# Study on Tiny Object Detection

Jiahao Song

Central South University, Changsha 410083, China

**Abstract.** Object detection has been one of the most challenging tasks in computer vision and a hot research topic in the world. With the rapid development of in-depth learning technology, researchers have obtained abundant research results in the field of object detection. However, most of the current mainstream object detection methods are based on the modeling of normal scale objects, and the performance of these methods is seriously degraded when they are directly applied to the detection of tiny objects, because the real scene is changing and unknown, and generally there are problems such as object occlusion, close connection and different scales. In this paper, the existing detection methods of tiny objects are summarized.

**Keywords:** Object detection, computer, deep learning.

## 1. Introduction

In recent years, tiny object detection has received wide application and had broad development prospect in many fields of real life, such as emergency rescue, intelligent monitoring, unmanned driving, work automation and medical diagnosis, etc., aiming at detecting tiny objects such as pedestrians, vehicles, industrial parts, cells and so on in a wide range of images.

In order to improve the detection of tiny objects, researchers have carried out a series of research in terms of network structure, training strategy, data processing and other aspects. However, at present, there is still a big gap between the performance of tiny object detection and normal scale object detection.

Although great progress has been made in object detection in recent years, tiny target detection is still a challenge. Current object detection algorithms are not suitable for tiny objects, and the reasons for the difficulties are as follows:

1) If the field of view of the whole input image is relatively large, but the object is very tiny, its characteristics are not obvious, which means the obtained target information is very limited, and there will be a lot of interference information from the background and low resolution. As a result, the detection process can be easily interfered by the image noise, leading to no detection and missed detection [1] [2].

2) The lack of relevant data sets and training data as well as the relatively small number of samples in the existing data sets containing tiny objects will lead to object detection models that focus more on the detection of normal scale objects.

3) Tiny objects cover only a small area; thus, the location of tiny objects lacks diversity.

Basic solutions

According to different ideas of improvement, the methods of tiny object detection can be divided into three parts. The first part is data enhancement, which makes some changes to the existing data sets to generate more equivalent (equally effective) data and increases the relevant data volume, so that the model obtained through the training set have stronger generalization ability. The second part is feature fusion, which means to fuse all the features together to show. The bottom feature has high resolution and contains a lot of position and detail information. However, it has less convolution and semantic information, and more noise. The high-level feature, on the contrary, has stronger semantic information, but the resolution is very low and the perception ability to details is poor. The fusion of the two can make the detection more efficient. The third part is resolution improvement, which aims to perceive more features and details so as to improve detection performance.

## 2. Definitions of Tiny Object

The definition of tiny object is mainly divided into two categories [3]: The first is the relative size definition, according to the Society of Photo-Optical Instrumentation Engineers (SPIE). It is defined as an object whose coverage is less than 80 pixels in an image of 256×256 pixels, that is, less than 0.12% of the 256×256 pixels [4]. The second is the absolute size definition, which is not clarified. Objects less than 32 x 32 pixels are defined as tiny objects in the COCO [5] data set. In 2016, Chen et al. defined a tiny object as an object from 16 x 16 pixels to 42 x 42 pixels in a VGA picture. Braun et al. [7] believe that for data such as pedestrians and non-motor vehicle drivers in traffic scenes, tiny objects refer to objects that are between 30 pixels and 60 pixels and blocked by less than 40%. In the image data set DOTA [8] in the field of aviation detection and the face recognition data set WIDER FACE [9], objects with pixel values in the range of 10 pixels to 50 pixels are considered as tiny objects. In the pedestrian detection data set City Persons [10], objects with a height of less than 75 pixels are defined as tiny objects.

## 3. Tiny Object Detection Methods

At present, the detection methods of tiny objects are based on depth learning, and some improvements are made on some mainstream object detection models for better performance in tiny objects detection.

### 2.1. Tiny object detection method based on data enhancement

Data enhancement means to use such ways as flipping, zooming, moving, oversampling, translation, rotation, noise suppression and so on to train samples for neural network learning.

A scale match method is proposed in reference [11], which is to match the scale of network pre-training data set (COCO) and detector learning data set (Tiny Person). The purpose is to align the object scales between the two data sets to achieve a desired representation of tiny objects. This method has won the champion in the First TOD Challenge [12] (In order to encourage people to study and develop innovative and accurate methods to detect tiny objects in images with wide field of view, the focus is on the detection of tiny people. A Tiny Person data set is provided for the game).

However, scale match algorithm adopts simple image-level zooming, which is only a simple approximation, and it is obviously inappropriate to take the average size of all objects in the image as the size of the image through simple scale matching. The SM algorithm is improved in reference [13], and SM+ is proposed, which extends the scale matching from image level to instance level, and effectively improves the similarity between the pre-training data set and the object data set. The method is to divide the image into two parts: labeled instances and backgrounds. Scale match has been conducted on these instances, and these images are damaged by some holes (backgrounds) in the shape of instances. The inputting of these holes in traditional ways leads to blurred or unreal images, which decrease the performance of the pre-training model. To solve this problem, the authors proposed a probabilistic structure inpainting (PSI) method to deal with the background, which achieves dynamic inpainting of the image by compressing the blurred images and preserving the context consistency of the edge of the holes, thus effectively balancing the information loss between the image structure and the semantic information. For sure, this method also has disadvantages. The pre-training data set is scaled to adapt to the object data set; The number of images in the pre-training data set far exceeds that in the object data set; And it will also limit the performance to a certain extent to reduce only the absolute size of the objects, but ignore their relative size.

In reference [14], a method of oversampling tiny objects is proposed to solve the problem that only a few images in the data set contain tiny objects; At the same time, in order to solve the problem that the area covered by tiny objects is much smaller and the position lacks diversity, the tiny object is copied and pasted in the same image for many times, so that the number of the tiny object in the image is increased, and the existing objects will not be covered in the pasting process. This increases the diversity of the location and ensures that the tiny objects appear in the right context. Compared

with the latest technology obtained by Mask R-CNN [15] on MS COCO [5], the recognition accuracy and segmentation accuracy of this method are improved by 7.1% and 9.7% respectively.

### 2.2. Tiny object detection methods based on feature fusion

In reference [16], a new multi-feature rotation detector called SCRDet, is proposed for tiny, densely arranged objects in arbitrary direction, which fuses the high and low level features. In particular, a sampling fusion network is designed, which consists of SF-Net, MDA-Net and rotation branch. Firstly, the feature map extracted from resnet backbone network is used to carry out multi-feature fusion and accurate feature sampling with SF-Net, so as to solve the problem of insufficient object information and anchor sample, and improve the sensitivity to tiny objects. Next, MDA-Net (consists of pixel attention and channel attention) is used to explore supervised pixel attention networks and channel attention networks for the detection of tiny, messy, densely arranged object by suppressing noise, enhancing object features and weakening non-object features. In order to solve the problem of large loss in the regression process, the IoU constant factor is added to the traditional Smooth L1 loss. Meanwhile, the IoU optimization regression task is the same as the measurement standard of the evaluation method, so that the regression is more effective and direct. A large number of experiments are carried out on two remote sensing public data sets DOTA [8], NWPU VHR-10 [17], natural image data set COCO [5], VOC2007 [18] and scene text data ICDAR2015 [19], and the method shows good performance in the detection of densely arranged and tiny objects.

The idea of fast detection of tiny objects in reference [20] is to use high-level feature maps to strengthen the semantic information of low-level feature maps, so as to increase the detection accuracy of tiny objects. SSD [21] is adopted as the basic structure, giving consideration to both speed and accuracy. SSD uses feature pyramid to take into account the detection of objects with different scales, and uses the shallowest features to detect tiny targets, because the receptive field is small, and the size of the receptive field exactly matches that of tiny objects. But the shallow features lack semantic information. The semantic information will influence the detector to judge whether the detection area is the object or the background, so the fusion of high-level features and low-level features will produce features suitable for the receptive field without lack of semantic information. In addition, in terms of how to fuse multiple features, this paper also tries two methods, namely Element-Sum module and concatenation module. The experiment results show that the former has better effect on blurred objects with less pixels, while the latter can reduce the influence of interference.

Reference [22] focuses on the research and improvement of the detector of FPN [23], so as to improve its performance in the detection of tiny objects. In this paper, a new concept of fusion factor is proposed to control the information transmitted from deep layer to shallow layer, so that FPN is suitable for the detection of tiny objects. The FPN-based detector has obvious effect on common object detection data sets, such as MS COCO [5], PASCAL VOC [24] and CityPersons [25], but has less satisfactory performance on tiny object detection data sets such as TinyPerson [26] and Tiny CityPersons [26]. It is proposed that adjusting the fusion factor between adjacent layers of FPN can promote the concentrated learning of tiny objects in shallow layers, thus improving the detection ability. The fusion factor is affected by the number of detectable objects in each layer, and the fusion factor can be learned by implicit methods. Most of the tiny objects are distributed in the P2 and P3 layers of FPN. At the same time, an important point of view is also put forward that feature fusion is affected by the object scale distribution of the data set, and most feature fusion methods overlook this factor.

### 2.3. Tiny object detection methods based on resolution improvement

The performance of tiny object detection can be improved by Generative Adversarial Network (GAN) or feature resolution improvement, so that the features of the tiny objects are similar to those of the large and medium-sized objects, which improves the detection performance of tiny objects.

1) Perpetual GAN

The method of using GAN for tiny object detection was first proposed in reference [28]. Perceptual GAN is divided into generator and perceptual discriminator. The generator converts the features of the tiny objects into super-resolved representation by introducing fine-grained features, and the discriminator judges whether the features come from the super-resolved representation features of tiny objects of the generator or the features of large objects, and calculates the loss from the detection gain in the generated super-resolved images. The output of the discriminator is the confrontation branch and the perception branch. The former is used for distinguishing features. The training instances are divided into two categories, namely large objects and tiny objects. Firstly, large objects are used to train the perception branch of the discriminator to obtain a higher detection rate. Next, based on the learned perception branch, tiny objects are used to train the generator. Thirdly, large objects and tiny objects are used to train the confrontation branch of the discriminator together. The training process of the confrontation branch of the generator and the discriminator network is alternately performed until a balanced point is finally reached. It shows improved performance on Tsinghua-Tencent 100K [29] data set.

2) MTGAN

Reference [30] proposed an end-to-end multitasking adversarial network MTGAN, which can be used in combination with any existing detector.

Tiny targets lack sufficient feature information to be distinguished from the background or similar objects. In MTGAN, the generator introduces a super-resolution network (SRN) to upsample the tiny object images to a larger scale (4× of the original image corresponding to the sample), and a multi-task discriminator network to distinguish the real and generated high-resolution images, predict the object category and refine the predicted bounding box at the same time. The category and regression losses are transmitted back to the generator to further guide the generative network to generate super-resolution images for easier classification and better localization. In this paper, the performance of this method is compared with that of Faster RCNN [31] and Mask-RCNN [32] in COCO [5] mini subset, and all indicators are improved.

The above two methods are tiny object detection methods based on GAN. The principle is similar, and tiny objects are difficult to detect. Insufficient feature is the root reason. These methods use GAN to learn an SR (super-resolution) network, introduce super-resolution of the image or features of the tiny objects, improve the information volume, enhance the features, so as to improve the detection accuracy. In order to reduce the calculation amount, both of them are based on RoI(Region of Interest) learning, instead of learning on the whole image or the whole feature map. The generator of P GAN learns the difference between the features of low-level and high-level of RoIs, while the generator of MTGAN learns the mapping of low-resolution RoIs images to high-resolution images. The difference lies in the level of super-resolved representation: MTGAN is the super-resolution of image-level, and the features must be extracted again after obtaining high-resolution image slices, so there will be more calculation; P GAN is the super-resolution of feature-level, thus there is no need for re-extraction of the features, resulting in less calculation.

3) HRDNet

Simply enhancing resolution has greatly increased the calculation amount of the model. In order to avoid this problem, HRDNet model is proposed in reference [33]. Its main idea is to use deep backbone network to process low resolution images, and shallow neural network to process high resolution images. The advantage of using shallow and small neural network to extract high-resolution images has been demonstrated in reference [34]. HRDNet includes two parts: MD-IPN and MS- FPN. MD-IPN inputs high-resolution images into shallow network, which keeps more location information and reduces the calculation amount. Convolution in shallow layer is input into deep layer to extract more semantic information. The performance of tiny object detection is improved by extracting different features of tiny objects through convolution from higher to lower layers, while recognition effect of the medium and large-sized objects has been maintained. MS-FPN is used to align and fuse the multi-scale feature map generated by MD-IPN to reduce the information imbalance between multi-scale and multi-depth features.

## 4. Data Sets Related to Tiny Object Detection

A good data set can effectively improve the performance of object detection, but most of the early object detection data sets are images of large and medium-sized objects, which is one of the reasons why tiny object detection is still a challenge. Therefore, in order to promote the progress of tiny object detection, many relevant data sets have emerged, including data sets in aviation: AI-TOD [35], DOTA [8] and relevant data sets of pedestrian detection such as TinyPerson [11] and so on.

Data set AI-Tod includes 8 categories of objects, with 28,036 images, 700,621 instance objects in total, and a size of 800×800 for each image.

DOTA uses 15 common target categories for annotation, with 2,806 aerial images and 188,282 instance objects. Each image is approximately 4000 x 4000 pixels.

TinyPerson includes 5 categories of objects, 1,610 images, and 72,651 instance objects. The resolution is 1920×1080, and some of which reach 3840×3160.

## 5. Conclusion

In this paper, the research of tiny object detection methods in recent years is summarized. Firstly, the concept of tiny object is introduced. Secondly, the difficulties of tiny object detection research are summarized. Thirdly, the methods proposed by existing literature are introduced from three aspects: data enhancement, feature fusion and resolution improvement. Lastly, several common tiny object data sets applied in different fields are introduced.

To conclude, progress has been obtained in the research of tiny object detection at this stage, but compared with the detection performance of normal objects, the performance of tiny object detection still has a long way to go. There are still a lot of problems to be solved in order to develop tiny object data sets of specific scene, and to design an excellent object detection algorithm with the existing depth learning theories and experience. Thus, tiny object detection still needs further development.

## References

[1] Kong, Fanjie, and Ricardo Henao. "Efficient Classification of Very Large Images with Tiny Objects." arXiv preprint arXiv:2106.02694 (2021).

[2] Van Etten, Adam. "You only look twice: Rapid multi-scale object detection in satellite imagery." arXiv preprint arXiv:1805.09512 (2018).

[3] Li, Kecen, et al. “Literature review on one-stage tiny object detection methods in depth learning.” Journal of Frontiers of Computer Science & Technology: 16.1 (2022): 41.

[4] Xie, Wenliang., Zhu, Dan., and Tong, Xinxin. “A tiny object detection method based on visual attention.” Computer Engineering and Applications: 49.12 (2013): 125-128.

[5] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." European conference on computer vision. Springer, Cham, 2014.

[6] Chen, Chenyi, et al. "R-CNN for small object detection." Asian conference on computer vision. Springer, Cham, 2016.

[7] Braun, Markus, et al. "The eurocity persons dataset: A novel benchmark for object detection." arXiv preprint arXiv:1805.07193 (2018).

[8] Xia, Gui-Song, et al. "DOTA: A large-scale dataset for object detection in aerial images." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[9] Yang, Shuo, et al. "Wider face: A face detection benchmark." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[10] Zhang, Shanshan, Rodrigo Benenson, and Bernt Schiele. "Citypersons: A diverse dataset for pedestrian detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[11] Yu, Xuehui, et al. "Scale match for tiny person detection." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020.

[12] Yu, Xuehui, et al. "The 1st tiny object detection challenge: Methods and results." European Conference on Computer Vision. Springer, Cham, 2020.

[13] Jiang, Nan, et al. "SM+: Refined Scale Match for Tiny Person Detection." ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021.

[14] Kisantal, Mate, et al. "Augmentation for small object detection." arXiv preprint arXiv:1902.07296 (2019).

[15] He, Kaiming, et al. "Mask r-cnn." Proceedings of the IEEE international conference on computer vision. 2017.

[16] Yang, Xue, et al. "Scrdet: Towards more robust detection for small, cluttered and rotated objects." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

[17] Cheng, Gong, Peicheng Zhou, and Junwei Han. "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images." IEEE Transactions on Geoscience and Remote Sensing 54.12 (2016): 7405-7415.

[18] Everingham, Mark, et al. "The pascal visual object classes (voc) challenge." International journal of computer vision 88.2 (2010): 303-338.

[19] Karatzas, Dimosthenis, et al. "ICDAR 2015 competition on robust reading." 2015 13th International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2015.

[20] Cao, Guimei, et al. "Feature-fused SSD: Fast detection for small objects." Ninth International Conference on Graphic and Image Processing (ICGIP 2017). Vol. 10615. International Society for Optics and Photonics, 2018.

[21] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.

[22] Gong, Yuqi, et al. "Effective fusion factor in FPN for tiny object detection." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021.

[23] Lin, Tsung-Yi, et al. "Feature pyramid networks for object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[24] Everingham, Mark, et al. "The pascal visual object classes (voc) challenge." International journal of computer vision 88.2 (2010): 303-338.

[25] Zhang, Shanshan, Rodrigo Benenson, and Bernt Schiele. "Citypersons: A diverse dataset for pedestrian detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[26] Yu, Xuehui, et al. "Scale match for tiny person detection." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020.

[27] Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems 27 (2014).

[28] Li, Jianan, et al. "Perceptual generative adversarial networks for small object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[29] Zhu, Zhe, et al. "Traffic-sign detection and classification in the wild." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[30] Bai, Yancheng, et al. "Sod-mtgan: Small object detection via multi-task generative adversarial network." Proceedings of the European Conference on Computer Vision (ECCV). 2018.

[31] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015): 91-99.

[32] He, K., Gkioxari, G., Dollar, P., Girshick, R.: Mask r-cnn. In: CVPR. pp. 2961–2969 (2017)

[33] Liu, Ziming, et al. "HRDNet: high-resolution detection network for small objects." 2021 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2021.

[34] Pang, Jiangmiao, et al. "$\mathcal {R}^ 2$-CNN: Fast Tiny Object Detection in Large-scale Remote Sensing Images." IEEE Transactions on Geoscience and Remote Sensing 57.8 (2019): 5512-5524.

[35] Wang, Jinwang, et al. "Tiny Object Detection in Aerial Images." 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021.