

PAPER • OPEN ACCESS

## Small Aircraft Detection in Remote Sensing Images Based on YOLOv3

To cite this article: Kun Zhao and Xiaoxi Ren 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **533** 012056

View the [article online](#) for updates and enhancements.

# Small Aircraft Detection in Remote Sensing Images Based on YOLOv3

**Kun Zhao, Xiaoxi Ren**

Hunan University, Changsha, China

happyrx@hnu.edu.cn

**Abstract.** Accurate and rapid detection in remote sensing images is of great significance in military navigation, environmental monitoring, and civil applications. Due to the small object detection problem of remote sensing images, higher requirements and challenges are put forward for object detection technology. In recent years, the one-stage object detector and the two-stage object detector based on convolutional neural network have made great achievements in the field of image classification and detection. The one-stage object detector is generally superior to the two-stage object detector in detection speed, but the detection accuracy is inferior to the two-stage target detector. In this paper, YOLOv3 is used as a one-stage target detector for small aircraft detection of remote sensing images. We chose appropriate anchors to cover the size distribution in our experimental data by using dimension clusters. The experimental results show that YOLOv3 not only exceeds the conventional one-stage object detector in speed, but also well match the accuracy of the two-stage object detector. The YOLOv3 achieved excellent detection accuracy and low processing time at the same time for small aircraft detection.

## 1. Introduction

With the development of space remote sensing technology, the resolution of remote sensing images has been greatly improved, and the information carried by remote sensing images is becoming more and more abundant. In recent years, aircraft detection in remote sensing images [1-3] has become one of the research hotspots in the fields of environmental monitoring, military and civil applications, etc. Most of the traditional aircraft detection methods use artificially extracted shallow features, such as the shape feature, or scale-invariant feature transform (SIFT) [4] and histogram of oriented gradients (HOG) [5], then select support vector machine (SVM) or Adaboost as the classifier. For example, Cheng et al. [6] used multi-scale histogram of oriented gradients (HOG) feature and trained a latent SVM for each object category. Luo et al. [7] utilized Object Proposal algorithm for airplane locating, and a SVM classifier was trained on the HOG features to detect the airplanes. However, the traditional methods extract the low-level features of remote sensing images, and can not fully exploit the high-level features. What is more, the selection and extraction of features are subjective and the extraction process is very complicated. The traditional methods are far from meeting the needs of large-scale automation applications in terms of detection speed and accuracy. For nearly 10 years, with the continuous development of big data and computing capacity, deep learning methods have been rapidly developed and widely used in various fields. Among them, convolutional neural network (CNN) proved to be a promising approach for feature extraction has caused great concern in the field of image recognition and detection. Many theoretical researches concerning CNN have been quite



mature [8-13]. There are many detection methods based on CNN for aircraft detection have been proposed. Zhang et al. [1] proposed a weakly supervised learning framework based on coupled CNN for aircraft detection. Wu et al. [14] put forward a new aircraft detection framework based on objectiveness detection techniques and CNN. A modified Faster R-CNN [15] for the task of small object detection in optical remote sensing images was proposed in [16]. Nonetheless, superior detection speed and accuracy can hardly be achieved at the same time for small aircraft detection. In this paper, YOLOv3 [17] is used as a object detector for small aircraft detection, and compared with two-stage object detector Faster R-CNN [15] and one-stage object detector SSD [18] in terms of performance. Experiment results show that the YOLOv3 outperforms the other two methods.

## 2. Methodology

### 2.1. Object Detection Based on CNN

Object detection method based on CNN was successfully applied to object detection as early as 1994 [19]. Due to lack of data, limitations in hardware performance, and over-fitting problem, object detection based on CNN has not made a breakthrough for a long time. Compared with the traditional object detection method at present, the object detection based on CNN has no great advantage in both detection accuracy and detection speed. Therefore, the research is gradually neglected. Until 2012, the convolutional neural network AlexNet [20] made a major breakthrough in image classification, and researchers began to refocus on how to apply CNN to object detection. Nowadays, CNN-based object detection methods occupy a dominant position in the field of object detection, which are mainly divided into two categories, one is two-stage object detector represented by R-CNN (SPP Net [21], Fast R-CNN [22], Faster R-CNN[15], Mask R-CNN [23]), and the other is one-stage object detector represented by SSD [18] and YOLO [24].

### 2.2. YOLOv3

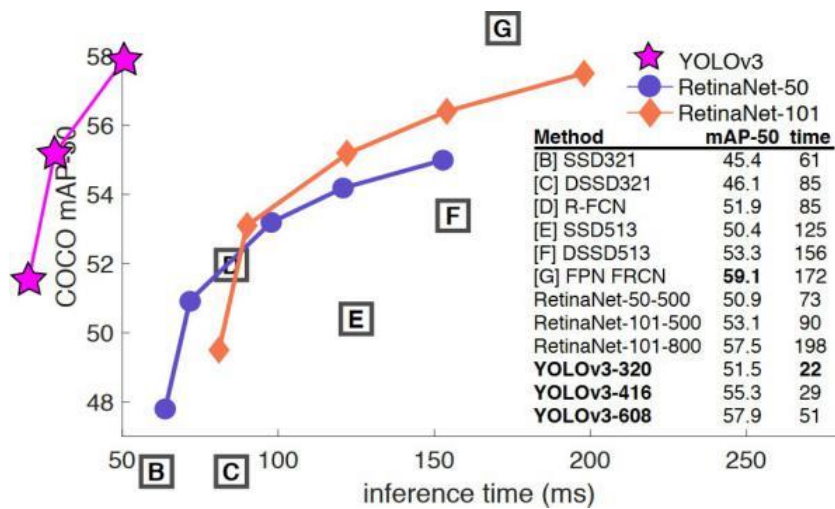
1) YOLOv3 has made some improvements on the basis of YOLOv2 [25], and these changes make it even better. The network structure of YOLOv3 is shown in Figure 2. The performance comparison of YOLOv3 with other detection methods in the COCO dataset [26] at 0.5 IoU(Intersection- of-Union) metric is shown in Fig 1. We can observe that YOLOv3 performs very well, both accuracy and speed can be considered at the same time. These improvements are mainly reflected in the following aspects.

2) Unlike the Darknet-19 used in the YOLOv2 [25], YOLOv3 uses a 53-layer convolutional network named Darknet-53 that is superimposed by residual units as feature extraction networks. According to the author ' s experiments in [17], this model performs better than Darknet-19 in terms of classification accuracy and efficiency.

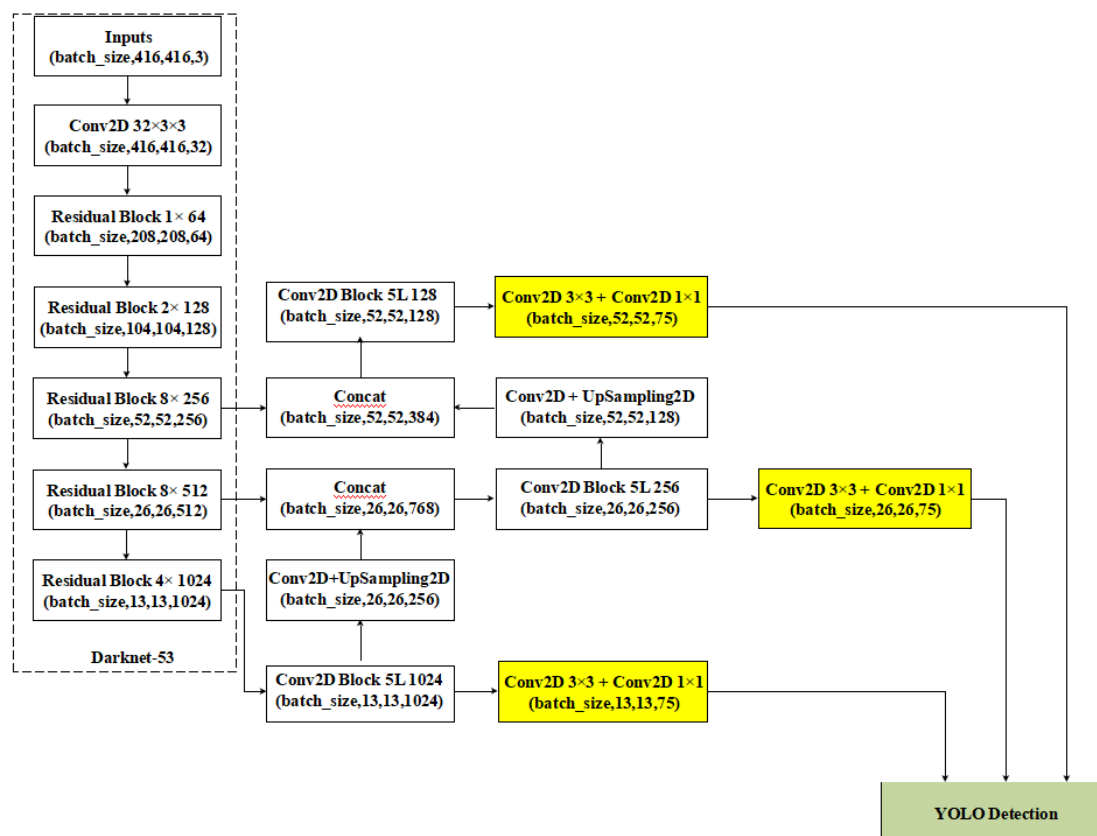
3) To achieve multilabel classification, use logistic classifier instead of softmax and use binary cross-entropy loss for the class predictions during training. YOLOv3 uses logistic regression to predict the score for each bounding box. If the bounding box prior overlaps a ground truth object by more than any bounding box prior. If the priori bounding box is not the best, but does overlap the ground truth by more than a certain threshold, then this prediction is ignored. YOLOv3 only assigns a bounding box to each ground truth. If the a priori bounding box does not match the ground truth, no coordinate or class prediction loss will occur, and only objectness.

4) Different from the previous YOLO and YOLOv2, YOLOv3 predicts tasks from three different scale feature maps, which better solve the small target detection problem.

YOLOv3 made some improvements on the basis of the original, which further improved the performance of the model.



**Figure 1.** The Speed/Accuracy Tradeoff on the Map of the Detection Methods at 0.5 Iou Metric in the COCO Dataset.



**Figure 2.** The Structure Diagram of YOLOv3.

### 3. Experiments

The experiment consists of two processes: the training phase and the detection phase. All the experiments were performed on Ubuntu with NVIDIA GTX 1080Ti GPU, and the evaluation metrics are average precision (AP) with IoU threshold as 0.5 and average detection time per image. If the

overlap between the bounding box and ground truth is greater than or equal to 0.5, the bounding box is a TruePositive; otherwise, it is a FalsePositive.

### 3.1. Dataset Description

In our experiments, we collected 350 remote sensing images from Google Earth and DOTA dataset [27], each image is of the size in the range from about 600x600 to 1500x1500 pixels and we manually labeled aircraft targets in each image. The training validation set and the test set are randomly generated in a ratio of 8:2. The training set and the verification set are randomly generated by the training verification set according to the ratio of 8:2. The experimental data statistics are shown in Table 1. Some samples of experimental data are shown in Figure 3.

**Table 1.** Statistics of the Data Sets.

Class	Image Number	Aircraft Number
Training set	224	4528
Verification set	56	965
Testing set	70	1372
Total	350	6865



**Figure 3.** Some Samples of Experimental Data

### 3.2. Evaluation Indicators

In the experiment, we use the AP and the average detection time of each image to evaluate the performance of the models. The indicators we used in the experiment are as follows.

- 1) True Positive (TP): The number of samples that are predicted to be positive and actually positive;
- 2) True Negative (TN): The number of samples that are predicted to be negative and practically negative;
- 3) False Positive (FP): The number of samples that are predicted to be positive and actually negative;
- 4) False Negative (FN): The number of samples that are predicted to be negative and practically positive.

The measures can be defined as following:

$$P(\text{Precision}) = \frac{TP}{TP + FP} \quad (1)$$

$$R(\text{Recall}) = \frac{TP}{TP + FN} \quad (2)$$

$$AP(\text{AveragePrecision}) = \int_0^1 P(R) dR \quad (3)$$

### 3.3. Experimental Results and Comparisons

We evaluated the performance of YOLOv3 object detector for small aircraft detection by comparing it with Faster R-CNN and SSD. For a fair comparison, the same arguments for training that are listed in Table 2 and the same test set are used in the experiment. Some experimental results is shown in Figs 4, 5 and 6.

**Table 2.** Arguments for training

Argument	Value
Learning rate	0.01(20k:0.001, 50k:0.0001)
Batch Size	32
Momentum	0.9
Weight decay	0.0005
Max iteration	80k

**Figure 4.** The detection results of SSD**Figure 5.** The detection results of Faster R-CNN**Figure 6.** The detection results of YOLOv3

The corresponding AP value and detection time are shown in Table 3.

**Table 3.** Detection Results in Term of the Metrics of AP and Average Detection Time

Method	AP	Average detection time
SSD	0.538	0.078s
Faster R-CNN	0.874	0.652s
YOLOv3	0.925	0.023s

Form Figs 4, 5 and 6 and Table 2, the following conclusions can be obtained: the YOLOv3 achieves much better performance than the Faster R-CNN and the SSD. The YOLOv3 outperforms the Faster R-CNN and the SSD by 0.051 and 0.387 in terms of AP, respectively. In terms of detection time, the average detection time of YOLOv3 is 0.023s. The time cost of Faster R-CNN is 27 times than that of YOLOv3. The YOLOv3 is 3.4 faster than SSD. As can be seen, a large number of mislabeling and missing labels appear in the SSD detection, followed by Faster R-CNN, and YOLOv3 can accurately detect the objects in different scenarios. To sum up, we believe that the YOLOv3 is more suitable and has obvious advantages for small aircraft detection in remote sensing images, and can meet the requirements of accurate detection and real-time in practical applications.

#### 4. Conclusion

In this paper, we presented an effective detection method to address the small object problem in remote sensing images by using YOLOv3. We analyzed some improvements that YOLOv3 made on the previous basis, and compared the performance with other object detection algorithms. Comparisons with Faster R-CNN and SSD approaches demonstrated the effectiveness and superiority.



YOLOv3 is real-time while accurately detecting small aircraft. In the future, we will continue to work on multi-class small object detection in remote sensing images based on the current research basis.

### Acknowledgments

The research was supported by the nation fund project for the study of key issues in high-availability on-board computing under Grant 61802122.

### References

- [1] Fan Zhang, Bo Du, Liangpei Zhang, and Miao Zhong Xu. Weakly supervised learning based on coupled convolutional neural networks for aircraft detection. *IEEE Transactions on Geoscience & Remote Sensing*, 54(9):5553–5563, 2016.
- [2] Shijin Li, Jianbin Qiu, and Yu Hui. Aircraft detection in high-resolution remote sensing imagery based on visual words selection. *Journal of Data Acquisition & Processing*, 2014.
- [3] Yuqingyang Hou, Jicheng Quan, and Yongming Wei. Valid aircraft detection system for remote sensing images based on cognitive models. *Acta Optica Sinica*, 2018.
- [4] Junwei Han, Dingwen Zhang, Gong Cheng, Lei Guo, and Jinchang Ren. Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. *IEEE Transactions on Geoscience & Remote Sensing*, 53(6):3325–3337, 2015.
- [5] Gong Cheng, Peicheng Zhou, Xiwen Yao, Chao Yao, Yanbang Zhang, and Junwei Han. Object detection in vhr optical remote sensing images via learning rotation-invariant hog feature. In *International Workshop on Earth Observation and Remote Sensing Applications*, pages 433–436, 2016.
- [6] Gong Cheng, Junwei Han, Lei Guo, Xiaoliang Qian, Peicheng Zhou, Xiwen Yao, and Xintao Hu. Object detection in remote sensing imagery using a discriminatively trained mixture model. *Isprs Journal of Photogrammetry & Remote Sensing*, 85(9):32–43, 2013.
- [7] Qinhan Luo and Zhenwei Shi. Airplane detection in remote sensing images based on object proposal. In *Geoscience and Remote Sensing Symposium*, pages 1388–1391, 2016.
- [8] Ali S Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. Cnn features off-the-shelf: An astounding baseline for recognition. pages 512–519, 2014.
- [9] Y Lan Boureau, Jean Ponce, and Yann Lecun. A theoretical analysis of feature pooling in visual recognition. In *International Conference on Machine Learning*, pages 111–118, 2010.
- [10] Pulkit Agrawal, Ross Girshick, and Jitendra Malik. Analyzing the Performance of Multilayer Neural Networks for Object Recognition. Springer International Publishing, 2014.
- [11] Jeffrey Dean, Greg S. Corrado, Rajat Monga, Kai Chen, Matthieu Devin, Quoc V. Le, Mark Z. Mao, Marc’Aurelio Ranzato, Andrew Senior, and Paul Tucker. Large scale distributed deep networks. In *International Conference on Neural Information Processing Systems*, pages 1223–1231, 2012.
- [12] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. pages 1–9, 2014.
- [13] Hossein Azizpour, Ali Sharif Razavian, Josephine Sullivan, Atsuto Maki, and Stefan Carlsson. Factors of transferability for a generic convnet representation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 38(9):1790–1802, 2014.
- [14] Hui Wu, Hui Zhang, Jinfang Zhang, and Fanjiang Xu. Fast aircraft detection in satellite images based on convolutional neural networks. In *IEEE International Conference on Image Processing*, pages 4210–4214, 2015.
- [15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: towards real-time object detection with region proposal networks. In *International Conference on Neural Information Processing Systems*, pages 91–99, 2015.
- [16] Yun Ren, Changren Zhu, and Shunping Xiao. Small object detection in optical remote sensing images via modified faster r-cnn. *Applied Sciences*, 2018.

- [17] Joseph Redmon and Ali Farhadi. Yolo3: An incremental improvement. 2018.
- [18] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. pages 21–37, 2015.
- [19] R Vaillant, C Monrocq, and Y Le Cun. Original approach for the localisation of objects in images. *Vision, Image and Signal Processing*, 141(4):245 – 250, 1994.
- [20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *International Conference on Neural Information Processing Systems*, pages 1097–1105, 2012.
- [21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell*, 37(9):1904–1916, 2015.
- [22] Ross Girshick. Fast r-cnn. *Computer Science*, 2015.
- [23] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, PP(99):1–1, 2017.
- [24] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788, 2016.
- [25] Joseph Redmon and Ali Farhadi. Yolo9000: Better, faster, stronger. pages 6517–6525, 2016.
- [26] Tsung Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. 8693:740–755, 2014.
- [27] Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.