A Review of Aerial Images Object Detection Based on Deep Learning

Xuechun Wang

Abstract— **Due to the close relationship between object detection and image understanding, it has attracted a lot of research attention in recent years. Driven by deep learning, the problem of object detection has been developed rapidly, especially in natural scenes, there have been a series of breakthroughs, but the progress in remote sensing images has been slow. Due to the fact that the detection object of the aerial image is generally small, the object may be rotated in the picture, and the detection instance is large in magnitude. As a result, the existing object detection algorithm directly used in the aerial images object detection effect is not ideal. This article first introduces several popular object detection algorithms and analyzes the characteristics of each algorithm. Secondly, the characteristics of the aerial images data set are introduced, and the existing aerial images object detection algorithms are analyzed and summarized. Finally, discuss the existing problems and some insights on future object detection work.**

Index Terms—**Deep Learning; Object Detection; Aerial Images.**

I. Introduction

Object detection is a challenging and important problem in the field of computer vision. The function of traditional hand-made object detection methods will limit the ability of representation and cannot provide the required accuracy. The object detection method based on deep learning has entered people's field of vision and achieved unexpected results. The application of current object detection algorithms in natural scenes, such as face recognition and vehicle detection, has made breakthrough progress, but the progress in aerial images has been relatively slow. Due to the large scale and diverse types of aerial images, and there are relatively few aerial image data sets with good annotations, the direct application of existing object detection algorithms to aerial images is not ideal.

The object detection of aerial images has great practical value in both civil and military fields, and it plays an important role in traffic control, emergency rescue and other fields. In some applications, there are high requirements for the speed and accuracy of aerial image recognition. The existing object detection directly used for aerial image recognition obviously does not meet our requirements, so aerial images based on deep learning object recognition technology has become the focus of computer vision research at home and abroad. This article will focus on analyzing the specific direction of the existing aerial image object detection, and provide ideas for improving the speed and accuracy in the future

Manuscript received August 08, 2020
Xuechun Wang, School of computer science and technology，Tiangong University, Tianjin, China

II. Review Of Related Studies

A. *Existing object detection methods*

The current object detection methods based on deep learning are mainly divided into two stage and one stage object detection algorithms. The former is based on the candidate region method, using neural networks for sample classification, such as R-CNN[1], SPP-Net[2], Fast R-CNN[3], Faster R-CNN[4], etc. The latter directly converts the problem of object frame positioning into regression problem processing, such as the YOLO [5]series algorithm and the SSD[6] algorithm.

In 2014, Girshick R et al. [1] combined Region Proposal and CNN for the first time, and proposed the R-CNN algorithm, which achieved unexpected performance effects. In 2015, Girshick R et al. [3] improved the R-CNN algorithm and added Region of Interest (ROI) on the original basis. Softmax replaced the SVM used by R-CNN for classification to achieve end-to-end detection, called Fast R- CNN. In 2017, Ren S et al. [4] proposed to add the Faster R-CNN algorithm of the Region Proposal Network (RPN) to the Fast R-CNN algorithm to further improve the speed and ensure the accuracy.

In 2015, Redmon J et al. [5] proposed the YOLO (You Only Look Once) algorithm, which uses DarkNet as a separate neural network for feature detection to complete the output from the image input to the object location and category information. In 2016, Liu W et al. [6] proposed SSD (Single Shot MultiBox Detector) to extract multi-scale features.

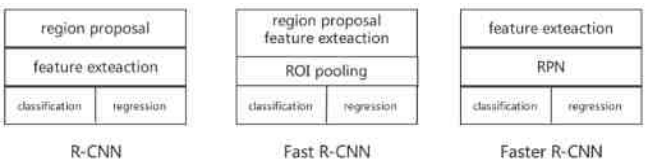

Fig 1.Structure of two-stage object detection algorithm

B. *Datasets*

For deep learning, the choice of data set is very important. For a specific object detection task such as aerial images, using current public image data sets (such as MSCOCO [7] and VOC [8]) for training is not ideal. Aerial images have the following characteristics: 1. The instances in the images are of

larger order of magnitude, and the resolution of aerial images is higher compared to ordinary data sets. 2. Objects in aerial images often appear in arbitrary positions, and the object itself may rotate. 3. The scale of the object in the aerial image varies greatly, most of which are small objects, and the detection is difficult.

In order to solve the above problems, in view of the characteristics of aerial images, aerial image data sets are proposed, UCAS-AOD [9] data sets, images are collected in Google Earth, and are used for object detection of vehicles and aircraft. The CARPK [10] data set is used for vehicle detection and counting detection, which contains nearly 90,000 vehicles. VEDAI (Vehicle Detection in Aerial Imagery) [11] data set is used for various types of vehicle detection. The DOTA [12] data set is a large-scale data set specially used for aerial image object detection. It contains 2806 aerial images, including 15 categories such as plane, ship, storage tank, and tennis court.

C. Aerial image object detection algorithm

For the difficulty of aerial image data sets, combined with the current application of deep learning-based object detection algorithms in natural scenes, it provides a lot of ideas for aerial image object detection. Many studies have proposed improved algorithms and achieved significant results. This article will analyze some representative methods. The following is an introduction to related literature.

The number of instances in aerial images is large, the resolution is high, and the object distribution is uneven. The processing cost of directly detecting these images is very high. Therefore, the first task is to reduce the cost of object search. There are many algorithms to improve efficiency from this perspective. Yang F et al. [13] proposed an end-to-end clustering detection framework (ClusDet) that combines object clustering and detection. The key components in ClusDet include a cluster proposal network (CPNet), a scale estimation network (ScaleNet), and a dedicated detection network (DetecNet). Given an input image, CPNet generates a object cluster area, and ScaleNet estimates the object scale of the cluster area. Then, the normalized cluster area of each scale is input into DetecNet for object detection. The experiments on the data sets of VisDrone, UA VDT, and DOTA show that ClusDet not only improves the efficiency of detection operation, but also improves detection accuracy. Uzkent B et al. [14] combined reinforcement learning and convolutional neural networks to detect small objects in large images. The task of reinforcement learning is to efficiently find small objects in the image. The algorithm first conducts a rough search on the image, first divides the image into sub-pictures of the same size, and calculates their respective gains after enlargement. Next is the fine search, further search optimization of the selected sub-pictures, and finally decide which sub-pictures to enlarge. The experimental results on the xView data set show that, while maintaining the detection accuracy, the operating efficiency is increased by 50%, and the high resolution time is only 30%.

Because aerial images are taken at high altitude, objects in the image may appear in any position and direction. Even the same type of objects have different shapes in the image. This makes the task of object detection more difficult. Some research on improved algorithms Specifically to solve this problem. Ma J et al. [15] proposed to add angle parameters on the basis of the Faster R-CNN algorithm, generate an anchor frame with angle information, and obtain a candidate region RRPN (Rotation Region Proposal Networks) that can have any direction, thereby using the rotated candidate region to perform Detection. Although this method improves the accuracy of object detection, it greatly increases the amount of calculation due to the generation of more rotating anchor frames. Jian Ding et al. [16] proposed a supervised rotating ROI, which uses a horizontal anchor frame, which is obtained through fully connected learning in the RPN stage. Under the supervision and annotation of OBB (Oriented Bounding Box), the ROI is spatially transformed and RT (ROI-Transformer) is learned, and then rotation invariant features are extracted from the rotating ROI for subsequent object classification and position regression. The algorithm is in DOTA And the detection performance on the HRSC data set has been significantly improved.

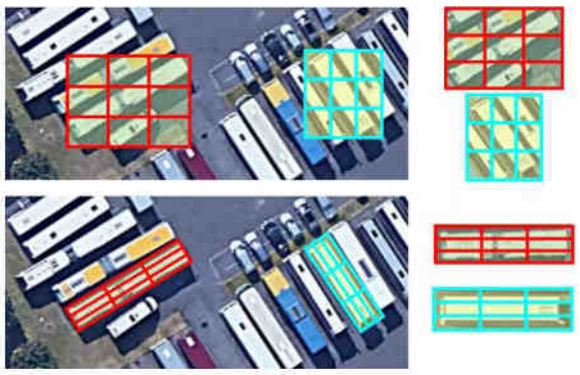

Fig 2.Horizontal v.s. Rotated RoI warping

Lin Y et al. [17] designed Feature Pyramid Networks (FPN), which realized the fusion of more detailed bottom-level features with rich semantic information. The FPN algorithm uses the intranet feature pyramid instead of the feature image pyramid, which greatly reduces the amount of calculation and solves the problem of inconsistency between training and testing time. Yang X et al. [18] used dense connections in DenseNet for the FPN algorithm, and combined dense connections and horizontal connections in a top-down network to improve the resolution characteristics. Wang J et al. [19] used an improved Inception module to replace the horizontal connection in FPN to improve speed and accuracy. Although the addition of the FPN module improves the effect of aerial image object detection, it reduces the speed of the algorithm. In general, it is not an ideal object detection method, and the speed needs to be improved. Yu X et al. [20] changed their thinking, added training samples under the existing algorithm, adjusted the scale of these training samples, and proposed a scale matching method to make it similar in scale to the statistical attributes of the object data set (TinyPerson), which improves object detection efficiency for small objects in complex backgrounds. However, this algorithm has very strict requirements on the data set, and cannot achieve general applicability. For the problem that most of the aerial images are small objects and difficult to detect, whether it is from the model itself,

combining the module with FPN to improve the algorithm, or from the perspective of the existing data set, improving the quality of training is practical.

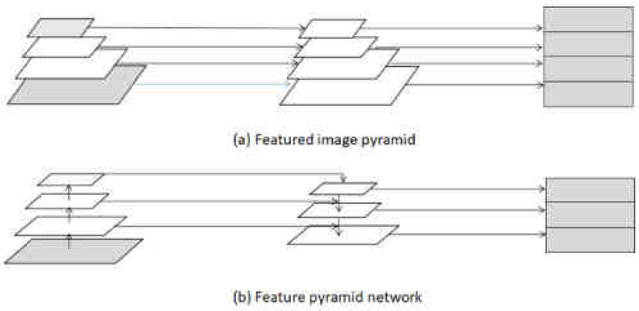

Fig 3.Pyramid structure in computer vision

III. Conclusions And Discussion

Aerial image object detection is widely used in disaster prediction, urban planning, and military. Although existing algorithms have achieved good detection results, there is still much room for improvement. The aerial image itself has a large order of magnitude, the object is rotated, and the object is small, which increases the difficulty of detection. Most of the algorithms are based on existing detection methods and add new modules and networks, which increase the limitations of detection and make it difficult to deal with complex aerial images. This paper summarizes the research results of deep learning in aerial image object detection, and analyzes the results obtained in this field. object detection algorithms in aerial images still face many challenges. We can try to combine deep learning methods with other methods to improve the effect of aerial image object detection.

References

[1] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

[2] HE K, ZHANG X, REN S. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[3] GIRSHICK R. Fast R-CNN [C]//CVPR 2015: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2015: 1440-1448.

[4] REN S, HE K, GIRSHICK R, Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[5] REDMON J, DIVVALA S, GIRSHICK R, You only look once: unified, real-time object detection[C]//CVPR 2016: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2016:779-788.

[6] LIU W, ANGUELOV D, ERHAN D, SSD: single shot multibox detector[C]//ECCV 2016: 2016 European Conference on Computer Vision. Berlin: Springer, 2016: 21-37.

[7] LIN T Y, MAIRE M, BELONGIE S.Microsoft coco: common objects in context[C]//ECCV 2014:2014 European Conference on Computer Vision. Berlin:Springer, 2014: 740-755.

[8] EVERINGHAM M, ESLAMI S M, VAN GOOL L, The pascal visual object classes challenge: a retrospective[J]. International Journal of Computer Vision, 2015, 111(1): 98-136.

[9] ZHU H, CHEN X, DAI W, Orientation robust object detection in aerial images using deep convolutional neural network[C]//2015 IEEE International Conference on Image Processing. Piscataway, NJ: IEEE,2015:3735-3739.

[10] HSIEH M, LIN Y, HSU W H, Drone-based object counting by spatially regularized regional proposal network[C]//ICCV 2017: Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 4165-4173.

[11] RAZAKARIVONY S, JURIE F. Vehicle detection in aerial imagery[J]. Journal of Visual Communication and Image Representation, 2016: 187-203.

[12] XIA G S, BAI X, DING J, DOTA: A large-scale dataset for object detection in aerial images[C]//CVPR 2018: Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2018: 3974-3983.

[13] YANG F, FAN H, CHU P, Clustered object detection in aerial images[C]//ICCV 2019: Proceedings of the 2019 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2019: 8311-8320.

[14] UZKENT B, YEH C, ERMON S. Efficient object detection in large images using deep reinforcement learning[C]//WACV 2020: 2020 IEEE Winter Conference on Applications of Computer Vision. Washington, DC:IEEE Computer Society, 2020: 1824-1833.

[15] MA J, SHAO W, YE H, Arbitrary-oriented scene text detection via rotation proposals[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111—3122.

[16] DING J, XUE N, LONG Y. Learning ROI transformer for oriented object detection in aerial images[C]//CVPR 2019: Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition.Washington, DC: IEEE Computer Society, 2019: 2849-2858.

[17] LIN T Y, DOLLAR P, GIRSHICK R. Feature pyramid networks for object detection[C]//CVPR 2017:Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2017: 2117-2125.

[18] YANG X, SUN H, FU K. Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks[J]. Remote Sensing, 2018,10(1):132-146.

[19] WANG J, DING J, GUO H. Mask OBB: a semantic attention-based mask oriented bounding box representation for multi-category object detection in aerial images[J]. Remote Sensing, 2019, 11(24):2930-2951.

[20] YU X, GONG Y, JIANG N. Scale match for tiny person detection[C]//WACV 2020: 2020 IEEE Winter Conference on Applications of Computer Vision.Washington, DC: IEEE Computer Society, 2020: 1257-1265.