International Journal of Recent Engineering Research and Development (IJRERD)
ISSN: 2455-8761
www.ijrerd.com || Volume 04 – Issue 07 || July 2019 || PP. 86-91

Improved small object detection algorithm in panoramic depth map

Bingxin Bai[1], Zhuangzhuang Mao[2], Hongpeng Zhou[3]

[1]*Henan Polytechnic University, School of Mechanical and Power Engineering, 2001 shiji road, Jiaozuo, China*

[2]*School of Mechanical and Power Engineering, Henan Polytechnic University, 2001 shiji road, Jiaozuo, China*

[3]*School of Mechanical and Power Engineering, Henan Polytechnic University, 2001 shiji road, Jiaozuo, China*

Abstract: In recent years, artificial intelligence has become more and more popular. It has been well applied in some fields. All of this is due to the rapid development of scientific research and the rapid improvement of computer computing speed, especially the training of neural networks has made great progress, in the object detection field, the Fast R-CNN has achieved good results for the detection of medium and large object, but the detection effect of the network for small objects is not very good. The training goal of this paper is the small object in the panoramic depth map. Therefore, it is necessary to improve the existing network. First, the improved Faster R-CNN is used for object detection, and the detection effect is improved. Then, the improved Pvanet is used to detect small objects in the panoramic depth map, and the final detection results have achieved a great improvement. The environment is more adaptable. In the similar research in the future, there is a good reference value.

Keywords: panoramic depth map; small object; improved Fast R-CNN; improved Pvanet

1. Introduce

Since the development of artificial intelligence, the various algorithms used to solve problems have emerged through the joint efforts of scientists all over the world. Some of the artificial intelligence industries have already achieved commercialization. It is believed that everyone who has taken the train knows that the staff will verify our identity before we enter the station. This process requires us to wait for a long time when there are many people. It has caused some troubles for passengers, but now it is possible for passengers to enter the station by “brushing face”. When passengers enter the station, they only need to put their ID cards in the designated place. After the system recognizes the face information successfully. The passengers will pass through, the whole process is fast, convenient and efficient, which enhances the travel experience of passengers. In addition, the object detection[1,2] in the field of artificial intelligence is also applied to road traffic flow monitoring, vehicle illegal capture, etc. These processes have greatly facilitated the traffic workers and promoted the modern development of the society.

In fact, the field of object detection appeared very early. Although the traditional detection methods have made some progress in detection, such methods are based on human extraction features and can only detect the object of fixed features. The whole process is time-consuming and complicated. The generalization ability is poor. In recent years, the rapid development of artificial intelligence, especially in the field of object detection based on deep learning, has made great progress. Through the joint efforts of scientists all over the world, various algorithms for solving problems have empvaerged one after another. Object detection needs to solve the problem of what the object is and where it is. After years of in-depth research, Girshick et al. proposed the

Fast-RCNN based on the predecessors and combined with the positioning idea. It mainly added a pooling layer (ROI pooling layer) behind the convolutional layer. The network structure is faster and more accurate than the previous structure. Then, the network structure of the Faster-RCNN was proposed by Ren Shaoqing et al. It mainly replaced the previous selective search[3] with the candidate region (RPN) method, which speeds up the training. At this point, object detection is quite mature in the detection of the medium and large objects.

The object detection in this paper is a small one in the panoramic depth map. The original Faster-RCNN structure and the improved one were tested and their effect were not very good.Therefore, this paper improved the two network frameworks, and improved the network by modifying the number and size of anchor[4] points in the network, making it more suitable for the size of the object. The accuracy of detection is further improved.

2. Related work

The Pvanet [5] is suitable for small object recognition[6,7]. The quality of the datasets affects the performance of the network training. Therefore, the datasets used for network training is very important. Aiming at the situation of small objects, a dataset for small object training and detection is made based on the KITTI dataset. This paper refers to the format of PASCAL VOC, and has the following improvements in the dataset:

(a) This paper proposes a method to training neural networks with depth pictures. Compared with traditional RGB pictures, the appearance of depth pictures is fuzzy and difficult to identify. However, the depth pictures have points that are smooth and easy to regularize [4]. It also reduces the adverse effects of reflection and nighttime environment in the RGB pictures on the detection. In the object detection field, this method avoids some disadvantages compared with using RGB images.

(b) The datasets used in the training of the model is panoramic depth picture. The picture format is 64×870 , and its size does not conform to the format of PASCAL VOC dataset. If it is used directly for training, it will be wrong during training. This paper proposes a solution, which is to splicing 5 original panoramic depth pictures into one large picture. The dataset of 1500 pictures was finally produced. The size of the spliced picture is 320×870 , which conforms to requirements of the network for datasets. This method solves the problem well.

3. Method of this paper

3.1 Improved Faster-RCNN

The development of object detection based on the Faster-RCNN has made great progress in both detection accuracy and speed, and some achievements have been made in commercial applications. However, this kind of detection also has certain limitations. Its good performance is mainly reflected in the detection of medium and large object, For small object. Because the pixel is low and its outline is blurred, the original Faster-RCNN is not ideal for detecting small objects. Since the anchor_box controls the size of the regression box[8], we need to adjust the size of the regression box when we detect small objects. In the paper, the improved Faster R-CNN[9] is proposed, and the reasonable value and number of anchors are set on the original algorithm framework. The goal of improving the recognition rate of small objects is achieved. In order to set the appropriate size of the anchor boxes, it is necessary to analyze the size distribution of the object and the aspect ratio in the dataset. The histogram of the distribution of size and aspect ratio is shown in Figures 1 and 2.

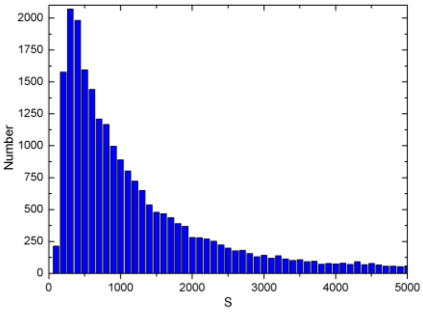


Figure.1 Size distribution column chart

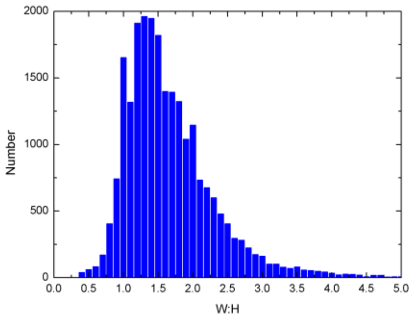


Figure.2 Aspect ratio distribution column chart

This column chart intuitively reflects the distribution of the area of the object object and the distribution of the aspect ratio in the image. It can be seen that the object is mainly distributed in position of the smaller values on the coordinate axis, which is a typical small object detection, Therefore, this paper had undergone many experiments for constantly adjusting the anchor value in the network, and finally adjust it to a suitable size.The network will provide 60 possible candidate windows for each position in the image, namely: base_size=6;ratios=[0.125, 0.25, 0.5, 1, 2, 4]; scales = 16×np. array ([0.75, 1, 1.25, 1.75, 2, 2.25, 3, 3.5, 4.25, 4.75]). The anchor point is an important hyperparameter. The anchor point has a great relationship with the candidate box. Therefore, this paper used 60 anchors in order to make the network detect more small objects.The experimental results show that the final effect is improved. The Region Proposal Network is shown in Figure 3.

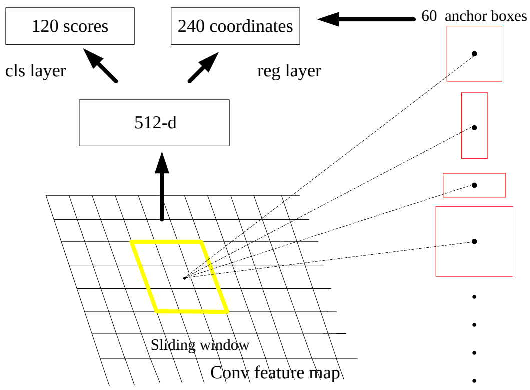


Figure. 3 Region Proposal Network

3.2. Improved Pvanet

In the process of small object detection, because small objects occupy a small proportion of the entire image, there are many problems in the detection process. On the one hand, the complicated background information will have a very adverse effect in the detection process; On the other hand, small objects tend to be vague and contain less useful information, which ultimately leads to unsatisfactory detection results. Although the Pvanet is mainly aimed at the detection of small objects, the small objects in the dataset of this paper have their particularity. In the experiment, the Pvanet was improved similar to the Fast-RCNN. The size distribution of small object vehicle is shown in Figure 1 and Figure 2, so the anchor value of the network is set as: scales= [0.25, 0.5, 0.75, 1, 1.5, 2] and ratios= [8,16, 24, 32]. The detection process of the improved Pvanet [10] is divided into two steps. First, the feature map is obtained through the feature extraction network to obtain, and then the feature map is input to the RPN for obtaining the object candidate frame. Second, the feature map and the object candidate frame are sent to the pooling layer, the fully connected layer and the subsequent layers, which are used to make the object classify and obtain more accurate coordinates of the object candidate frame. The structural schematic diagram of the Pvanet is shown in Figure 4.

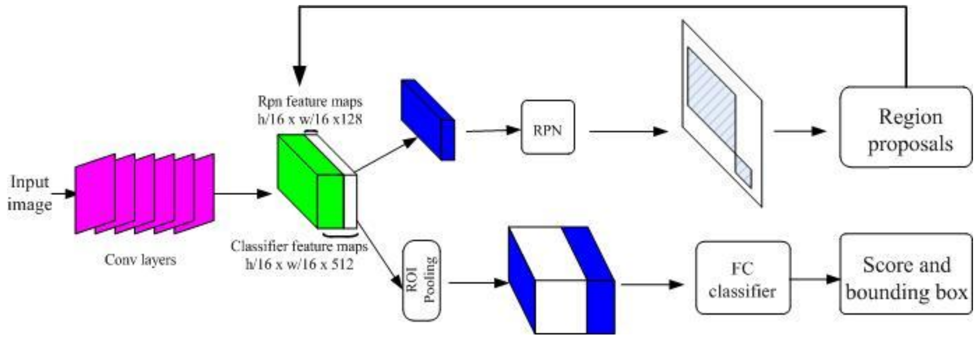


Figure.4 Pvanet Structure

4. Experimental results and analysis

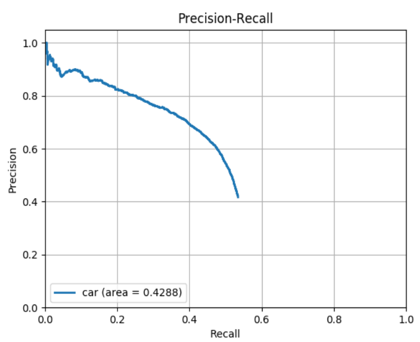
In this paper, we train the neural network on an Ubuntu 18.04 server with an GTX 1070 GPU, using CUDA 10.0 for acceleration. The experiment uses the CAFFE[11] frame to train the network that uses the improved Faster RCNN and improved Pvanet to train separately. In order to evaluate the effectiveness of the trained model, the performance can be measured by setting corresponding indicators. The value of the average precision (AP) represents the performance of the model and its calculation formula is shown in Equation (1). The recall rate[12] indicates how many positive examples in the sample are predicted correctly, and the calculation formula is expressed in Equation (2).

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{1}$$

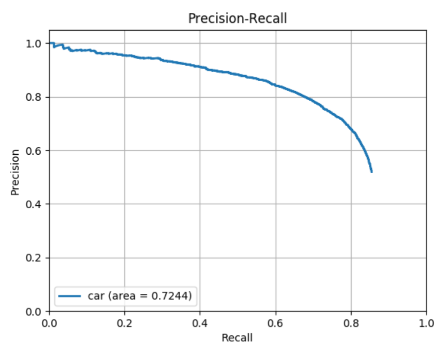
$$P = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN} \tag{2}$$

where TP represents a positive sample predicted to be positive by the model, FN represents a positive sample predicted to be negative by the model, and FP represents a negative sample predicted to be positive by the model.

The experimental results of the two improved network PR curves are shown in Figure 4.



(a) PR curve of improved Faster-RCNN



(b) PR curve of improved Pvanet

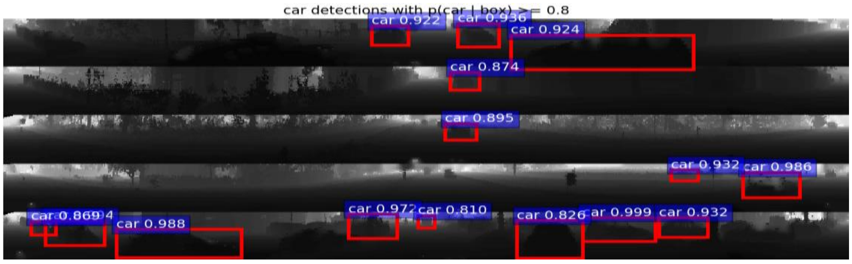
Figure 4. PR curves

Table 1 gives the comparison of training average accuracy, recall rate and speed when several methods are used to train the small object dataset. It can be concluded from the data in the table that the detection effect of the improved network has been greatly improved, especially the improved Pvanet has great advantages in small object detection.

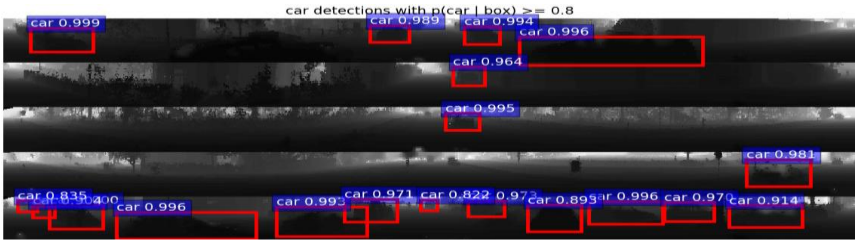
Table 1 Quantitative comparative analysis of various methods

Method	AP	Recall	Speed(s·/per)
Improved Faster-RCNN	42.88%	37.02%	0.061 s·/per
Improved Pvanet	72.44%	57.91%	0.185 s·/per

In order to verify the performance of the improved two networks, we randomly selected a few pictures from the dataset to test, and the final results are shown in Figure 5.



(a) Improved Faster R-CNN



(b) Improved Pvanet

Figure.5 Detection effect of the Faster R-CNN +VGG16 and Pvanet in the testset

5. Conclusion

The detection object in this paper is mainly the object in the depth map. The paper uses the Faster R-CNN with good performance in the object detection field to detect the small object. If the object in the image is too small, the network will not perform its best performance. Based on this, this paper modifies the ratio and size of the anchor of the RPN in the Faster R-CNN, which improves the detection accuracy. Then, due to the advantages of Pvanet for small object detection, this paper also further optimizes it. The improved Pvanet is used to detect small object vehicles, and the final result is improved. In this process, because of the particularity of the size of the panoramic depth map, this paper proposes a splicing method that can be used for Pvanet training, and breaks the application limitations of the Pvanet, making it not only for small object objects in color maps. Moreover, the application of the defective dataset in the panoramic depth map is realized, which makes the Pvanet used more widely, and provides ideas for subsequent academic research. In the experiments, the improved network does not reach the ideal state in detection accuracy. Incorrect marking of images in the dataset may cause errors in the detection results, and the performance of the trained model may not achieve the desired effect due to the limitations of the panoramic depth map itself. So to solve the above problem is the next important problem to be studied.

References

[1] A. Ouaknine, "Review of deep learning algorithms for object detection," Medium, February, 2018.

[2] S. Ren, K. He, R. Girshick, X. Zhang and J. Sun, "Object detection networks on convolutional feature maps," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 7, pp. 1476-1481, 2016.

[3] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," International journal of computer vision, vol. 104, no. 2, pp. 154-171, 2013.

[4] Y. Zhong, J. Wang, J. Peng, and L. Zhang," Anchor boxoptimization for object detection," arXiv preprint arXiv:1812.00469, 2018.

[5] K. H. Kim, S. Hong, B. Roh, Y. Cheon, and M. Park, "Pvanet: Deep but lightweight neural networks for real-time object detection," arXiv preprint arXiv:1608.08021, 2016.

[6] B. Yang, Y. Zhang, J. Cao, and L. Zou, "On road vehicle detection using an improved faster RCNN framework with small-size region up-scaling strategy," in Pacific-Rim Symposium on Image and Video Technology, pp. 241-253, Springer, Cham, November, 2017.

[7] C. Chen, M. Y. Liu, O. Tuzel, and J. Xiao, "R-CNN for small object detection," in Asian conference on computer vision, Springer, Cham. pp. 214-230, November, 2016.

[8] Y. He, X. Zhang, M. Savvides, and K. Kitani, "Softer-nms: Rethinking bounding box regression for accurate object detection," arXiv preprint arXiv:1809.08545, 2018.

[9] C. Zhang, X. Xu, and D. Tu, "Face detection using improved faster rcnn," arXiv preprint arXiv:1802.02142. 2018.

[10] C. Zaharia, D. Dinu, and A. Caliman, "PVANet optimization for person detection," in International Conference on Optimization of Electrical and Electronic Equipment (OPTIM) and Intl Aegean Conference on Electrical Machines and Power Electronics (ACEMP), pp. 959-964, May,2017.

[11] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in Proceedings of the 22nd ACM international conference on Multimedia, pp. 675-678. November, 2014.

[12] Y. H. Byeon, and K. C. Kwak, "A performance comparison of pedestrian detection using Faster RCNN and ACF," in 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), pp. 858-863, July, 2017.