# Specific Object Detection And Recognition in Optical Remote Sensing Images

Tao Hong hongt@myorbita.net

Zhuhai Orbita Aerospace Science & Technology Co., Ltd.

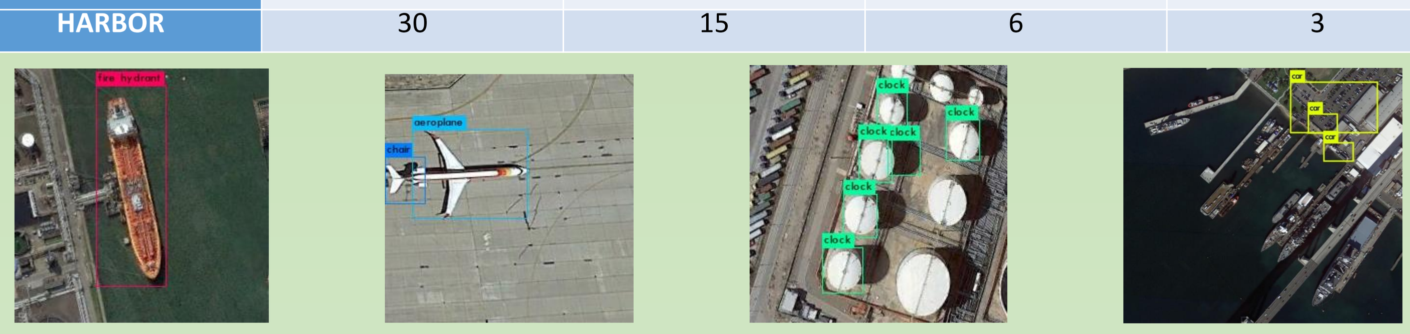
## ABSTRACT

This article uses ten typical man-made targets such as airplane, ship, storage tank, harbor, bridge, and vehicle as objects of detection. Specific object detection algorithms for optical remote sensing image are studied to meet the application needs of more recognition accuracy and higher speed. Applying Faster R-CNN algorithm and YOLO v2 algorithm to the object detection task, adaption changes are applied at the application level targeted at specific requirements of optical remote sensing images. Faster R-CNN realized the detection and recognition of remote sensing multi-category man-made objects with an average accuracy of 71.2%. The recognition speed of YOLO v2 detector can reach 67FPS with a detection time of about 15ms/image, and meet the real-time detection requirements.

## INTRODUCTION

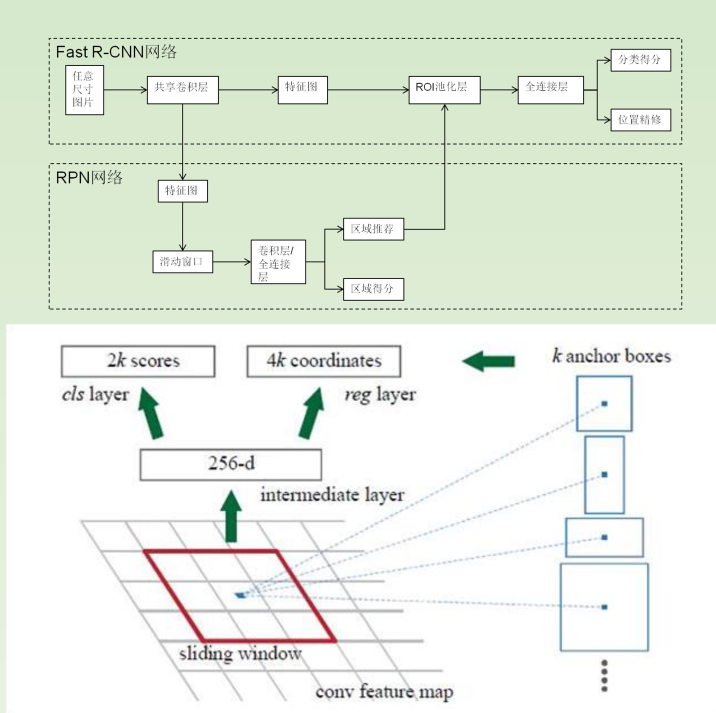
The degree of discrimination of different targets at different resolutions (unit: meter)

| OBJECT | Detection | Identification | Classification | Description |
|---|---|---|---|---|
| BRIDGE | 6 | 4.5 | 1.5 | 0.9 |
| ROAD | 9 | 6 | 1.8 | 0.6 |
| VEHICLE | 1.5 | 0.6 | 0.3 | 0.05 |
| AIRPLANE | 4.5 | 1.5 | 0.9 | 0.15 |
| HARBOR | 30 | 15 | 6 | 3 |

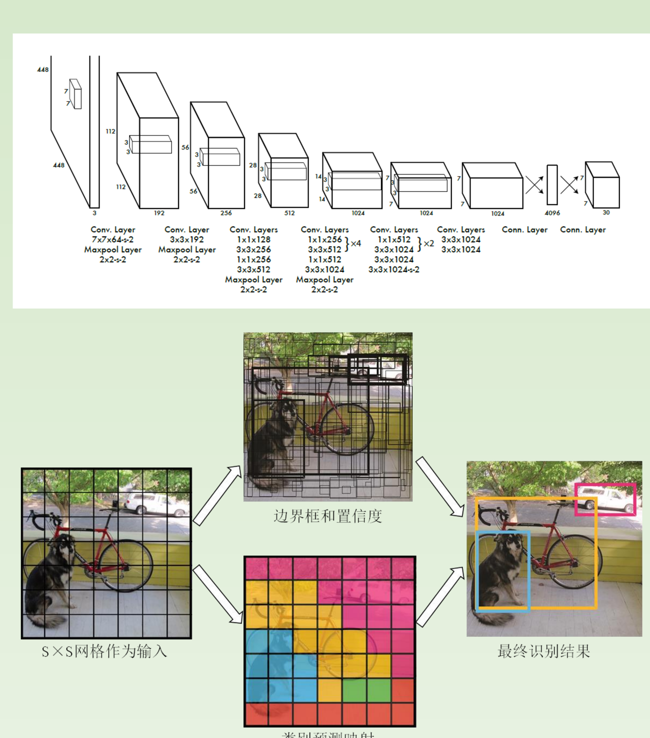


- The scale variations of object instances in aerial images are huge.
- Many small object instances are crowded in aerial images. Moreover, the frequencies of instances in aerial images are unbalanced.
- Objects in aerial images often appear in arbitrary orientations.

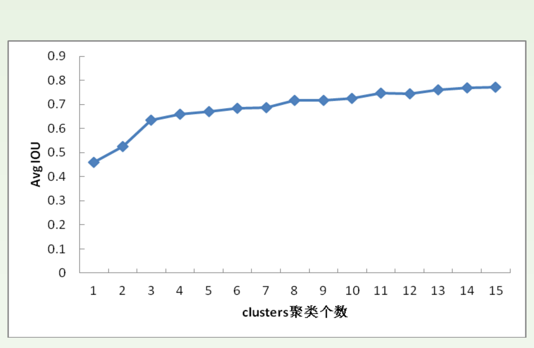
## METHODS

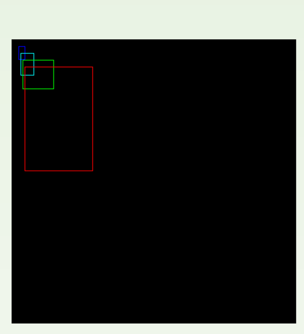



FASTER R-CNN



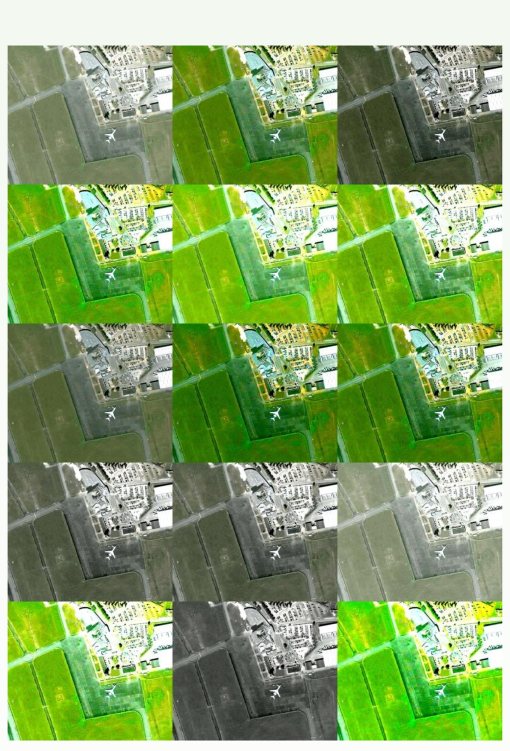
YOLO



**Clustering analysis of remote sensing dataset NWPU VHR-10**

Four kinds of anchor boxes of different sizes calculated under the optimal clustering result, one of which is a square and the other three are thin and high rectangles, and the four kinds of frames are basically small for the overall image size, which can reflect the remote sensing image object detection dataset is more of a small-scale target.



data augmentation 15 times

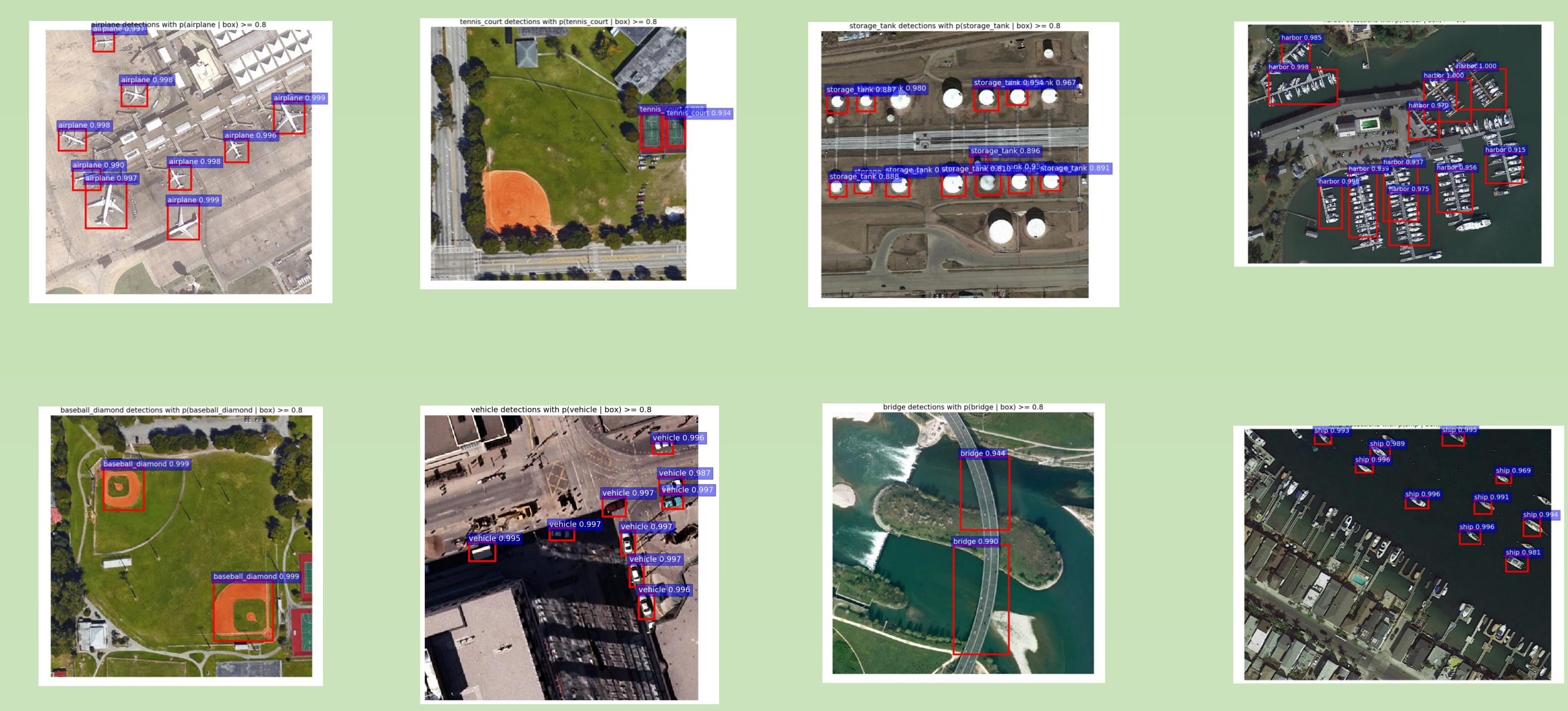
For the problems existing in optical remote sensing image of color-texture interference, rotation-scale variance and shape-similar interference, eight enhancement processing for training data are performed, such as flipping, panning, rotation, and color jitter, etc.

The expanded sample size significantly improves the model training results, and the average accuracy is achieved by using model migration learning.

**39.3% → 47.2% → 71.2%**

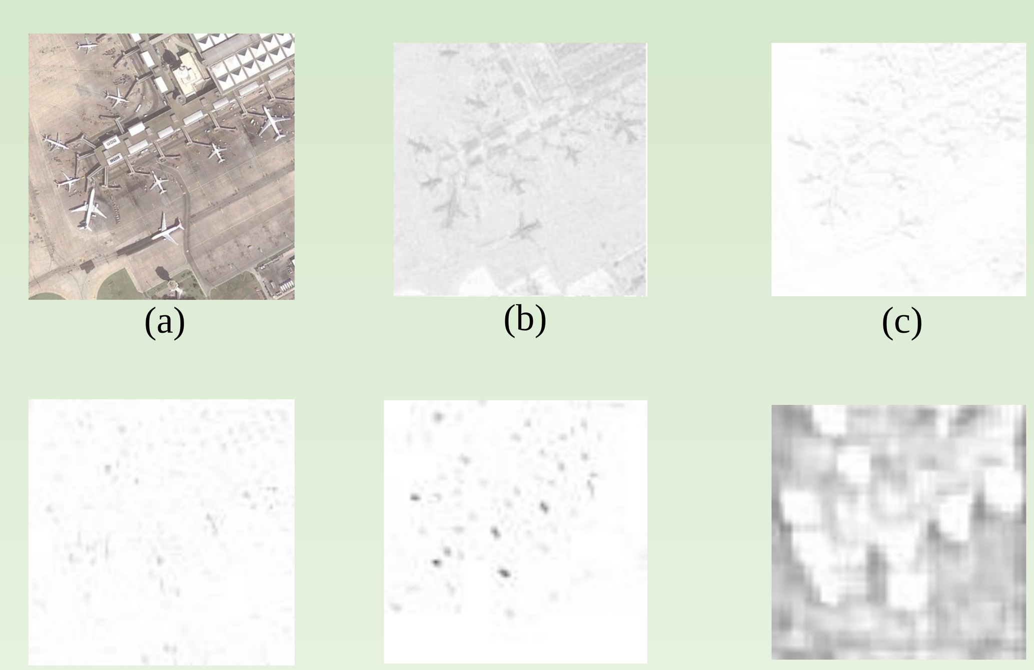
| Object Category | AP（ImageNet pre-training） | AP（data augmentation） | AP（original dataset） |
|---|---|---|---|
| Airplane | 80.3 | 45.5 | 36.1 |
| Ship | 68.1 | 45.4 | 44.3 |
| Storage tank | 45.9 | 45.0 | 25.7 |
| Baseball diamond | 90.6 | 50.0 | 45.3 |
| Tennis court | 71.5 | 45.4 | 40.4 |
| Basketball court | 67.7 | 45.5 | 40.7 |
| Ground track field | 89.2 | 50.0 | 45.4 |
| Harbor | 76.9 | 50.0 | 40.8 |
| Bridge | 57.2 | 50.0 | 40.8 |
| Vehicle | 64.6 | 45.2 | 33.9 |
| Mean AP | $\boxed{71.2}$ | $\boxed{47.2}$ | $\boxed{39.3}$ |

## RESULTS



| | Test parameter | | | |
|---|---|---|---|---|
| Iteration times | IOU | Recall | Proposals | Precision |
| 50,000 | 50.57% | 64.38% | 431 | 80.10% |
| 100,000 | 66.89% | 80.22% | 5176 | 88.90% |
| 150,000 | 67.38% | 80.94% | 5087 | 91.33% |

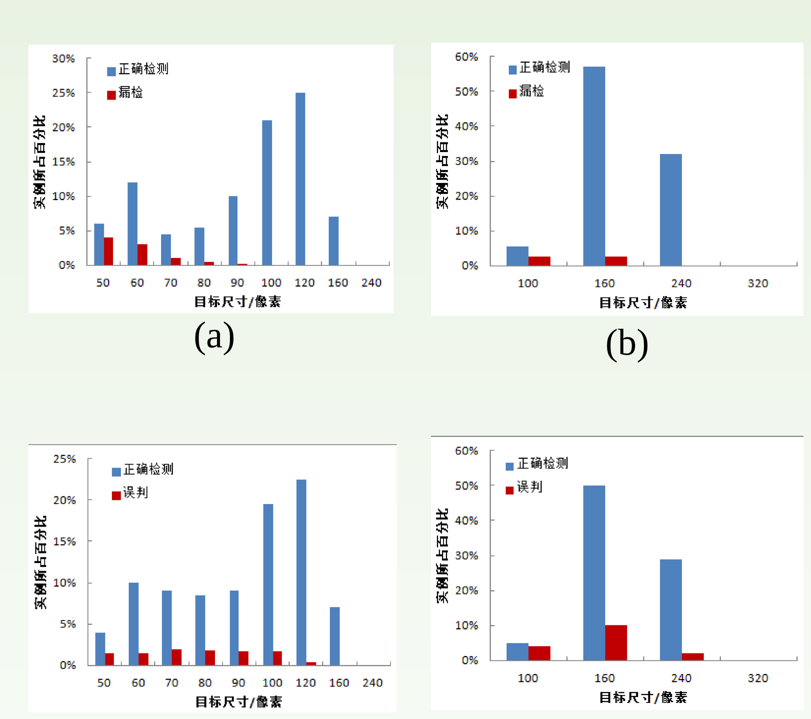
## DISCUSSION



(a) (b) (c) (d) (e) (f)

(a) The original input image;
(b) Convolution layer 1_2 feature map;
(c) Convolution layer 2_2 feature map;
(d) Convolutional layer 3_3 feature map;
(e) Convolutional layer 4_3 feature map;
(f) Convolutional layer 5_3 feature map

The first two layers mainly extract the **edge structure feature information**, after reaching the upper level gradually becomes **more comprehensive semantic information**, which is not easy to understand.



(a) (b) (c) (d)

(a) size distribution of aircraft recalls and missed inspections;
(b) size distribution of ship recalls and missed inspections;
(c) size distribution of aircraft correct detection and misjudgment;
(d) size distribution of ship correct detection and misjudgment

The bounding box that is not properly recalled mainly exists on the **smaller target**. This phenomenon is more obvious on the aircraft target. The lowest recall rate is the target with the size between 50-60 pixels. The detection accuracy of small targets are also lower.

## MAJOR REFERENCES

[1] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.

[2] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[J]. 2015:779-788.

[3] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[J]. 2016:6517-6525.

[4] Cheng G, Zhou P, Han J. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images[J]. IEEE Transactions on Geoscience & Remote Sensing, 2016, 54(12):7405-7415.

[5] Cheng G, Han J, Lu X. Remote Sensing Image Scene Classification: Benchmark and State of the Art[J]. Proceedings of the IEEE, 2017, 105(10):1865-1883.