Research on the Algorithm of Detection for Small and Weak Target Based on the Mechanism of High-Resolution Amplification

Xianyi Cheng\textsuperscript{1,2}, Lili Sun\textsuperscript{1} and Yachong Tian\textsuperscript{1}

\textsuperscript{1}Silicon Lake College, Kunshan, Jiangsu 215300, China
\textsuperscript{2}Nantong University, Nantong, Jiangsu 226019, China

Abstract. Aiming at the problem that the existing target detection framework based on big data and deep learning has poor recognition effect for small and weak targets with low resolution in complex scenes, in this paper, we improve the small targets detection accuracy and speed, based on the FCOS algorithm for target detection and the mechanism of high resolution amplification. The basic principle is to increase amplification precision gain regression network (R-NET) and amplification region selection algorithm, on the basis of enhanced learning. The former learns the correlation between coarse and fine detection, and the latter calculates the information gain of the amplified region. The results show that this method can detect small and weak targets in complex environment.

Keywords. Target detection; deep learning; FCOS; reinforce learning; weak and small target detection; high resolution amplification.

1. Introduction

In recent years, computer vision technology based on deep learning has developed independently, both in civilian and military aspects, the target detection in video images plays an important role. In the civil field, high precision target detection is helpful for traffic management and the urban planning; in the military field, high-precision target detection is helpful to accurately lock the invasion and harm of hostile targets and maintain national security [1]. However, the low resolution of the video image and the small size of the target are the key and difficult points in current target detection [2]. The context in which small target sizes are defined varies greatly. In order to ensure the consistency of later content description, this paper refers to the definition of small target in COCO data set, and defines a target whose total pixel number is less than 32\times32 as a small target [3].

2. Related Research

2.1. Traditional Methods of Target Detection

Due to small target size, low resolution and fuzzy edge information, the research of small target detection has great challenges. In the traditional method, the target detection task in the broad sense is mainly solved, and the target detection process is shown in figure 1 [4].

![Figure 1. Flow chart of traditional target detection.](image-url)
The most classic screening strategy is the non-maximum inhibition method proposed by Neubeck and Van Gool et al. [5].

2.2. Research on Dim Target Detection Based on Deep Learning
Deep learning has almost become a standard part of computer vision research for three reasons [6]: figure 2 shows the development of target detection.

![Figure 2. Development history of target detection.](image)

2.3. FCOS Algorithm
FCOS algorithm [7] is a full-convolution single-stage target detection method different from popular target detection methods (Retina Net, SSD, Faster RCNN, etc.), which does not need to define an anchor in advance. This feature makes FCOS greatly reduce the memory required for training, so this method can reach or even exceed the Anchor-based detection method.

As shown in figure 3 is FCOS network’s overall structure, the structure and most of the detection network have the same full convolution. First input image through feature extraction network to extract the C3, C4 and C5 layer, then use obtained five FPN layered structure of layer, finally making forecast in each layer [8].

The biggest innovation of FCOS lies in the regression strategy. Compared with other anchor-based detection methods, FCOS makes full use of the foreground sample to train the network, instead of using the anchor whose truth box exceeds the IOU threshold as the positive sample. This is also one of the reasons why FCOS method is better than other detection algorithms.

![Figure 3. Overall structure of FCOS.](image)

The biggest innovation of FCOS lies in the regression strategy. Compared with other anchor-based detection methods, FCOS makes full use of the foreground sample to train the network, instead of using the anchor whose truth box exceeds the IOU threshold as the positive sample. This is also one of the reasons why FCOS method is better than other detection algorithms.
Aiming at the problem that the existing target detection framework has poor recognition effect on small targets with low resolution in high resolution and complex large scenes, this paper, based on FCOS target detection algorithm, reduces the missed rate of small targets through high-resolution amplification mechanism, and improves the detection accuracy and speed of small targets on the whole. This method consists of a magnification accuracy gain regression network (R-NET) and a magnification region dynamic selection algorithm.

3. The Mechanism of High Resolution Amplification

The sub-sampling version of the image is used to perform rough detection, and then the regions of low resolution small targets are selected for amplification in order to ensure the recognition accuracy of low resolution small targets. Reinforcement learning method was used to model the amplification region from two aspects of detection accuracy and computational cost, and a series of regions were dynamically selected to be amplified to high resolution for analysis. The framework is shown in figure 4.

The algorithm adopts a detection strategy from coarse to fine. The coarse detector is applied in P3 and P4 under the low resolution of figure 4, and the output results of the detector are used to guide the in-depth search of high-resolution targets. Although the crude detector is not as accurate as the fine detector, it will identify areas of the image that require further analysis, thus detecting dim targets only in the promising areas. The algorithm mainly consists of two mechanisms [9]:

- Learn the statistical relationship mechanism between the coarse detector and the fine detector to predict which regions need amplification given the output of the coarse detector. A sequence mechanism for selecting a high-resolution analysis region in the case of a rough detector output but requiring analysis by a fine detector.
- The selection of weak target region is described as a Markov decision process. At each step, the algorithm looks at the current state and then selects the action with the greatest long-term cost-perceived return. In the deep reinforcement learning mechanism, the method of learning while acquiring samples is adopted. After obtaining the samples, the current learning model is updated and the next action is guided. After the next action is obtained, the learning model is updated and repeated iteratively until the learning model reaches convergence.

In order to determine whether the high resolution detection improves the overall detection results, a matching layer CR is introduced to correlate the detection results generated by the two detectors. In this layer, if object I in the subsampled image and object j in the high-resolution image are found to have a large enough intersection IOU (LDI, HDJ) (IOU >0.5), I and j are defined as corresponding to each other, and a set of corresponding relations are generated [10].

AG Map is generated according to the learning accuracy gain of each target. Each pixel in the candidate border is assumed to contribute equally to its accuracy gain. Therefore, AG map is defined as:
\[ AG(x, y) = \begin{cases} \alpha \Phi(W, \lfloor \cdot \rfloor_k) / b_k & \text{if } (x, y) \in \text{LDK}, \\ 0 & \text{otherwise} \end{cases} \]

where \((x, y)\) is point in the bounding box \(\text{LDK}\), \(b_k\) denotes the number of pixels contained in \(\text{LDK}\), \(\alpha\) is a constant, \(W\) represents the estimated parameter of the CR layer. The \(AG\) map represents status and naturally contains information about rough inspection quality. After zooming in and detecting the region, all values in the region are set to 0 to prevent future zooming in the same region again.

The \(AG\) map was divided into equal rectangular areas according to the \(8\times 8\) grid, and the sum of pixel values in each rectangle was counted. The total pixel value of RTGI in the rectangular region in the \(AG\) Map after mesh partition is:

\[ \text{Sum}_{p_{x_j}} = \sum_{j \in \text{rect}_k} p_{x_j} \]

where \(p_{x_j}\) represents the pixel value of the \(j\)th pixel point in the RTGI region. The larger \(\text{SUMP}_{x_i}\) is, the greater the amplification payoff of RTGI representing the rectangular region is, and the region with high amplification payoff is taken as the center to conform to the understanding of correlation.

4. Experiment and Analysis

4.1. Experimental Data Set

In this paper, the FCOS was improved, and this improvement has a better effect. Experiments are conducted on MSCOCO data set. Table 1 lists the proportion of all kinds of target.

| The target category | Statistical number | Pixel value statistics |
|---------------------|--------------------|-----------------------|
| Small target        | 41.43%             | 15.3%                 |
| Middle target       | 34.32%             | 34.2%                 |
| Big target          | 24.24%             | 50.5%                 |

Table 1 shows that although the number of small targets on the MSCOCO dataset is large, however, in terms of the number of pixels occupying the image, it is far less than the large and medium target. This is similar to our life, the number of small targets is large but the pixel occupied is very small, which is also one of the reasons for choosing this data set.

Table 2 shows the top 5 results of the dataset challenge, with APS representing the average accuracy of the small target and APL representing the average accuracy of the large target. It can be seen from table 2 that the detection effect of large and medium-sized targets is good, and good results can be obtained when applied in real life. However, the detection effect of small targets in this data set is not ideal, and the highest is only 0.283, which is far from being applied in production activities.

| Methods    | AP  | AP_{50} | AP_{75} | AP_{S} | AP_{M} | AP_{L} |
|------------|-----|---------|---------|--------|--------|--------|
| DANet      | 0.457| 0.674   | 0.510   | 0.283  | 0.484  | 0.587  |
| BUPT-Priv  | 0.435| 0.662   | 0.474   | 0.253  | 0.477  | 0.560  |
| DL61       | 0.421| 0.632   | 0.467   | 0.245  | 0.458  | 0.544  |
| DeNet      | 0.421| 0.612   | 0.452   | 0.223  | 0.460  | 0.580  |
| IL         | 0.416| 0.629   | 0.456   | 0.231  | 0.449  | 0.546  |
4.2. Evaluation Standard
Target detection evaluation criteria generally use positive and negative samples of the predicted results. The definition is obtained by calculating IOU. When the IOU of the sample and the truth value of each prediction result is greater than the given threshold, the sample is defined as a positive sample, otherwise, it is defined as a negative sample. Usually, only one IOU threshold is defined as 0.5, because the change of this threshold will lead to the change of the final MAP. This paper adopts a more rigorous approach, using multiple IOU threshold values, namely the IOU = 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95 the 10 values to calculate mAPi, then the average of the 10 mAPi, get the final mAP, formula is as follows [11]:

\[ \text{mAP} = \frac{\sum_{i=1}^{10} \text{mAP}_i}{10} \]

4.3. Experimental Comparison
Table 3 shows the comparison of experimental data between FCOS and the improved FCOS.

| Methods             | mAP  | mAP_{50} | mAP_{75} | mAP_{S} | mAP_{M} | mAP_{L} |
|---------------------|------|----------|----------|---------|---------|---------|
| FCOS                | 0.356 | 0.557    | 0.387    | 0.197   | 0.407   | 0.480   |
| DANet               | 0.457 | 0.674    | 0.510    | 0.283   | 0.484   | 0.587   |
| Method in this paper| 0.451 | 0.672    | 0.501    | 0.324   | 0.491   | 0.580   |

The method in this paper adopts low-resolution detection for large targets, which has the phenomenon of missing detection. Therefore, the method in this paper is better than FCOS in both the overall MAP and each separately calculated standard. So, the method in this paper is the best for the experimental data of small and medium-sized targets.

5. Conclusions
The research on small and weak targets has the following significances:

- If use this research to the radar equipment of air defense missile system, will improve the system performance of radar in the most convenient and feasible way, and greatly reduce the radar equipment and the whole air defense missile system development costs.
- If this research was applied to the traffic surveillance video, the greatest possible to detect illegal vehicles, it will create more safe and civilized traffic environment, help auxiliary transportation management and urban planning; If the research results are applied to on-board cameras, it will realize more effective detection of pedestrians, vehicles and road conditions in front of the vehicle, and effectively assist drivers to make corresponding measures to deal with emergencies.
- If the research results are applied to medical diagnosis, the small target detection technology can be used to quickly detect the diseased cells in the computed tomography images, which can help doctors diagnose the condition and effectively reduce the misdiagnosis rate.

Acknowledgment
This research was financially supported by the National Science Foundation (61340037, 61171132).

References
[1] Li J 2020 Research on Small Target Detection Algorithm Based on Deep Learning (Shenyang University of Technology) Master Dissertation.
[2] Bu F 2019 Research on Small Target Detection and Segmentation Algorithm Based on Convolutional Neural Network (Xi’an: Xidian University). (in Chinese)
[3] Lin T Y, Maire M, Belongie S, et al. 2014 Microsoft COCO: Common objects in context ECCV 2014 pp 740-755.

[4] Guo C Z 2019 Deep DPM Algorithm and Its Application in Target Detection (Guangdong: Guangdong University of Technology).

[5] Neubeck A and Van Gool L 2016 Efficient non-maximum suppression 18th International Conference on Pattern Recognition (ICPR'06) pp 850-855.

[6] Wang H L, Qi X L and Wu G S 2018 Research progress of target detection based on deep convolutional neural network Computer Science 45 (9) 11-19.

[7] Bodla N, Singh B, Chellappa R, et al. 2017 Soft-NMS: Improving object detection with one line of code Proceedings of the IEEE International Conference on Computer Vision pp 5561-5569.

[8] Huang J, Shi Y and Gao Y 2019 Multiscale fast RCNN detection algorithm for small targets Computer Research and Development 56 (02) 319-327. (in Chinese)

[9] Ren S, He K, Girshick R, et al. 2017 Faster R-CNN: Towards real-time object detection with region proposal networks IEEE Trans. on Pattern Analysis & Machine Intelligence 39 (6) 1137-1149.

[10] Z Tian, C Shen, H Chen, et al. 2019 FCOS: Fully convolutional one-stage object detection IEEE International Conference on Computer Vision (ICCV) (Seoul, South Korea) pp 9626-9635.

[11] Zhou J and Chen J 2019 Research on identification and detection of OCS bird’s nest based on DSSD Journal of East China Jiaotong University 36 (06) 70-78. (in Chinese)