ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0121

Generative Adversarial Network-based Small Object Detection Method

Li Bing*, Yanrong Cui

School of Computer Science, Yangtze University, Jingzhou 434000, China

*479277813@qq.com

Abstract

In computer vision, object detection has always been an important hot topic, and it has a wide range of applications in intelligent video surveillance, robot navigation, and aerospace. Currently, although object detection techniques show good results on datasets related to large or medium-sized objects, the performance on small objects is far from satisfactory, because small objects lack sufficient detailed appearance information. Aiming at the detection of such small targets, this paper proposes a method that uses the advantages of generative adversarial networks in generating super-resolution images to upsample the blurred images of small targets into fine-scale images and restore relevant details to achieve more accuracy.

Keywords

Small Object Detection; Generative Adversarial Network; Super-resolution Reconstruction.

1. Introduction

Object detection is a basic task in computer vision, and it is a key step to realize many practical application scenarios, including obstacle detection, unmanned driving and other fields. In practical applications, the target of the detection area is usually too small due to the inconspicuous target features or the lack of partial feature information, which leads to classification errors. Currently, the detection of small objects is still a challenge for object detection tasks.

At present, the main method in target detection is to use convolutional neural network to extract image features, and to classify and locate the target by establishing a detection model. There are two main categories: candidate region-based and regression-based. Target detection based on candidate regions is also called a two-stage method, such as Faster-RCNN, FPN, etc., and regression-based target detection has only one stage, directly regressing the predicted target objects, such as YOLO series and SSD series algorithms.

This paper proposes a small target detection[1] method based on the idea of generative confrontation, which improves small target detection by reducing the representation difference between small targets and large targets. Traditional small target detection methods cannot reconstruct low-resolution images into high-definition images. Through the idea of generating adversarial network, the representation of small targets can be transformed into a super-resolution representation of the original target similar to the large target, so as to achieve better detection results. The detector in this paper adopts the detection network of the Faster-RCNN model, and proposes a generative adversarial network model that can be added as a component to any two-stage target detection model, so as to realize the target detection of small targets.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0121

2. Related Work

2.1 Definition of Small Goals and Technical Difficulties

At present, there is no clear definition of the concept of small goals in the world. The MS COCO dataset proposed by Microsoft defines the absolute scale of small targets. When the area of the target area is less than 32*32pixel values, it is considered as a small target; the other is the definition of relative scale, that is, the length of the target size. When the width accounts for 0.1 of the original image size, it is a small target. Therefore, for an image with a resolution of 416*416, from the absolute scale, a bird with a pixel value of 20*30 in the image is counted as a small target, and from a relative scale, a car with a pixel value of 100*100 in the image is counted as a small target. It can also be counted as a small goal.

(1) The many factors or main technical difficulties that currently cause the poor performance of small target detection are as follows:

(2) Lack of support for large-scale small target datasets. At present, the public datasets used for image classification, detection, segmentation and other tasks, such as ImageNet, PASCAL VOC series, etc., are mostly for the detection of objects of common scale. The MS COCO dataset contains some small-scale targets, but on average each image contains a large number of instance targets, resulting in uneven distribution of small targets, which cannot be used as a good dataset to support.

(3) The interference caused by the complex environment to the detection of small targets. Since the detection of small objects mostly depends on the specific scene, the information of the small objects will be covered by the noise of other larger objects, or integrated with the background, which is also one of the factors that make the detection of small objects difficult.

(4) Small targets have low resolution and a small proportion of pixels. For small targets, their inherent low resolution, only occupying a few or dozens of pixel values and other characteristics make the effective information that can be extracted during target detection very limited, which is the fundamental reason for the poor detection effect of small targets. Therefore, in practical applications, how to accurately detect cigarette butts, mobile phones, small-scale faces, etc. is extremely challenging for small target detection technology.

(5) There is a bottleneck in the feature extraction of small objects by convolutional neural networks [2]. For the current general target detection model, in order to increase the receptive field, it will go through several downsampling operations, and at the same time, the features will be continuously reduced in dimension and the feature map will be reduced. However, due to the fuzzy edge information and little semantic information of small objects, the information loss of small objects after convolution is serious or even cannot be transmitted to the object detector.

2.2 Generative Adversarial Network GAN

Generative adversarial networks (GAN) [3] is a generative model proposed by ian Goodfellow in 2014. The structure of GAN is inspired by the two-player zero-sum game in game theory (that is, the sum of the interests of two people is zero, and the gain of one party is the loss of the other party), and the system consists of a generator and a discriminator. The optimization process of GAN is a Minimax game problem, and the optimization goal is to achieve Nash equilibrium, so that the generator can estimate the distribution of data samples as much as possible. Its network model is shown in Figure 1.

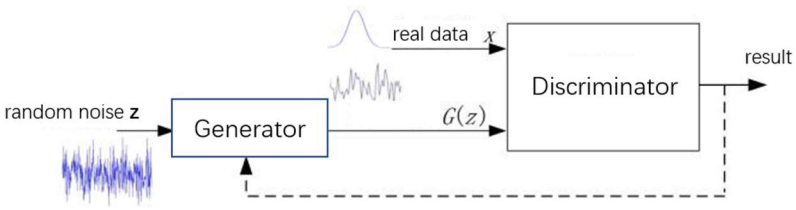

Figure 1. the structure of the GAN network model

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0121

The generator G receives random variable z and generates fake sample data G(z). The purpose of the generator is to make the generated samples as close to the real samples as possible. The input of the discriminator D consists of two parts, namely the real data x and the data G(x) generated by the generator. The output is usually a probability value, indicating the probability that D determines that the input is a real distribution. If the input comes from real data, then output 1, otherwise output 0. At the same time, the output of the discriminator will be fed back to G to guide the training of G. Ideally, D cannot determine whether the input data comes from real data x or generated data G(z), that is, the output probability value of D is 1/2 each time (equivalent to random guessing), and the model is optimal at this time. In practical applications, generative networks and discriminative networks are usually implemented with deep neural networks. The generator and discriminator can be seen as two players in the game. In the process of model training, the generator and the discriminator will update their own parameters to minimize the loss. Through continuous iterative optimization, a Nash equilibrium state is finally reached, and the model is optimal at this time. The objective function of GAN is defined as:

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))].$$

It can be seen from the objective function that the goal of the generator and the discriminator is to minimize and maximize the value of the objective function, resulting in a confrontational process.

At present, GAN has been widely used in images, and can already generate objects such as numbers and faces. It can also form various realistic indoor and outdoor scenes, restore original images from segmented images, color black and white images, restore object images from object contours, and generate high-resolution images from low-resolution images.

3. Implementation Ideas

3.1 Generator

The generator is an important part of the GAN model. In this paper, in order to solve the problem of small object detection, the generator is defined as a super-resolution network, which can upsample small blurred images to fine images, and restore detailed information for accurate detection. In view of the insufficient feature detail extraction at a single scale, the generator used is a multi-scale recurrent network model, which uses multi-scale convolution to extract detail information, and then uses recursive structure to reduce the loss of texture details. Finally, sub-pixel convolution is used for reconstruction to generate High-resolution target image.

3.2 Discriminator

The discriminator is another important part of the GAN model, where the discriminator inputs the generated high-resolution target image and the real high-resolution image into the discriminator at the same time, which can detect whether the generated high-resolution target image can be as real images.

3.3 Detector

The baseline detector can be any type of detector used to crop the target object and background from the input image to train the generator and discriminator.

The final target category detector is proposed to adopt the detection structure of Faster RCNN, which consists of two layers of shared full connections, a classification branch composed of a layer of full connections, and a regression branch composed of a layer of full connections. In this paper, the high-resolution target image output by the generative adversarial network model is used as the input of the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202206_8(6).0121

detector, and the output of the detector is the class probability of the target. That is, the detection of the target is completed.

3.4 Improved GAN - PGGAN

For the generation of super-resolution images, PGGAN[4] has achieved good results in this regard. PGGAN (Progressive Growing of GANs) is a progressively enhanced GAN. PGGAN introduces a new training method, from 4x4 generation Starting with images, up to 1024x1024 faces, the generator and discriminator are progressively grown, starting with low-resolution images and gradually adding new layers to complicate the network model to learn better detailed features.

This approach both speeds up and stabilizes training, and produces more detailed super-resolution images. Applying this model to predict better results in the super-resolution generation of small objects needs further research.

4. Conclusion

The idea of this paper is mainly derived from learning the game ideas involved in generative adversarial network models. It can be seen from the description in the text that the training process of the Generative Adversarial Network GAN is very similar to the minimax algorithm. Through the research on the generative confrontation network GAN, I know its good effect in generating super-resolution images, then apply this idea to related projects of target detection, such as the transformation of small target detection into a target detection mentioned in this article. A super-resolution representation similar to a large target can also be an idea to solve the detection of small targets, but the specific implementation needs to be further discussed and studied.

References

[1] Bai Y, Zhang Y, Ding M, et al. Sod-mtgan: Small object detection via multi-task generative adversarial network[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 206- 221.

[2] Liu Xiaonan, Wang Zhengping, He Yuntao, et al. A review of small target detection based on deep learning [J]. Tactical Missile Technology, 2019 (1): 100-107.

[3] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.

[4] Karras T, Aila T, Laine S, et al. Progressive growing of gans for improved quality, stability, and variation[J]. arXiv preprint arXiv:1710.10196, 2017.