Object Detection Algorithm for Small Objects Based on Residual Branch

Xiaoling Feng

***Abstract*—In recent years, object detection has been widely developed, and small object detection in object detection has received more and more attention. Feature pyramid networks are often used in object detection. Because the feature pyramid network can be detected on feature maps of different scales, more objects can be detected using the feature pyramid network for detection. However, in the pyramid feature network, due to the reduction of channels in feature fusion, the high-level features of the pyramid will lose the detailed information of the object. At the same time, because the background information at the bottom of the pyramid is complicated, it is not conducive to detecting small object ions. In order to better improve the performance of feature pyramid detection for small objects, we propose an object detection method based on residual branch (SORB), which improves the detection accuracy of small objects while maintaining the existing object detection accuracy. Our method improves the network structure of the traditional feature pyramid. We also recalculated the weights of the network to reduce the semantic gap between different features in the feature pyramid. Our method is validated on the VOC2012 dataset, and the experimental results show that our method has good results**

***Index Terms*—Feature Pyramid Network, Residual Branch, Object Detection**

I. Introduction

Existing object detection algorithms can be divided into two categories: two-stage detectors and one-stage detectors. Two-stage detectors such as [1], [14], [15] first generate some regions of interest in the first stage, followed by object classification and bounding box regression. One-stage detectors, such as YOLO [2] and SSD [3], detect objects directly

In recent years, many object detection algorithms based on Feature Pyramid (FPN) have been proposed. Previous object detection networks used only a single extracted feature for prediction. The proposal of feature pyramid enables people to make full use of feature pyramids of multiple scales to extract candidate boxes. FPN can combine low-level feature information and high-level edge information, which greatly increases the probability of objects of different scales being detected. Therefore, small objects in the picture will not be too different in scale from other objects to be ignored by the detector as noise information. Mou et al. [4] proposed a method to build feature pyramid networks with strong semantic feature maps at all scales using top-down paths and horizontal connections. Feature maps at different layers are responsible for objects of different sizes. Yang et al. proposed a dense feature pyramid network (DFPN) [5] to achieve automatic ship detection: each feature map is closely connected and combined through cascades.

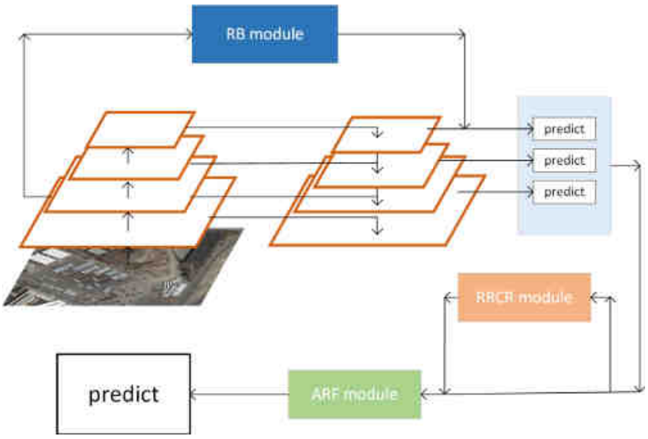

Figure. 1 The figure shows the SORB network architecture.

II. Methods

Existing object detection methods are difficult to deal with the compatibility problem between feature maps of different scales, especially the feature maps of higher layers. The information contained in the feature maps can be well enhanced if a proper fusion method is used for the feature maps of higher layers. In this way, it is convenient to extract the small target feature area and to detect small target objects. So we propose a residual branch-based SORB network to improve the detection accuracy of small objects in object detection. Figure 1 shows the network structure of our method. Our designed residual branch performs multiple operations on tensors in order to better fuse feature maps from higher layers in the feature pyramid.

A. *Network structure*

Our method is described in detail below. The object detection algorithm we designed consists of two parts - RB module and RRCR module. The RB module is used to extract the feature map and send it to the subsequent network for detection, and the RRCR module is used to generate candidate boxes and perform regression. The RB module uses the residual branch to generate a new feature map F5. The residual branch utilizes the three higher layers of the feature pyramid, recalculates the weights and fuses them with the original feature map. Finally, a new feature map is generated to replace the original top-level feature map to form a new pyramid network. The RRCR module consists of an anchor classification branch and an anchor regression branch, which

Manuscript received March 08, 2022

Xiaoling Feng, Computer Science and Technology, Tiangong University, Tianjin, China

are used to calculate the position of the candidate box and perform regression. We then send the candidate boxes and input feature maps into deformable convolution [6] to extract aligned features. Finally, Active Rotation Filter [7] (ARF) is used to extract invariant orientation features and produce final detection results.

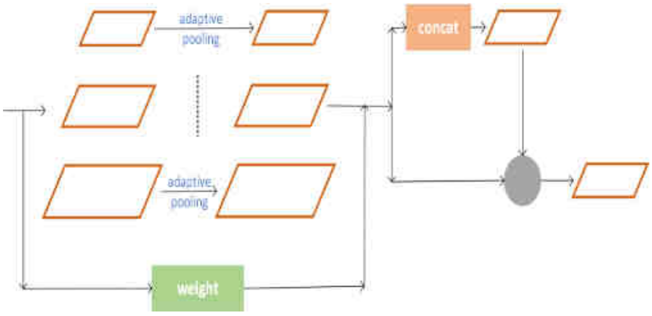

Figure. 2 The diagram shows the detailed structure of the residual branch that we propose.

B. Residual Branch

The feature pyramid reduces the number of channels after multiple convolutions during top-down feature fusion. Figure 2 shows the structure of our designed residual branch. As a result, the feature maps that are closer to the higher layers are more likely to lose detailed information. To this end, we use adaptive pooling of different scales in the upper layers of the feature pyramid to generate feature pyramids of different scales with multiple contextual features. To avoid aliasing effects caused by interpolation, we set four different scales to accommodate these context functions instead of simply summing them. And two scales are specially set for the small target in order to better extract the feature information of the small target. Each contextual feature is then independently passed through a 1×1 convolutional layer, reducing the channel dimension to 256 feature maps. Finally, to build a feature pyramid, we use a 3×3 convolutional layer on each feature map.

III. EXPERIMENT

Our method is validated on the Pascal VOC2012 dataset. The trained model identifies objects of the corresponding category from a given image. The target frame given by the model is calculated, and the experimental results can be obtained.

Figure. 3 Example of an image with acceptable resolution

A. Data Set

The PASCAL VOC Challenge (The PASCAL Visual Object Classes) is a world-class computer vision challenge. The PASCAL VOC 2007 and 2012 datasets are divided into 4 categories: vehicle, household, animal, person, with a total of 20 subcategories (plus 21 background categories). Currently, the VOC2007 and VOC2012 datasets are commonly used for target detection. There are 11,540 images used for classification and detection in the VOC2012 dataset, including 27,450 labeled instances. The target object will be marked with 4 coordinate values of the upper left corner and the lower right corner of the target frame. The evaluation standard of PASCAL is mAP (mean average precision). For PASCAL, each category has such a PR curve. The area enclosed by the PR curve and the x-axis is called average precision. There is one AP, and the average of APs of 20 categories is mAP.

B. Experiment Result

The SORB method is compared with other popular methods GoogleNet [8], VGG-16 [9], FRCN [10] in the VOC2012 dataset. The experimental results are shown in Table 1 below. The mAP in the last row of the table is the average of multiple object detections. From the results, our method significantly outperforms some previous detection methods. With default input size, e.g. 1024×1024, SORB can run at 386ms per image on RTX2080. Single-scale tests can run as fast as 62 ms per image.

From the data in the table, we can see that the accuracy of most target objects is higher than several other methods. Only individual classes of target objects have lower accuracy than several other methods, and the average accuracy of our method is also higher than several other methods. Among them, the accuracy of birds is improved by 5.4%, the accuracy of bus type is improved by 1.5%, the accuracy of dogs is improved by 1.4%, and the accuracy of plants is improved by 2%. The accuracy of other kinds of targets is not improved much, less than 1%. It can also be seen that our method improves the detection accuracy of some small objects, and the average accuracy of all types of objects also improves.

Table I VOC 2012 Test Detection Average Precision (%)

precision	**Method**			
	GoogleNet	**VGG16**	**FRCN**	**Ours**
aero	74.1	73.6	67.2	76.8
bike	68.9	60.7	71.8	72.0
bird	59.9	55.3	51.2	65.3
boat	36.7	35.6	38.7	35.5
bottle	35.4	33.5	20.8	35.8
bus	71.6	72.5	65.8	73.5
car	62.4	60.3	67.7	68.5
cat	81.8	80.1	71.0	79.4
chair	36.1	34.6	28.2	36.3
cow	58.5	57.7	61.2	65.2
table	40.0	42.5	61.6	55.7
dog	77.5	76.2	62.6	78.9
horse	67.8	68.2	72.0	73.4
motor	74.8	75.1	66.0	75.2
person	61.0	60.0	54.2	62.6
plant	30.8	30.0	21.8	32.8
sheep	61.2	61.5	52.0	60.6
sofa	58.0	56.3	53.8	58.3

train	67.0	64.8	66.4	68.1
tv	64.1	63.5	53.9	64.3
mAP	59.4	58.6	55.4	61.9

CONCLUSIONS

In this paper, a new method for object detection based on residual branch is proposed. Our method uses residual branches to improve the pyramid network structure and reduce the feature loss that occurs during feature fusion. Our method uses focal loss to better rebalance different scales of bounding boxes. Using multi-scale testing in experiments can significantly improve detection performance. Our SORB is trained using ResNet-50-FPN and ResNet-101-FPN as backbone networks, and both the resulting models achieve good performance on the VOC2012 dataset. I hope our method can play some role in the field of object detection or data statistics.

REFERENCES

[1] Shah M, Kapdi R. Object detection using deep neural networks[C]//2017 International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2017: 787-790.

[2] Redmon J, Divvala S, Girshick R, and Farhadi A, You only look once: Unified, real-time object detection, in CVPR, pp. 779–788, (2016).

[3] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, and Berg A C , “SSD: Single shot multibox detector,” in ECCV, pp. 21–37, (2016).

[4] Mou L and Zhu X X, Vehicle instance segmentation from aerial image and video using a multitask learning residual fully convolutional network, IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 11, pp. 6699–6711, (2018).

[5] Yang X, Sun H, Fu K, Yang J, Sun X, Yan M, and Guo Z, Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks, Remote Sensing, vol. 10, no. 1, (2018).

[6] Law H and Deng J, Cornernet: Detecting objects as paired keypoints, in ECCV, (2018).

[7] Zhou Y, Ye Q, Qiu Q, and Jiao J, Oriented response networks, in CVPR, pp. 4961–4970, (2017).

[8] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. 2015.

[9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

[10] Ren S, He K, Girshick R, and Sun J, Faster r-cnn: Towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, (2017).

[11] Lin T-Y, Dollar P, Girshick R, He K, Hariharan B, and Belongie S, Feature pyramid networks for object detection, vol. 2017-January, (Honolulu, HI, United states), pp. 936–944, (2017).

[12] Lin T-Y, Goyal P, Girshick R, He K, and Dollar P, Focal loss for dense object detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 2, pp. 318–327, (2020).

[13] Yang X, Sun H, Fu K, Yang J, Sun X, Yan M, and Guo Z, Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks, Remote Sensing, vol. 10, no. 1, (2018).

[14] Han J, Ding J, Li J, and Xia G-S, Align deep features for oriented object detection, IEEE Transactions on Geoscience and Remote Sensing, (2021).

[15] Yang X, Liu Q, Yan J, and Li A, R3det: Refined single-stage detector with feature refinement for rotating object, CoRR, vol. abs/1908.05612, (2019).

Xiaoling Feng **Feng Xiaoling is from Tianjin University of Technology, majoring in computer science and technology. Her main research direction is object detection in computer vision. She has published two EI conference papers.**