Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Improved YOLOv5 Network Method for Remote Sensing Image Based Ground Objects Recognition

**xue jie** ( 201983060062@sdust.edu.cn )
Shandong University of Science and Technology https://orcid.org/0000-0003-1446-3949
**Shandong University of Science and Technology**
Shandong University of Science and Technology
**Yongguo Zheng**
Shandong University of Science and Technology https://orcid.org/0000-0002-8859-5920
**Changlei Dong-Ye**
Shandong University of Science and Technology
**Ping Wang**
Shandong University of Science and Technology
**Muhammad Yasir**
China University of Petroleum Huadong - Qingdao Campus

**Research Article**

**Keywords:** high resolution remote sensing image, object recognition, YOLOv5, attention mechanism

**Posted Date:** February 2nd, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1224458/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

# Abstract

High resolution remote-sensing images have the characteristics of complex background environment, clustering of objects, etc., which lead to the problem of low accuracy in recognition of large ground objects such as airports, dams, golf field, etc. Based on this problem, this paper proposes a remote sensing image object detection method based on YOLOv5 network. By improving the backbone extraction network, the network structure can be deepened to get more information about large objects, the detection effect can be improved by adding attention mechanism and adding output layer to enhance feature extraction and feature fusion. The pre-training weight is obtained by transfer learning and used as the training weight of the improved YOLOv5 to speed up the network convergence. The experiment is carried out on DIOR dataset, the results show that the improved YOLOv5 network can significantly improve the accuracy of large object recognition compared with the YOLO series network and the EfficientDet model on DIOR datadet, and the mAP of the improved YOLOv5 network is 80.5%, which is 2% higher than the original YOLOv5 network.


# 1. Introduction

With the development of remote sensing technology, remote sensing has been widely used in modern industry and life [1]. At present, the object recognition of remote sensing images has also been paid more attention remote sensing image target detection is widely used in many fields, such as military reconnaissance [2], environmental monitoring [3], urban construction [4], climatic data analysis [5], building damage detection [6]. The traditional object detection algorithms can be divided into three parts. Firstly, the candidate regions are obtained by image segmentation technology or sliding window method, then the features from each region are extracted, and finally extracted features are put into the classifier for classification and recognition. Among them, the common feature extraction methods are Scale-invariant feature transform (SIFT) [7], Speeded Up Robust Features(SURF) [8], Histogram of Oriented Gradient (HOG) [9], etc. the common classifiers include support vector machines□SVM)[10], Haar [11], Adaboost [12], etc. Remote sensing images have the characteristics of dense ground objects and complex environment, so the traditional target detection algorithm requires a lot of calculation and it is inefficient.

In recent years, with the continuous development of deep learning, object detection based on Convolutional Neural Networks (CNN) [13] has been widely used in various fields, and gradually replaces the traditional detection methods. There are two kinds of object detection algorithms based on deep learning: those based on candidate regions and those based on regression algorithms. The former, also known as the two-stage model, divides object detection into two phases: generating region proposals, classifying region proposals in classifier and correcting positions, such as R-CNN [14], SPP-Net [15], Fast R-CNN [16], Faster R-CNN [17], Mast R-CNN [18], etc. The latter directly regresses the predicted object to generate the bounding box, such as You Only Look Once (YOLO) [19][20][21][22][23], Single Shot MultiBox Detector (SSD) [24], CenterNet [25], EfficientDet[26],etc.

According to the problems of remote sensing images, researchers have made unremitting efforts on the basis of CNN. Chen Jinyong et al. [27] proposed Domain Adaptation Faster R-CNN algorithm to improve robustness of the model and widen scope of application and proved its effectiveness. Han qinzhen et al. [28] proposed a remote sensing image building detection algorithm combining Mask R-CNN with traditional object detection algorithm. This method improves the detection accuracy and reduces the calculation time. Li Yangyang et al. [29] proposed a lightweight keypoints-based oriented object detector for remote sensing images in view of the complex background. Xu Danqing et al. [30] proposed a detection algorithm based on YOLOv3 for the detection of remote sensing targets at different scales, so that the detection targets dominated by small targets can maintain the detection speed and improve the average accuracy. Zhou Liming et al. [31] proposed the Multiscale Detection Network (MSDN) to solve the problem that aircraft size is small, and proposed the Deeper and Wider Module (DAWM) to resist the background noise. Finally, the DAWM is introduced into the MSDN and the novel network structure is named the Multiscale Refined Detection Network (MSRDN). Lu Qixiang [32] proposed a detection algorithm based on improved SSD to solve the problem of small and dense objects in remote sensing images, proposed a new loss function to speed up network convergence, and proposed Laplace-NMS method, which had good post- treatment effect on dense objects. Many researches are focused on small objects detection in remote sensing images, but experiments show that the recognition effect of some large objects is not ideal in complex environment. Therefore, the purpose of this paper is to propose an algorithm for large objects detection.

In the relatively complex environment, the recognition effect of large ground objects is poor, this paper presents a method of remote sensing image ground object recognition based on YOLOv5 [23]. This paper improves the YOLOv5s network structure by adding a set of CSP structure and attention mechanism to the backbone extraction network, and changing the output layer to four feature layers. Experiments show that the improved YOLOv5s network improves the accuracy of large ground objects detection compared with the original YOLOv5s.

# 2. Yolov5 Network

The YOLOv5 network structure consists of four parts: input, backbone, neck and prediction. The network structure is shown in Figure 1. There are four kinds of networks, YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x. They have the similar network structure with different depths and widths. The yolov5s network has the smallest depth and the shallowest depth, the fastest running speed and the lowest accuracy. On this basis, the other three network structures become gradually deeper and wider, the accuracy becomes continuously improved, on the contrary, the operation speed becomes slow.

The input uses mosaic data enhancement, adaptive anchor box operation and picture scaling to process the input dataset; The backbone adopts focus structure and CSP structure. Focus improves the network speed and reduces floating-point operations (FLOPs) by slicing the input picture. The focus structure is shown in Figure 1.1. YOLOv5 uses two CSP [33] structures: CSP 1_ X and CSP 2_ X, CSP1_ X is used for down-sampling in backbone, CSP 2_ X is used in neck. CSP can improve the learning ability of the

network and ensure the accuracy of the network while reducing the operations. The structure diagrams of the two CSPs are shown in Figure 1.1; Neck adopts SPP-net and FPN+PAN structure to enhance the feature fusion effect of the network; Prediction adopts GIOU_ Loss [34], GIOU not only pay attention to the overlapping area between the prediction box and the ground truth, but also to the non-overlapping areas. GIOU solves the problems of IOU [35] while maintaining the advantages of IOU. The calculation formula is as follows:

$$IOU = \frac{|A \cap B|}{|A \cup B|}$$

1

$$GIOU = IOU - \frac{\left| \frac{C}{(A \cup B)} \right|}{|C|}$$

2

In addition to four different networks, the version of YOLOv5 is also constantly updated. This article uses v5.0 of yolov5s, compared with v4.0, v5.0 of YOLOv5 changed all activation functions in the network to SiLU()[36], deleted the conv in the CSP and renamed it C3 as shown in Figure 2. Therefore, the network structure of v5.0 is smaller and faster than that of v4.0.

# 3. Yolov5 Network Improvement

### 3.1. Backbone improvement

Feature fusion of features with different scales can often get more useful object information. The low-level features have higher resolution, smaller receptive field, more texture information and more noise, which is suitable for detecting small objects; The high-level feature has lower resolution and poor perception of object details, but the receptive field is larger, which is suitable for detecting large objects. In DIOR dataset, the complex background environment leads to the unsatisfactory detection effect of some large ground objects. In this paper, a group of C3 structure is added to the backbone network of YOLOv5s, and the original three groups of C3 are changed to four groups of C3, so as to deepen the overall network structure, which can effectively improve the expression ability of the network and the learning ability of larger ground objects, and then improve the detection accuracy of the model.

### 3.2. Attentional mechanism

Attention mechanism refers to human visual attention mechanism, it focuses on local information and suppresses redundant information, in other words, the attention mechanism enables the network to find significant information among multitudinous information. In this paper, the attention mechanism [37] is added to the backbone of YOLOv5s to obtain the information of different characteristic channels and the

importance of different channels, and then suppress useless channels for the detection object. In this way, the network performance is enhanced by adding a small amount of computation. The improved backbone structure is shown in Figure 3.

### 3.3. Neck improvement

Neck adopts the FPN [38] +PAN structure. This structure adds a bottom-up feature pyramid network after FPN, which enhances the semantic expression and location information on multiple scales. Neck of YOLOv5 integrates C32_X structure to enhance the feature fusion effect of the network. Since a group of C3 structures are added in this paper, an output layer is added in neck to further improve the feature extraction of the network. The improved FPN+PAN structure is shown in Figure 4.

### 3.4. Remote sensing image ground objects recognition method based on improved YOLOv5 network

Remote sensing image ground objects recognition method based on improved YOLOv5 network needs to recalculate the anchor box firstly, and modify the data configuration file and network configuration file, and then the pre-training weight is obtained by means of migration learning. Then start training the network, read the data configuration file, analyze and load the network model, get the training weight and save it. Finally, the network detection is carried out, the trained network model is used to predict the training set, and the image with prediction class and boundary box is generated and output. The process of remote sensing image ground objects recognition method based on improved YOLOv5 network is shown in Algorithm 1.

Algorithm 1. remote sensing image ground objects recognition method based on improved YOLOv5 network

| |
|---|
| Input:800*800 size remote sensing image |
| Output: result image of object detection |
| The algorithm process: |
| 1: begin |
| 2: K-means to calculate anchor boxes and modify configuration files |
| 3: Transfer learning get pre-training weight |
| 4: if method = train |
| 5: then |
| 6: read_data.yaml// read the data configuration file |
| 7: parse_network// parse and load the network |
| 8: read_ network.yaml// read the network configuration file |
| 9: load_data |
| 10: training_network |
| 11: save_ training_weight |
| 12: if method = test |
| 13: then |
| 14: load_ network// load the trained network model |
| 15: read_data.yaml |
| 16: load_data |
| 17: pre_network// network to predict |
| 18: get_pre_box |
| 19: input_result |
| 20: end |

# 4. Experiment And Result Analysis

In this paper, DIOR dataset [39] is used to verify the detection effect of the improved YOLOv5s network; the precision, recall, mAP@.5 and mAP@.5:.95 are used as performance indicator of the evaluate model. Finally, the improved YOLOv5s network is compared with the original YOLOv5s network, YOLO series network and EfficientDet model on the DIOR dataset.

### 4.1. Dataset introduction

DIOR dataset is an open large-scale dataset proposed by Northwestern Polytechnical University for object detection of optical remote sensing images. The image size of the dataset is 800*800, including 23463 images and 190288 object instances, involving 20 object classes, including airplane, airport, basketball court, baseball field, bridge, chimney, dam, expressway toll station, expressway service area, golf field, ground track field, harbor, overpass, ship, stadium, storage tank, tennis court, trains station, vehicle and windmill. The dataset example is shown in Figure 5.

### 4.2. Evaluation index

Precision represents the accuracy of the correct target in the prediction result, and the formula is shown in formula (3); Recall represents the proportion of correct target in all prediction results, and the formula is shown in formula (4). Both tend to rise in one value and fall in the other. The P-R curve can be obtained by taking Precision as the vertical axis and Recall as the horizontal axis. The area surrounding the curve with x axis and y axis is called the average precision (AP). The formula is shown in formula (5). The P-R curve of the improved YOLOv5s under DIOR data set is shown in Figure 6. mAP is the mean average precision of multi-objective, and the formula is shown in formula (6). mAP@.5 indicates the mAP of all categories when the IOU is 0.5. mAP@.5:.95 refers to the average mAP of IOU between 0.5 and 0.95.

$$Precision = \frac{TP}{TP + FP}$$

3

$$Recall = \frac{TP}{TP + FN}$$

4

$$AP = \int_0^1 P(R)\,dr$$

5

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i$$

6

Where $TP$ is the number of instances where the correct object is identified as correct; $FP$ is the number of instances where the wrong target is identified as correct; $FN$ is the number of instances where the correct

target is identified as wrong; $N$ is the total number of object classes, and in this paper, $N$ is 20; $i$ is the AP of the $i$th object.

### 4.3. Training and result analysis

#### 4.3.1. Network training

The model training experimental environment of this paper: the graphics card is GeForce GTX 1080 Ti, GPU driver, CUDA version is 11.2, cuDNN version is 8.1, compilation language is Python3.8, batch size is 16 and epochs is 200.

Firstly, the overpass in RSOD dataset [40] and NWPU VHR-10 dataset [41] are combined into a new dataset by the transfer learning method, and this dataset is used to pre-train the improved YOLOv5s, and the generated pre-training weight is used as the training weight of the improved YOLOv5s. This method of migrating the data parameters of the pre-training model to the new model to help the new model training can improve the training speed of the improved YOLOv5s and accelerate the network convergence.

YOLOv5 network has the function of automatically calibrating the anchor box. Because an output feature layer is added to the network model in this paper, k-mean algorithm is used to recalculate the anchor box, and the original 9 anchor boxes are changed to 12 anchor boxes, which are (20,14), (19,38), (54,25), (38,76), (97,68), (84,189), (151,138), (208,230), (435,115), (207,491), (474,317), (476,590).

#### 4.3.1. Result analysis

In this paper, YOLOv5s is compared with the improved yolov5s on DIOR dataset, and precision, recall, mAP@.5, mAP@.5:.95 as performance indicators to evaluate the quality of the network as shown in Table 1. According to the table, the improved YOLOv5s is improved in precision, mAP@.5, mAP@.5:.95 compared with original YOLOv5s, and the recall is not as effective as original YOLOv5s.

The test results are shown in Figure 7. The original YOLOv5s test images are on the left and the improved YOLOv5s test images are on the right. It can be seen from images in the first row that the ship objects are small, densely distributed and cover the larger object harbors. The modified YOLOv5s can identify the harbors more accurately than the original YOLOv5s; the second row show that under complex background, the accuracy of the improved YOLOv5s for the large object of golf field is nearly 50% higher than the original YOLOv5s; images in the third row show that overpass is particularly similar to the expressway. The original YOLOv5s classified the expressway as overpass, while the improved YOLOv5s identifies overpass more accurately.

In order to further verify the effectiveness of the improved YOLOv5s, in this paper, the improved YOLOv5s is compared with the YOLO series network and EfficientDet model on DIOR dataset, and AP of 20 types of ground objects and mAP are used as the indicators of the evaluation algorithm. The experimental results are shown in Table 2.

It can be seen from the comparison results in Table 2 that the mAP of the improved YOLOv5s has significantly improved compared with YOLOv3, YOLOv4, the original YOLOv5s and EfficientDet. Among them, the detection accuracy of large ground objects with complex environment has been greatly improved, such as airport, dam, golf field, harbor and trains station, and the detection accuracy of other ground objects has also been improved to varying degrees. It is concluded that the improved YOLOv5s network has a good detection effect on large ground objects in YOLO series network.

# 5. Conclusion

Aiming at the problems existing in the recognition of large ground objects in remote sensing images base on YOLOv5s, this paper proposes a Remote sensing image ground objects recognition method based on YOLOv5 network. This paper uses DIOR dataset, firstly, the anchor box is recalculated, then CSP and attention mechanism are added to the backbone, and finally output layer is added to the neck, which improve the effect of YOLOv5s network on ground object recognition. The experimental results show that compared with the original YOLOv5s network, YOLO series network and EfficientDet model, the improved YOLOv5s network improves the recognition accuracy of large ground objects in complex environment, and its mAP reaches 80.5%. In the next step, on the basis of maintaining the existing advantages, we will further carry out research on improving the accuracy of small ground objects recognition.

# Declarations


### Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.


### Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

### Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Jie Xue. The first draft of the manuscript was written by Jie Xue and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

### Data Availability

The RSOD datasets generated during and/or analysed during the current study are available in GitHub - RSIA-LIESMARS-WHU/RSOD-Dataset-: An open dataset for object detection in remote sensing images.

The DIOR datasets and NWPU VHR-10 dataset during and/or analysed during the current study are not publicly available due to the link failure but are available from the corresponding author on reasonable request.

# References


1. Wei Z, Liu Y (2021) Construction of super-resolution model of remote sensing image based on deep convolutional neural network. Computer Communications 178: 191-200.https://doi.org/10.1016/j.comcom.2021.06.022
2. Liu F, Zhu J, Wang W (2021) Surface-to-air missile sites detection agent with remote sensing images. Sciece China. Information Sciences 64(9). https://doi.org/10.1007/s11432-019-9920-2
3. Zhang Y, Ning G, Chen S, Yang Y (2021) Impact of Rapid Urban Sprawl on the Local Meteorological Observational Environment Based on Remote Sensing Images and GIS Technology. Remote Sens 13:2624. https://www.mdpi.com/2072-4292/13/13/2624
4. Guo M, Shu S, Ma S, Wang L (2021) Using high-resolution remote sensing images to explore the spatial relationship between landscape patterns and ecosystem service values in regions of urbanization. Environ Sci Pollut Res Int 28(40):56139-56151. https://doi.org/10.1007/s11356-021-14596-w
5. Zhou Q (2021) RETRACTED ARTICLE: Climatic data analysis and computer data simulation of inland cities based on cloud computing and remote sensing images. Arab J Geosci 14: 1010. https://doi.org/10.1007/s12517-021-07275-0
6. Wu C, Zhang F, Xia J, Xu Y, Li G, Xie J, Du Z, Liu R (2021) Building Damage Detection Using U-Net with Attention Mechanism from Pre- and Post-Disaster Remote Sensing Datasets. Remote Sens 13: 905. https://doi.org/10.3390/rs13050905
7. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60: 90–110. https://doi.org/1023/B:VISI.0000029664.99615.94
8. Herbert B, Andreas E, Tinne T, Luc VG (2008) Speeded-up robust features (SURF). Computer Vision and Image Understanding 110(3): 346–359. https://doi.org/10.1016/j.cviu.2007.09.014
9. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In:IEEE Computer Society Conference on Computer Vision & Pattern Recognition. https://doi.org/10.1109/CVPR.2005.177
10. Melgani F, Bruzzone L (2004) Classification of hyperspectral remote sensing images with support vector machines. IEEE Transactions on Geoscience and Remote Sensing 42(8): 1778–1790. https://doi.org/10.1109/TGRS.2004.831865
11. Viola P, Jones MJ (2004) Robust Real-Time Face Detection. International Journal of Computer Vision 57: 137–154. https://doi.org/10.1023/B:VISI.0000013087.49260.fb
12. Soui M, Mansouri N, Alhamad R, Kessentini M, Ghedira K (2021) NSGA-II as feature selection technique and AdaBoost classifier for COVID-19 prediction using patient's symptoms. Nonlinear

Dynamics 106: 1453–1475. https://doi.org/10.1007/s11071-021-06504-1

13. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521: 436–444. https://doi.org/10.1038/nature14539
14. Girshick R, Donahue J, Darrell T (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In:IEEE Conference on Computer Vision and Pattern Recognition (CVPR).https://doi.org/10.1109/CVPR.2014.81
15. He K, Zhang X, Ren S, Sun J (2014) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. In: Computer Vision – ECCV 2014. https://doi.org/10.1007/978-3-319-10578-9_23
16. Girshick, R (2015) Fast R-CNN. In: IEEE International Conference on Computer Vision ( ICCV). https://ui.adsabs.harvard.edu/abs/2015arXiv150408083G
17. Ren S, He K, Girshick, R (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. 39:1137–1149.https://doi.org/10.1109/TPAMI.2016.2577031
18. Kaiming H, Georgia G, Piotr D, Ross G (2017) Mask R-CNN. In: IEEE international conference on computer vision (ICCV), pp 2961–2969. https://doi.org/10.1109/ICCV.2017.322
19. Redmon J, Divvala S, Girshick R (2016) You Only Look Once: Unified, Real-Time Object Detection. In:IEEE International Conference on Computer Vision (ICCV). https://doi.org/10.1109/CVPR.2016.91
20. Redmon J, Farhadi A (2017) YOLO9000: Better, Faster, Stronger. In:IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2017.690
21. Redmon J (2018) YOLOv3: An Incremental Improvement. https://arxiv.org/abs/180-4.02767. Accessed 8 April 2018
22. Bochkovskiy A, Wang CY, Liao HYM (2020) Yolov4: Optimal speed and accuracy of object detection. arXiv, arXiv:2004.10934.
23. Ultralytics (2020) yolov5. https://github.com/ultralytics/yolov5. Accessed 18 May 2020
24. Wei L, Anguelov D et al (2016) SSD: Single Shot MultiBox Detector. In:European Conference on Computer Vision. https://doi.org/10.1007/978-3-319-46448-0_2
25. Zhou X, Wang D, Krähenbühl P (2019) Objects as Points. CoRR. abs/1904.07850
26. Tan M, Pang R, Le Q (2020) EfficientDet: Scalable and Efficient Object Detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR42600.2020.01079
27. Chen J, Sun J, Li Y, Hou C (2021) Object detection in remote sensing images based on deep transfer learning. Multimed Tools Appl. https://doi.org/10.1007/s11042-021-10833-z
28. Han Q, Yin Q, Zheng X, Chen Z (2021) Remote sensing image building detection method based on Mask R-CNN. Complex & Intelligent Systems. https://doi.org/10.1007/s40747-021-00322-z
29. Li Y, Mao H, Liu R, Pei X, Shang R (2021) A Lightweight Keypoint-Based Oriented Object Detection of Remote Sensing Images. Remote Sensing 13(13): 2459.https://doi.org/10.3390/rs13132459

30. Xu D, Wu Y (2020) Improved YOLO-V3 with DenseNet for Multi-Scale Remote Sensing Target Detection. Sensors 20(15): 4276. https://doi.org/10.3390/s20154276
31. Zhou L, Yan H, Shan Y, Zheng C, Liu Yang, Zuo X, Qiao B, Li Y (2021) Aircraft Detection for Remote Sensing Images Based on Deep Convolutional Neural Networks. Journal of Electrical and Computer Engineering 2021:4685644. https://doi.org/10.1155/2021/4685644
32. Lu Q (2021) An Improved Object Detection Algorithm Based on SSD in Remote Sensing Image. Computer Science and Application 11(05):1579-1587. https://doi.org/10.12677/CSA.2021.115163
33. Wang C, Liao HYM, Wu Y et al (2020) CSPNet: A New Backbone that can Enhance Learning Capability of CNN. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). https://doi.org/10.1109/CVPRW50498.2020.00203
34. Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I, Savarese S (2019) Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2019.00075
35. Yu J, Jiang Y, Wang Z, Cao Z, Huang T (2016) UnitBox: An Advanced Object Detection Network. Proceedings of the 24th ACM international conference on Multimedia. https://doi.org/10.1145/2964284.2967274
36. Elfwing S, Uchibe E, Doya K (2017) Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Neural Networks 107: 3-11. https://doi.org/10.1016/j.neunet.2017.12.012
37. Jie H□Li S□Gang S, Albanie S (2020): Squeeze-and-Excitation Networks. IEEE Trans Pattern Anal Mach Intell 42(8):2011-2023.https://doi.org/10.1109/TPAMI.2019.2913372
38. Lin T, Dollár P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. in: IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2017.106
39. Li K, Wan G, Cheng G et al (2020) Object detection in optical remote sensing images: A surveyand a new benchmark. ISPRS Journal of Photogrammetry and Remote Sensing 159:296-307.http://dx.doi.org/10.1016/j.isprsjprs.2019.11.023
40. Long Y, Gong Y, Xiao Z, Liu Q (2017) Accurate Object Localization in Remote Sensing Images Based on Convolutional Neural Networks. In: IEEE Transactions on Geoscience and Remote Sensing. http://dx.doi.org/10.1109/TGRS.2016.2645610
41. Cheng G, Han J, Zhou P, Lei G (2014) Multi-class geospatial object detection and geographic image classification based on collection of part detectors. ISPRS Journal of Photogrammetry and Remote Sensing 98(dec.):119-132.

# Tables

Table 1
Performance indicators of YOLOv5s and improved YOLOv5s on DIOR dataset

| Model | Precision | Recall | mAP@.5 | mAP@.5:.95 |
|---|---|---|---|---|
| YOLOv5s | 0.85 | 0.762 | 0.785 | 0.563 |
| improved YOLOv5s | 0.894 | 0.753 | 0.805 | 0.574 |

Table 2
Comparison of experimental results of different algorithms

| | YOLOv3 | YOLOv4 | YOLOv5s | EfficientDet | improved YOLOv5s |
|---|---|---|---|---|---|
| Airplane | 84.96% | 90.43% | 94.7% | 52.61% | 95.2% |
| Airport | 44.05% | 52.35% | 75.4% | 71.12% | 84.2% |
| Baseballfield | 89.01% | 91.38% | 93.5% | 68.14% | 94.8% |
| Basket-ballcourt | 70.64% | 77.1% | 85.7% | 81.53% | 85.2% |
| Bridge | 32.88% | 25.54% | 53.4% | 26.45% | 54% |
| Chimney | 61.95% | 87.35% | 90.1% | 73.30% | 90.5% |
| Dam | 28.92% | 45.54% | 61.5% | 63.20% | 71% |
| Expressway-service-area | 48.49% | 59.96% | 71% | 72.52% | 75.3% |
| Expressway-toll-station | 55.47% | 54.21% | 70.9% | 50.69% | 70.7% |
| Golffield | 39.47% | 48.43% | 73.2% | 77.29% | 82% |
| Ground-trackfield | 53.92% | 65.04% | 80.6% | 73.19% | 82.1% |
| Harbor | 46.77% | 50.23% | 66.8% | 46.55% | 70.6% |
| Overpass | 53.33% | 48.96% | 64.4% | 49.26% | 67.3% |
| Ship | 93.73% | 85.09% | 95% | 29.23% | 95% |
| Stadium | 92.23% | 78.09% | 92.7% | 65.26% | 94.3% |
| Storagetank | 78.35% | 79.82% | 86% | 25.82% | 83.8% |
| Tenniscourk | 88.65% | 88.77% | 92.6% | 79.29% | 91.6% |
| Trainstation | 19.15% | 22.22% | 56.2% | 47.39% | 61.2% |
| Vehicle | 65.28% | 63.61% | 81.5% | 15.84% | 79.8% |
| Windmill | 71.26% | 56.98% | 84% | 56.94% | 81.8% |
| mAP@.5 | 60.92% | 63.56% | 78.5% | 56.28% | 80.5% |

# Figures

**Figure 1**

YOLOv5 network

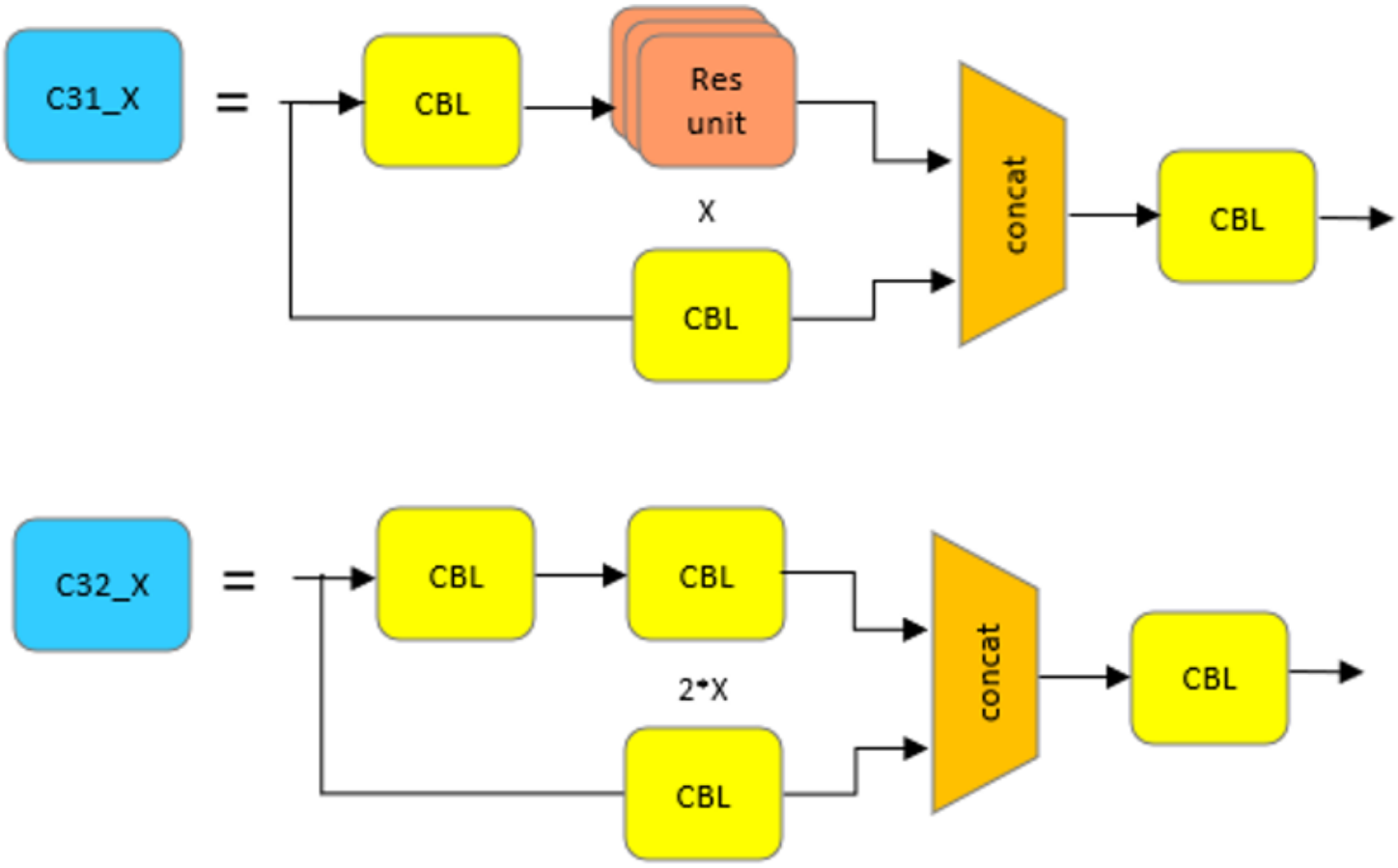
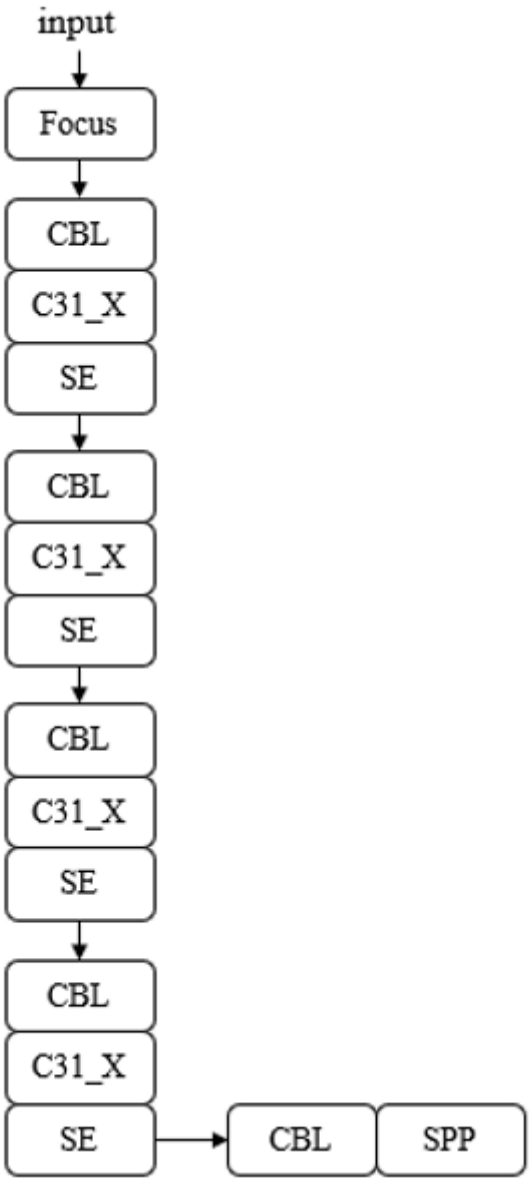



**Figure 2**

C3 structure




## Figure 3

Improved backbone structure

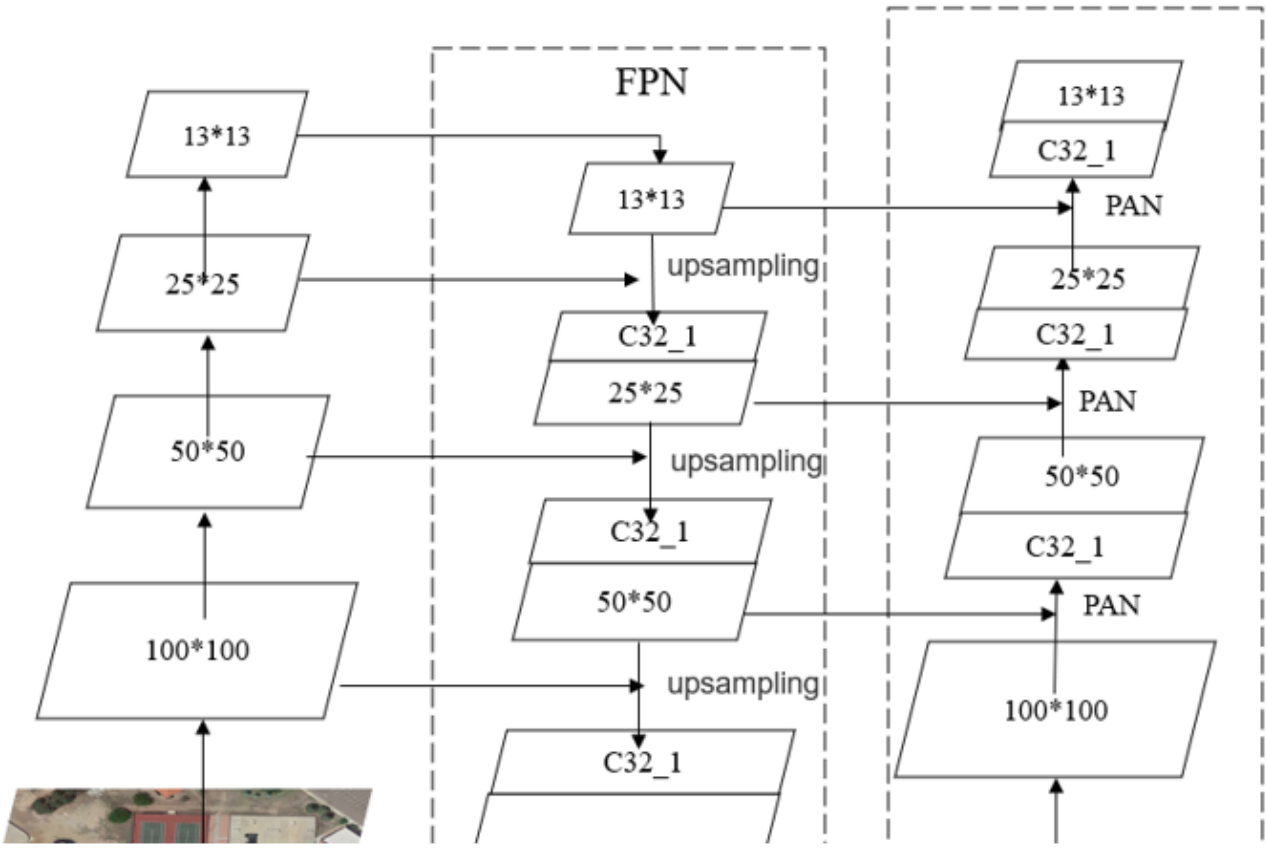



**Figure 4**

Improved FPN+PAN structure

## Figure 5

DIOR dataset instance

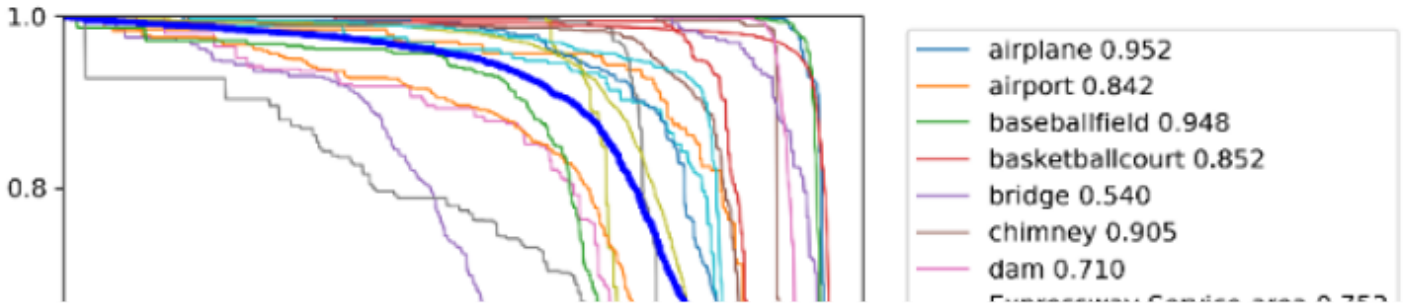



**Figure 6**

P-R curve


## Figure 7

Comparison of detection results between YOLOv5s and improved YOLOv5s