177/1550147719852036.
3. Zhang, Q., Bai, C., Chen, Z., et al. (2019). Deep learning models for diagnosing spleen and stomach diseases in smart Chinese medicine with cloud computing. *Concurrency and Computation: Practice and Experience*. https://doi.org/10.1002/cpe.5252
4. Peng, L., Chen, Z., Yang, L. T., et al. (2018). Deep convolutional computation model for feature learning on big data in internet of things. *IEEE Transactions on Industrial Informatics, 14*(2), 790–798.
5. Ren, S., He, K., Girshick, R., et al. (2015). Object detection networks on convolutional feature maps. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *39*(7), 1476–1481.
6. Pham, P., Nguyen, D., Do, T., et al. (2017). Evaluation of deep models for real-time small object detection. In *International conference on neural information processing* (pp. 516–526). Springer, Cham.
7. Chen, C., Liu, M. Y., Tuzel, O., et al. (2016). R-CNN for small object detection. ACCV 2016. Lecture Notes in Computer Science, vol. 10115. Springer, Cham, pp. 214–230. https://doi.org/10.1007/978-3-319-54193-8_14.
8. Teng, L., Li, H., & Karim, S. (2019). DMCNN: A deep multiscale convolutional neural network model for medical image segmentation. *Journal of Healthcare Engineering*.
9. Yu, J., & Li, H. (2019). Modified immune evolutionary algorithm for IoT big data clustering and feature extraction under cloud computing environment. *Journal of Healthcare Engineering*.

10. Cheng, P., Liu, W., Zhang, Y., et al. (2018). LOCO: Local context based faster R-CNN for small traffic sign detection. In *International conference on multimedia modeling*. Springer, Cham.
11. Zhang, L., Lin, L., Liang, X., et al. (2016). Is Faster R-CNN doing well for pedestrian detection?. In ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9906. Springer, Cham, pp. 443–457. https://doi.org/10.1007/978-3-319-46475-6_28.
12. Chen, Y., Li, W., Sakaridis, C., Dai , D. & Van Gool L. (2018). Domain adaptive faster R-CNN for object detection in the wild. In *2018 IEEE/CVF conference on computer vision and pattern recognition*, Salt Lake City, UT, pp. 3339–3348. https://doi.org/10.1109/CVPR.2018.00352.
13. Zhu, Z., Liang, D., Zhang, S., Huang, X., Li, B., & Hu, S. (2016). Traffic-sign detection and classification in the wild. In *2016 IEEE conference on computer vision and pattern recognition (CVPR)*, Las Vegas, NV, pp. 2110–2118. https://doi.org/10.1109/CVPR.2016.232.
14. Yun, R., Changren, Z., & Shunping, X. (2018). Small object detection in optical remote sensing images via modified faster R-CNN. *Applied Sciences, 8*(5), 813.
15. Rabbi, J., Ray, N., Schubert, M., et al. (2020). Small-object detection in remote sensing images with end-to-end edge-enhanced GAN and object detector network. *Remote Sensing, 12*(9), 1432.
16. Kim K H, Hong S, Roh B, et al. PVANET: Deep but Lightweight Neural Networks for Real-time Object Detection[C]\\Thirtieth Annual Conference on Neural Information Processing Systems (NIPS), 2016. arXiv:1608.08021
17. Deepak, K., Yang, C., & Kyungnam, K. (2014). A neuromorphic system for video object recognition. *Frontiers in Computational Neuroscience*, 8.
18. Pham, P., Nguyen, D., Do, T., et al. (2017). Evaluation of deep models for real-time small object detection. ICONIP 2017: Neural Information Processing (pp. 516–526).
19. Duan, B., Wen, P., & Li, P. (2020). Real-time small object detection method based on improved pvanet [J]. *Application Research of Computers., 37*(2), 279–283. https://doi.org/10.19734/j.issn.1001-3695.2018.06.0577
20. Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.
21. Smith, L. N. (2017). Cyclical learning rates for training neural networks. 2017 IEEE winter conference on applications of computer vision (WACV), 464–472. https://doi.org/10.1109/WACV.2017.58
22. Steno, P., Alsadoon, A., Prasad, P., et al. (2020). A novel enhanced region proposal network and modified loss function: threat object detection in secure screening using deep learning. *The Journal of Supercomputing, 8*, 1–30.
23. Chen, C., Yang, X., Huang, R., et al. (2020). Region proposal network with graph prior and IoU-balance loss for landmark detection in 3D ultrasound. IEEE ISBI, 2020. IEEE.
24. Zhu, J., Zhang, G., Zhou, S., et al. (2021). Relation-aware Siamese region proposal network for visual object tracking. *Multimedia Tools and Applications*, 9.
25. Guo, P., Xie, G., Li, R. (2019). Object detection using multiview CCA-based graph spectral learning. *Journal of Circuits, Systems and Computers*, 4.
26. Zhai, X., Cheng, Z., Wei, Y., et al. (2019). Compressive sensing ghost imaging object detection using generative adversarial networks. *Optical Engineering, 58*(1), 1.
27. Gao, M., Yujie, Du., Yang, Y., et al. (2019). Adaptive anchor box mechanism to improve the accuracy in the object detection system. *Multimedia Tools and Applications, 78*, 27383–27402.
28. Long, Y., Gong, Y., Xiao, Z., & Liu, Q. (2017). Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing, 55*(5), 2486–2498.
29. Dong, R., Xu, D., Zhao, J., Jiao, L., & An, J. (2019). Sig-NMS-based faster R-CNN combining transfer learning for small target detection in VHR optical remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing, 57*(11), 8534–8545. https://doi.org/10.1109/TGRS.2019.2921396