FAST AIRCRAFT DETECTION BASED ON REGION LOCATING NETWORK IN LARGE-SCALE REMOTE SENSING IMAGES

Zhongxing Han, Hui Zhang, Jinfang Zhang, Xiaohui Hu

Institute of Software Chinese Academy of Sciences(ISCAS)

ABSTRACT

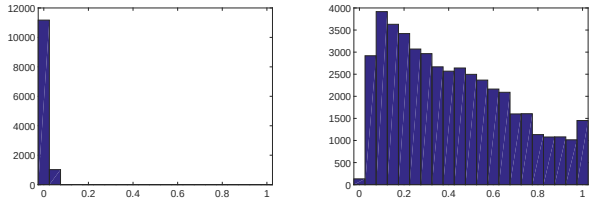
Nowadays, we get more and more remote sensing (RS) images which cannot be well processed or used by manual analysis or existing automatic methods. In the past few years, the object detection technology has greatly developed, especially after the usage of CNN in object detectors. However, Object detection in large scale RS images is still a challenging tasks which needs further study. Compared to natural images, RS images include much more objects in different sizes with a larger scope. Therefore, it is extraordinarily time-consuming to detect small objects in a large-scale RS image, since this work needs more scale and location traverses. Algorithms for common images cannot tackle the problem of some special object detection, like aircraft detection, in RS images. In this paper, we introduce an extra Region Proposal strategy named Region Locating Network (RLN) to improve the Faster RCNN framework. The proposed RLN locates spectacular areas where aircrafts are usually found, like parts of the runway and the parking apron. Based on the locating result, we can use Faster RCNN to detect airplanes in several smaller image regions. Extensive experiments show that the proposed method has oblivious improvement in recall rate, accuracy and computing efficiency.

Index Terms— Object Detection, Region Locating Network, Large-Scale Remote Sensing Image, Faster R-CNN

1. INTRODUCTION

Nowadays, RS images' number is becoming much more than before with their resolution growing higher. Many RS images can reach to 0.1 meters per pixel or even better. The issue comes with that how to process these images and extract the needed information with acceptable processing speed. To take full advantage of these images, it is important to promote the capacity of intelligent image process and analysis in large scale and at high speed. The common sliding windows and cropping method is hard to complete this work, which has some difficulties like long running time and redundancy process. In this paper, we are introducing a method named Region Locating Network (RLN) based aircraft detection, which can detect aircrafts in large-scale RS images at remarkable speed.

This work is supported by the Natural Science Foundation of China (U1435220) (61503365)



(a) (b)

Fig. 1. The sizes of annotated objects compare to the image sizes. The abscissa axis is the relative ratios and the vertical axis is the number of annotated objects. (a) is the ratios distribution histogram in our dataset, (b) is in Pascal VOC 2012.

Object detection plays a important role in many RS applications. There were many research works about object detection. [1] transformed images into saliency maps for unsupervised feature extraction and applied deep belief networks(DBNs) to detect objects. [2] used rotation-invariant CNN to solve the problem of object rotation variations. [3] tended to employ rotation invariant parts based model. However, these papers didn't refer to some problems resulted by large-scale RS images. Most directly, we can progress the whole images by common methods, but it will cost too much time since of the large size of images. To speed up the image progress, image resizing is a common approach. The problem also comes with that some objects like aircrafts are too small to be detected after image resizing. We set R as relative ratio which shows the ratio of sizes of annotated objects to the sizes of images. $R = L_1/L_2$, L_1 is the size of an annotated object, and L_2 is the size of image. Seeing Fig 1, compared to the common object detection dataset like Pascal VOC, the annotated objects in our dataset, a set of the images of airports, have much smaller relative ratios, which means that these aircrafts are too small to be detected after the image resized to small ones. Another approach to deal with this problem is to slide the RS image into several small ones, but this method still has some problems. First, the same as progressing the whole images directly, image segmentation method also costs much time. Second, most of the separated images are useless

for object detection, because it is almost impossible to detect particular objects in some areas. Third, some objects may be separated into two or more parts, which makes detection work more difficult.

Facing the same problem, how do humans do this work? To detect aircrafts in a large-scale RS image, we tend to glance at the whole image first. In this step, we just observe the image generally rather than watch the details. Then we can be well aware of where the aircrafts appear with high probability and where we don't need to seek anymore. In this paper, we follow this idea to improve the object detection algorithm for large-scale RS images. In generic object detection methods, people tend to "teach" computer "what does the objects look like" (the features of particular objects). However, before "teaching" them this, we can teach them "where can these objects be detected" . We introduce an additional R-CNN, named Region Locating Network, to select the regions where aircrafts locate with high probabilities, and then send these regions into a trained aircraft detector in the approach of Faster R-CNN to detect these aircrafts. The contribution of this paper includes: 1) we introduce a new thought of multiple region proposal locating method to detect mini-size object in large-scale RS images; 2) we speed up the whole detection remarkably; 3) the accuracy also has obliviously risen with much less false detection.

The rest of the paper is organised as follows. Section 2 introduces the related works. Section 3 describes the fast aircraft detection based on Region Locating Network. Section 4 shows experiments and results. Section 5 concludes the whole paper.

2. RELATED WORKS

In the past few years, techniques in object detection had remarkable improvement. Nowadays, most of object detection methods contains two different thoughts : the region proposal based method and the end-to-end method.

As the name tells, the region proposal based method contains two steps in object detection: obtaining the proposals from an image (where are the objects), and classifying which category of the object in every proposal belongs to (which kind is the object). R-CNN [4], Fast R-CNN [5] and SPP-net [6] used region-based convolutional neural networks(R-CNNs) to complete the step of object classification, which can be trained and test by GPUs to get remarkable speed acceleration. On the basis of these works, Faster R-CNN [7] pointed out that the region proposal methods are implemented on the CPU, which limits the speed of object detection, and implemented the region proposal progress on GPU using extra CNNs. As a result, the accuracy and the detection speed are both promoted.

On the other hand, the end-to-end method has only one neural network while the former has two. YOLO [8] is a representative method of end-to-end object detection. YOLO separates the whole image into 7x7 parts and lets every part take charge of an object's location. SSD [9] improved the performance based on YOLO. Instead of using only one default separation number(7x7), SSD uses several different separation numbers like 4x4 and 8x8. The bigger boxes are proper to detect big objects and the smaller ones are suitable for small objects. Compared to the region proposal based methods, YOLO and SSD have outstanding processing speed. On the contrary, they have a same problem that their separations are set before training so that it is hard for them to detect some special small objects like aircrafts in large-scale airport images. Considering all the above, we choose the Faster R-CNN as the basic object detection method.

3. OUR METHOD

In general, the proposed algorithm includes mainly three steps. Step 1, locate the region of aircrafts using Region Locating Network (RLN). Step 2, locate aircrafts with Region Proposal Network(RPN). Step 3, classify the aircrafts with Fast R-CNN and final process.

3.1. Aircraft Annotations Based Clustering

As we know, the original training data only marks the location of aircrafts. It is hard to deal with these citations directly. Before the RLN training, data preprocessing is needed. To transform the object location into the region location, we use clustering algorithm to combine several adjacent bounder boxes into one field boxes. The common clustering algorithms include partitioning methods, hierarchical methods, density-based methods, grid-based methods and so on. Since the numbers of objects in images are usually less than 200, we don't need to use a complex algorithm. In this letter, we use hierarchical based methods to complete this work.

For each RS image, we suppose each marked aircraft is an independent cluster at first. During each iteration, we check the minimum one in every two clusters' distance is whether less than a set threshold. If not, break the loop. In our method, the approach of distance calculation is Euclidian distance. While clustering, we set two different thresholds to limit the size of every cluster. Multiple sizes of clusters contribute to increase the amount of training data as well as avoiding over-fitting. In several images, some aircrafts are so far from others that they cannot be clustered. In this case, we treat these aircrafts as outlier and delete them from clustering results, since many of them have negative effects on the recognition of the special field with aircrafts. Because clustering annotations aims to get the information of where the aircrafts often locate, we expand every cluster to cover more background. To avoid over-fitting, the sizes of expansions are set from 45 to 75 pixels randomly.

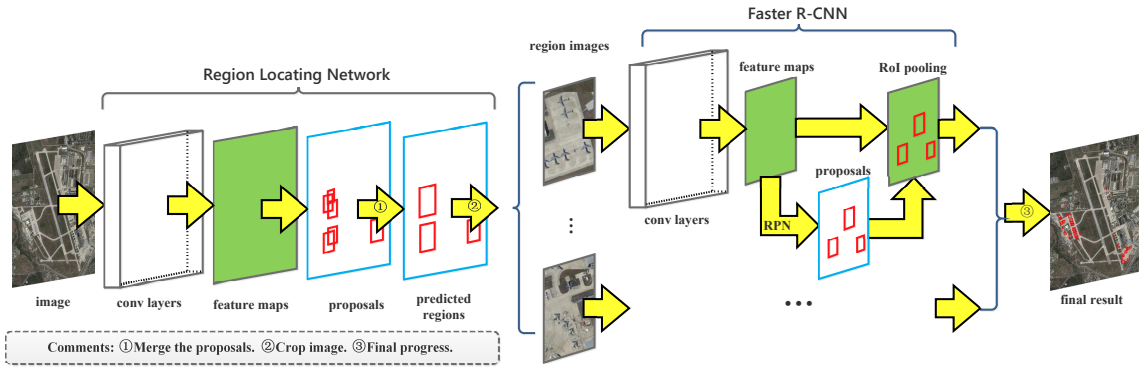


Fig. 2. The structure of the aircraft detector with proposed RLN

3.2. Region Locating Network Training

The RLN is a fully convolutional network that predicts the region bounds and the objectness scores. Its structure is the same as the RPN in Faster R-CNN(VGG16 [10]), which has 16 convolution layers in total. Different from the original version, in our method, its input can be parking aprons, areas of several aircrafts and some parts of runway, which are the regions that aircrafts locate in with high possibility.

For RLN training, we use the preprocessed training data. In [7], they used multi-scale anchors as regression references to train the RPN. The same with them, we also use default 3 scales(128^2, 256^2 and 512^2) and 3 aspect ratios(1:1, 1:2 and 2:1), yielding k = 9 anchors at each sliding position. We use Intersection-over-Union(IoU) overlap to estimate two bound boxes's overlap ratio. The formula to calculate IoU overlap is $IoU = S_o/(S_{box1} + S_{box2} - S_o)$. In this formula, S_{box1} and S_{box2} are the areas of two proposals, and S_o is the overlap area. For each sliding window, if this window has IoU higher than 0.7 with any cluster box, we treat it as the positive input. If its IoU ratio is lower than 0.2 with all cluster boxes, we treat it as the negative one and the anchors that are neither positive nor negative have no contribution to train the network. In general, the RLN training needs the labeled anchors and the image as inputs and the region bounds and the objectness scores as outputs, then we use back-propagation and stochastic gradient descent to update the parameters. Finally, we can obtain the RLN which can locate the areas with high possibility that the aircrafts locate in.

3.3. Faster R-CNN

Faster R-CNN [7] includes RPN and an object detection network(based on Fast R-CNN [5]). RPN is used to generate proposals and their objectness scores, and Fast R-CNN takes the former output as input and operates classification for each proposal. Faster R-CNN shares the convolutional layers of RPN and Fast R-CNN by 4 steps training method. In the first 2 steps, we train RPN and Fast R-CNN separately. In the next 2 steps, we train RPN and Fast R-CNN again by using the same convolutional layers and just fine-tuning their respective layers. Finishing these 4 steps of training, we can obtain an aircraft detector, whose inputs are the images and outputs are the predicted category labels, the bounding boxes and the objectness scores for all boxes.

Fig. 3. Images in our dataset

4. EXPERIMENTS

4.1. Our Dataset

In this paper, the training and testing datasets are annotated by ourselves. The whole dataset contains 265 images of 12 different airports and over 6,000 annotated aircrafts in total. Every image is a snapshot of a whole or a large part of airport, downloaded from Google Earth. The original RS images' sizes are from 2,048x2,048 to over 40,000x30,000. The most important characteristic of this dataset is the relative ratios in our dataset are much less than other datasets for object detection. To increase the amount of training data, the images in training set are reversed horizontally and vertically. For the test data, we randomly select 1/10 images in our dataset to test the performance.

4.2. Experiment Setup

For every image, the RLN locates several boxes as the possible fields which aircrafts situate in. Afterwards, we merge these boxes into some separated field boxes. The merging method is similar to the former cluster method. We check whether every two boxes are overlapped or not and combine them together if overlapped, until these boxes are all separat-

Table 1. Result compare

	Accuracy	Recall Rate	Mean Time
Our Method	**53.64%**	**65.71%**	**1.506s**
F.RCNN(6x6)	19.04%	46.73%	4.872s
F.RCNN(8x8)	17.46%	61.56%	8.445s
F.RCNN(10x10)	13.84%	63.87%	13.166s

ed. After this, all boxes are expanded by 100 pixels. This step of expansion aims to prevent some aircrafts to locate in an edge of region boxes, which may be missed during detection. For each field box, we crop the relevant part of image as the input image of Faster R-CNN detector and get the final output. If the IoU overlap of two boxes is larger than 0.6, we think them to be the same aircraft and combine them together. We also use IoU overlap to measure whether an output proposal is right or not. If an proposal can get more than 0.6 IoU overlap with any annotated aircrafts, this proposal is thought to be a right answer.

4.3. Experimental Analysis

Since compared to the whole image, the sizes of aircrafts are too small, it is pointless to process this work using common object detection methods directly. Seeing the results in Table 1 and the P-R curve in Fig 4, compared to the simple segmentation and detection method, the performance of our method is much better.

In the comparison experiments, we separate the images into 6x6, 8x8 and 10x10 parts and use Faster R-CNN to detect the aircrafts. The contrast of results are shown in Table 1. The accuracy, the recall rate and the running speed of our method are all better than others. For the simple separation methods, fewer separations(like 6x6 parts) cannot make the aircrafts big enough in every part of image to be detected. By increasing the number of separations(from 6x6 to 8x8 and 10x10), the recall rate is risen. However, new problems limit the detector to have better performance. First, the running time of every part of images is generally equal. More separations means more running time. Another problem caused by more separations is more useless parts of images, which have negative effects on aircraft detection. The aircraft detector may recognise some other objects as aircrafts, which cause the accuracy to be lower. At last, more separations lead to that more aircrafts are cropped into two or more parts which are hard to detected.

Compared to the simple separation methods, the RLN based method has many advantages. First of all, our method has fewer segmentations of images. Since the running time of the first region proposal network is much less than the runtime of aircraft detection, it is a linear relation between the total running time and the number of segmentations. The number of parts of regions is approximately between 6 to 12, which is obviously fewer than the simple separation methods'. Furthermore, fewer aircrafts are cut due to fewer separations, which is beneficial to higher recall rate. Another strength is the purposiveness of detection. The RLN promptly limits the detection range of images which the aircrafts have high possibility to locate in. By this way, the number of final detection proposals decreases remarkably, and the accuracy also extremely increases. Seeing Fig 5, compared to the simple separation methods, our method has much less false detection. Meanwhile, the strong purposiveness of detection ensures the recall rate.

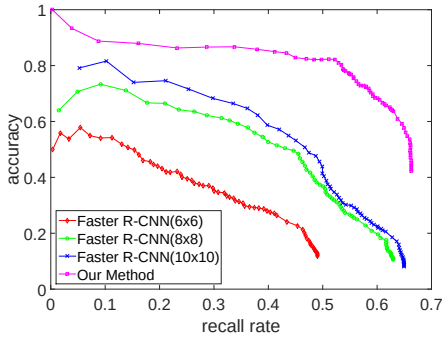
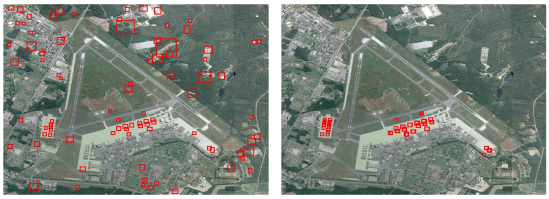


Fig. 4. P-R curve of several methods



(a) (b)

Fig. 5. Compare between two methods. (a) is the result of Faster R-CNN(6x6), (b) is from our method.

5. CONCLUSION

In this paper, we present a new method named Region Locating Network (RLN) for aircrafts detection in large-scale RS images. By using RLN, the aircrafts detector can locate rough areas of aircrafts in a large-scale RS image and crop these regions for further aircrafts detection. Compared to the common object detection method, our method has remarkable improvements in accuracy, recall rate and running speed. The RLN can also be used in the detection of other particular objects, like oil tanks, ships and so on, and we only used aircrafts to give an example. This method also has some limitations. It can only be used in detection of some special objects which often locate in specific regions. Moreover, this method doesn't apply for single aircraft detection which may be treated as outlier in the RLN.

6. REFERENCES

[1] W. Diao, X. Sun, X. Zheng, F. Dou, H. Wang, and K. Fu, "Efficient saliency-based object detection in remote sensing images using deep belief networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 2, pp. 137–141, Feb 2016.

[2] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in vhr optical remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7405–7415, Dec 2016.

[3] W. Zhang, X. Sun, K. Fu, C. Wang, and H. Wang, "Object detection in high-resolution remote sensing images using rotation invariant parts based model," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 1, pp. 74–78, Jan 2014.

[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142–158, Jan 2016.

[5] R. Girshick, "Fast r-cnn," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015, pp. 1440–1448.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, Sept 2015.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–1, 2016.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 779–788.

[9] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," *ArXiv e-prints*, Dec. 2015.

[10] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.