SMALL OBJECT DETECTION METHOD

Tang Yi, Gourinovitch.A.
KDepartment of Information Technologies in Automated Systems,
Belarussian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus
E-mail: tangyijcb@163.com, gurinovich@bsuir.by

In computer vision, significant advances have been made on object detection with the rapid development of deep convolutional neural networks (CNN). When it comes to small object, the accuracy of deep learning methods is low. This paper mainly introduces the difficulties in small object detection and proposes two methods to enhance feature representation of small objects.

I. Introduction

Small object detection is a fundamental computer technology related to image understanding and computer vision that deals with detecting instances of small objects of a certain class in digital images and videos.

There are mainly two definitions of small objects.One refers to objects with smaller physical sizes in the real world. Another definition of small objects is mentioned in MS-COCO metric evaluation. Objects occupying areas less than and equal to 32×32 pixels come under "small objects" category.

Small object detection has been widely used in academia and real world applications, such as robot vision, autonomous driving, intelligent transportation, drone scene analysis, military reconnaissance and surveillance[1].

II. Difficulties and Challenges

Small object detection has long been a difficult problem in object detection. It aims to accurately detect small objects with few visible features in the image. Relative to regular-sized objects, small objects often lack sufficient appearance information, making it difficult to distinguish them from background or similar objects[2].

There are three difficulties in small object detection. First, small objects lack appearance information needed to distinguish them from background or similar categories. Then the locations of small objects have much more possibilities. That is to say, the precision requirement for accurate localization is higher. Furthermore, the experiences and knowledge of small object detection are very limited because the majority of prior efforts are tuned for the large object detection problem. Besides, the annotations in the existing datasets are not very friendly for small object detection. For example: a large part of the small target annotations in the COCO dataset are hard case annotations. They are not only small, but have varying degrees of occlusion, blur, and incompleteness

Fig.1 below shows an example of the real life small object detection.

Figure 1 – small object detection in real life

III. Advantages and Disadvantages of existing method and model

Currently in many object detection tasks, in order to detect small objects, the usual method is to divide the image into smaller grids, but this will increase a lot of computation.

Handling feature scale issues is of crucial importance for small object detection. Feature pyramid network(FPN) is one of the main paradiams addressing multi-scale feature learning problem.

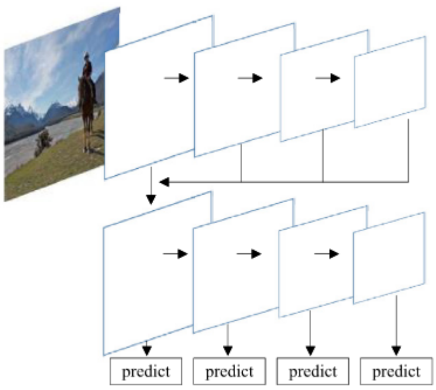

Figure 2 – Feature pyramid network

Above in Fig.2 shows a typical structure fo Feature pyramid network(FPN).The improvement in accuracy is very significant, especially for small targets. FPN can be embedded in any network structure and task as a freely pluggable module, and is widely used in the industry[3].

A large amount of calculation requires a lot of memory: a huge network structure similar to UNET is inserted into the target detection network, which inevitably reduces the speed of the network.

IV. Method improvement and description

According to the difficulty in small object detection mentioned earlier. This paper proposes to optimize small object detection from two aspects. The first is to optimize the loss function. Take the YOLO loss function as an example: in the process of training and predicting the data set, the part of the loss function (here without λ_{coord}) about the bounding box coordinates is as follows[4]:

$$\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbf{1}_{ij}^{obj}[(x_i-\hat{x_i})^2+(y_i-\hat{y_i})^2] \tag{1}$$

$$\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbf{1}_{ij}^{obj}[(\sqrt{w_i}-\sqrt{\hat{w_i}})^2+(\sqrt{h_i}-\sqrt{\hat{h_i}})^2] \tag{2}$$

In the YOLO1-3 series, when performing bounding box regression, the MSE (Mean Square Error) loss function is set directly according to the center point coordinates and width and height information of the predicted box and the real box. Here, the difference between the width and height of large objects is larger than the difference value detected by small objects, which will cause the loss function to be more inclined to correct the coordinate position of large objects rather than the position of small objects.

To directly estimate the coordinate values of each point of the BBox(Bounding Box) is to treat these points as independent variables, but in fact does not consider the integrity of the object itself.Therefore, it is proposed to use IOU (Intersection over union) as part of the loss function.

$$L_{IOU} = 1 - IOU(A, B) \tag{3}$$

Below in Fig.3 shows how to calculate the IOU value of ground truth box and predicted box.

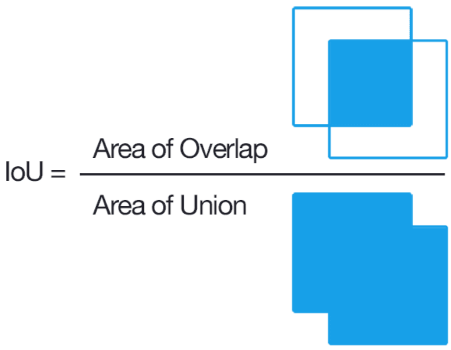

Figure 3 – Intersection over union

In addition to increasing the loss weight for small object detection by optimizing the loss function, the GAN network can also be used to enhance the data[5].

Generative adversarial learning methods aim to achieve the same detection performance as larger-sized objects by mapping the features of low-resolution small objects into features equivalent to those of high-resolution objects. Although the FPN methods mentioned above can effectively improve the performance of small target detection, the performance gains brought by these methods are often limited by the computational cost. Therefore, an effective method is proposed to improve the resolution of small targets by combining Generative adversarial network (GAN), reduce the feature difference between small targets and large/medium-scale targets, and enhance the feature expression of small targets. In turn, the performance of small target detection is improved.

V. Conclusion

In the target detection under computer vision, compared with the detection of large objects, small objects are difficult to detect because of their small physical area and contain fewer pixels.

This paper proposes two methods. The first method is to modify the loss function to enhance the loss weight of small object detection; the other method is to use a GAN network to map the resolution of small object detection. Enter high-resolution objects. The ultimate goal of both methods is to enhance the feature representation of small objects, thereby improving detection accuracy and performance.

VI. References

1. Zou Z, Shi Z, Guo Y, et al. Object detection in 20 years: A survey[J]. arXiv preprint arXiv:1905.05055, 2019.
2. Tong K, Wu Y, Zhou F. Recent advances in small object detection based on deep learning: A review[J]. Image and Vision Computing, 2020, 97: 103910.
3. Gong Y, Yu X, Ding Y, et al. Effective fusion factor in FPN for tiny object detection[C]//Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2021: 1160-1168.
4. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
5. Cubuk E D, Zoph B, Shlens J, et al. Randaugment: Practical automated data augmentation with a reduced search space[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 702-703.