Improved YOLOv5 network method for remote sensing image-based ground objects recognition

Jie Xue¹ · Yongguo Zheng¹ · Changlei Dong-Ye¹ · Ping Wang² · Muhammad Yasir³

Accepted: 3 April 2022 / Published online: 19 May 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
High-resolution remote sensing images have the characteristics of complex background environment, clustering of objects, etc., the complex background makes the remote sensing image contain a large number of irrelevant ground objects with a high similarity or overlap, which makes the edge and texture of the objects not clear enough, and this leads to low recognition accuracy of ground objects such as airports, dams, and golf field, although the size of this object is large. Based on this problem, this paper proposes a remote sensing image object detection method based on the YOLOv5 network. By improving the backbone extraction network, the network structure can be deepened to get more information about large objects, and the detection effect can be improved by adding an attention mechanism and adding an 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 the 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 dataset, and the mAP of the improved YOLOv5 network is 80.5%, which is 2% higher than the original YOLOv5 network.

Keywords High-resolution remote sensing image · Object recognition · YOLOv5 · Attention mechanism

1 Introduction
With the development of remote sensing technology, remote sensing has been widely used in modern industry and life (Wei and Liu 2021), and the object recognition of remote sensing images has been attached importance. At present, the object recognition of remote sensing images is widely used in many fields, such as military reconnaissance (Liu et al. 2021), environmental monitoring (Zhang et al. 2021), urban construction (Guo et al. 2021), climatic data analysis (Zhou 2021), and building damage detection (Wu et al. 2021). 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) (Lowe 2004), speeded up robust features (SURF) (Herbert et al. 2008), histogram of oriented gradient (HOG) (Dalal and Triggs 2005), etc., and the common classifiers include support vector machines (SVM) (Melgani and Bruzzone 2004), Haar (Viola and Jones 2004), AdaBoost (Soui et al. 2021), etc. Remote sensing images have the characteristics of dense ground objects and complex environments, 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 (CNNs) (LeCun et al. 2015) 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, divide object detection into two phases: generating region proposals, classifying region proposals in classifier and correcting positions, such as R-CNN (Girshick et al. 2014), SPP-Net (He et al. 2014), Fast R-CNN (Girshick 2015), Faster R-CNN (Ren et al. 2017), and Mast R-CNN (Kaiming et al. 2017). The latter directly regresses the predicted object to generate the bounding box, such as You Only Look Once (YOLO) (Redmon et al. 2016; Redmon and Farhadi 2017; Redmon 2018; Bochkovskiy et al. 2020; Ultralytics 2020, Liu et al. 2022), Single Shot MultiBox Detector (SSD) (Wei et al. 2016), and CenterNet (Zhou et al. 2019), EfficientDet (Tan et al. 2020).

According to the problems of remote sensing images, researchers have made unremitting efforts based on CNN. Chen et al. (2021) proposed Domain Adaptation Faster R-CNN algorithm to improve the robustness of the model and widen the scope of application and proved its effectiveness. Han et al. (2021) 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 et al. (2021) proposed a lightweight keypoints-based oriented object detector for remote sensing images in view of the complex background. Xu et al. (2020) 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 et al. (2021) 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 (2021) 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 the Laplace-NMS method, which had a good post-treatment effect on dense objects. Lots of research 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 a complex environment. Therefore, the purpose of this paper is to propose an algorithm for large objects detection.

In the relatively complex environment, some large objects in remote sensing images have similar colors to the background and unclear textures, which makes it difficult to extract their features, resulting in low accuracy of target recognition, and this paper presents a method of remote sensing image ground object recognition based on YOLOv5 (Ultralytics 2020). This paper improves the YOLOv5s network structure by adding a set of CSP structure and attention mechanisms 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

YOLOv5 network is the latest research achievement of YOLO series algorithms. Its network structure is similar to that of the YOLOv4 network; however, it is smaller than the YOLOv4 network, has faster convergence speed and running speed, and belongs to the lightweight algorithm. At the same time, it also improves accuracy. Therefore, this paper adopts the YOLOv5 algorithm for target detection of remote sensing images.

The YOLOv5 network structure consists of four parts: input, backbone, neck, and prediction. The network structure is shown in Fig. 1. There are four kinds of networks, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. They have a similar network structure with different depths and widths. The YOLOv5s network has the smallest structure 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, and 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 Fig. 1.1. YOLOv5 uses two CSP (Wang et al. 2020) structures: CSP 1_ X and CSP 2_ X; CSP1_ X is used for down-sampling in backbone, and 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 Fig. 1.1; neck adopts SPP-net and FPN + PAN structure to enhance the feature fusion effect of the network; prediction adopts GIOU _ Loss (Rezatofghii et al. 2019), and GIOU not only pays 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 (Yu et al. 2016) while maintaining the advantages of IOU. The calculation Eqs. (1) and (2) are as follows:

\[
\text{IOU} = \frac{|A \cap B|}{|A \cup B|}
\]

(1)

\[
\text{GIOU} = \text{IOU} - \frac{|C|}{|A \cup B|}
\]

(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() (Elfwing et al. 2017), deleted the conv in the CSP, and renamed it C3 as shown in Fig. 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 the DIOR dataset, the complex background environment leads to the unsatisfactory detection effect of some large ground objects. In this paper, a group of C3 structures 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, and in other words, the attention mechanism enables the network to find significant information among multitudinous information. In this paper, the attention mechanism (Jie et al. 2020) 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 Fig. 3.

3.3 Neck improvement

Neck adopts the FPN (Lin et al. 2017) + PAN structure. This structure adds a bottom-up feature pyramid network after FPN, which enhances the semantic expression and...
location information on multiple scales. The neck of YOLOv5s integrates the C32_X structure to enhance the feature fusion effect of the network. Since a group of C3 structures is added in this paper, an output layer is added in the neck to further improve the feature extraction of the network. The improved FPN + PAN structure is shown in Fig. 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 was 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 Experimental results and analysis

In this paper, the DIOR dataset (Li et al. 2020) is used to verify the detection effect of the improved YOLOv5s network; the precision, recall, mAP@0.5, and mAP@0.5:0.95 are used as a performance indicator of the evaluation 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 23,463 images and 190,288 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 Fig. 5.

4.2 Evaluation index

Precision represents the proportion of the predicted correct objects that are actually the correct objects, Eq. (3); recall represents the proportion of correct objects in the prediction results to the number of actually correct objects, Eq. (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 the x-axis and y-axis is called the average precision (AP), Eq. (5). The P-R curve of the improved YOLOv5s under the DIOR dataset is shown in Fig. 6. mAP is the mean average precision of multi-object, Eq. (6). mAP@0.5 indicates the mAP of all categories when the IOU is 0.5. mAP@0.5:0.95 refers to the average mAP of IOU between 0.5 and 0.95.

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (3)
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (4)
\]

\[
\text{AP} = \int_0^1 P(R) \, dr \quad (5)
\]

\[
mAP = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i \quad (6)
\]

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 ith object.

4.3 Training and result analysis

4.3.1 Network training

The model training experimental environment of this paper is as follows: 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, the batch size is 16, and epochs is 200.

Firstly, transfer learning is adopted. The overpass in RSOD dataset (Long et al. 2017) and NWPU VHR-10
Fig. 4 Improved FPN + PAN structure

Fig. 5 DIOR dataset instance
dataset (Cheng et al. 2014) are combined into a new dataset, 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), and (476,590).

### 4.3.2 Result in analysis

In this paper, YOLOv5s are compared with the improved YOLOv5s on the DIOR dataset, and precision, recall, mAP@0.5, mAP@0.5:0.95 as performance indicators to evaluate the quality of the network as shown in Table 1. According to the table, the improved YOLOv5s are improved in precision, mAP@0.5, mAP@0.5:0.95 compared with original YOLOv5s, and the recall is not as effective as original YOLOv5s.

![Fig. 6 Precision and recall curve. a Dense objects. b Complex background. c The object is similar to the background](image)

The test results are shown in Fig. 7. In each group of images, the upper row is the improved YOLOv5s test images, and the lower row is the original YOLOv5s test images. It can be seen from images in the group (a) of figures that the ship objects are small, densely distributed, and cover the larger object harbors, which interferes with the recognition of the harbors. The modified YOLOv5s can identify the harbors more accurately than the original YOLOv5s; in group (b) of figures, under the complex background, it is difficult to extract the features of the train station. The improved YOLOv5s can successfully identify the train station. The similarity between the two targets of bridge and overpass is high, and the original YOLOv5s recognition category is wrong, while the improved YOLOv5s can correctly identify the bridge with higher accuracy. The ground track field is inconsistent with other target sizes, and the improved YOLOv5s effectively improve the detection accuracy; in the group (c) of figures, the overpass, golf field, and airport are integrated with the background environment, and the object edge is not clear, which increases the difficulty of object recognition, and the accuracy of the improved YOLOv5s is significantly improved.

In order to further verify the effectiveness of the improved YOLOv5s, in this paper, the improved YOLOv5s are compared with the YOLO series network and

| Model        | Precision | Recall | mAP@.5 | mAP@.5:0.95 |
|--------------|-----------|--------|--------|-------------|
| YOLOv5s      | 0.85      | 0.762  | 0.785  | 0.563       |
| Improved YOLOv5s | 0.894    | 0.753  | 0.805  | 0.574       |
Fig. 7 Comparison of detection results between YOLOv5s and improved YOLOv5s.

(a) Dense objects

(b) Complex background

(c) The object is similar to the background
EfficientDet model on the DIOR dataset, and AP of 20 types of ground objects and mAP is 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 the complex environment has been greatly improved, such as airports, dam, golf fields, 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 the YOLO series network.

## 5 Conclusion

Aiming at the problems existing in the recognition of large ground objects in remote sensing images, such as in complex background, the object edge is not clear and the object feature extraction is difficult, resulting in poor target detection effect, and this paper proposes a remote sensing image ground objects recognition method based on the YOLOv5 network. This paper uses the DIOR dataset; firstly, the anchor box is recalculated, then C3 and attention mechanism are added to the backbone to extract more features of large ground objects by deepening the network structure, and finally, output layer is added to the neck to enhance feature fusion, which improves the effect of the 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 a complex environment, and its mAP reaches 80.5% which is 2% higher than the original YOLOv5s network. In the next step, based on maintaining the existing advantages, we will further carry out research on improving the accuracy of small ground objects recognition.

### Table 2 Comparison of experimental results of different algorithms

| Ground Object             | 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                 |
| Baseball field            | 89.01      | 91.38      | 93.5        | 68.14            | 94.8                 |
| Basketball court          | 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                 |
| Golf field                | 39.47      | 48.43      | 73.2        | 77.29            | 82                   |
| Ground track field        | 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                 |
| Storage tank              | 78.35      | 79.82      | 86          | 25.82            | 83.8                 |
| Tennis court              | 88.65      | 88.77      | 92.6        | 79.29            | 91.6                 |
| Train station             | 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                 |

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

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.
Data availability. The RSOD datasets generated during and/or analyzed 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 analyzed during the current study are not publicly available due to the link failure but are available from the corresponding author on reasonable request.

Declarations

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

References

Bochkovskiy A, Wang CY, Liao HYM (2020) Yolov4: Optimal speed and accuracy of object detection. arXiv, https://arxiv.org/abs/2004.10934.

Chen J, Sun J, Li Y, Hou C (2021) Object detection in remote sensing images based on deep transfer learning. Multimed Tools Appl 81:12093–12109. https://doi.org/10.1007/s1142-021-10833-z

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 J Photogram Remote Sens 98:119–132

Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE Computer society conference on computer vision and pattern recognition, pp.886–893. https://doi.org/10.1109/CVPR.2005.177.

Elfwing S, Uchibe E, Doya K (2017) Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Neural Netw 107:3–11. https://doi.org/10.1016/j.neunet.2017.12.012

Girshick, R (2015) Fast R-CNN. arXiv e-prints. arXiv:1504.08083 https://ui.adsabs.harvard.edu/abs/2015arXiv150408083G.

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), pp. 580–587. https://doi.org/10.1109/CVPR.2014.81.

Guo M, Shi 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/s40747-021-0322-z

Han Q, Yin Q, Zheng X, Chen Z (2021) Remote sensing image building detection method based on Mask R-CNN. Complex Intell Syst. https://doi.org/10.1007/s40747-021-00322-z

He K, Zhang X, Ren S, Sun J (2014) Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell 37(9):1904–1916. https://doi.org/10.1109/TGRS.2016.2577031

He K, Zhang X, Ren S, Sun J (2014) Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell 37(9):1904–1916. https://doi.org/10.1109/TGRS.2016.2577031

Herbert B, Andreas E, Tinne T, Luc VG (2008) Speeded-up robust features (SURF). Comput Vis Image Underst 110(3):346–359. https://doi.org/10.1016/j.cviu.2007.09.014

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

Kaiming H, Georgia G, Plot D, Ross G (2017) Mask R-CNN. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TPAMI.2017.322

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539

Li K, Wan G, Cheng G et al (2020) Object detection in optical remote sensing images: a survey and a new benchmark. ISPRS J Photogram Remote Sens 159:296–307. https://doi.org/10.1016/j.isprsjprs.2019.11.023

Li Y, Miao H, Liu R, Pei X, Shang R (2021) A lightweight keypoint-based oriented object detection of remote sensing images. Remote Sens 13(13):2459. https://doi.org/10.3390/rs13132459

Lin T, Dollar P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. IEEE Comput Soc. https://doi.org/10.1109/CVPR.2017.106

Liu F, Zhu J, Wang W (2021) Surface-to-air missile sites detection agent with remote sensing images. Sci China Inf Sci. https://doi.org/10.1007/s11432-019-9920-2.

Liu S, Kong W, Chen X, Xu M, Yasir M, Zhao L, Li J (2022) Multi-scale ship detection algorithm based on a lightweight neural network for spaceborne SAR images. Remote Sens 14(5):1149

Long Y, Gong Y, Xiao Z, Liu Q (2017) Accurate object localization in remote sensing images based on convolutional neural networks. IEEE Trans Geosci Remote Sens 55(5):2486–2498. https://doi.org/10.1109/TGRS.2016.2645610

Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vision 60:99–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Lu Q (2021) An improved object detection algorithm based on SSD in remote sensing image. Comput Sci Appl 11(05):1579–1587. https://doi.org/10.12677/CSA.2021.115163

Meliani F, Bruzzone L (2004) Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans Geosci Remote Sens 42(8):1778–1790. https://doi.org/10.1109/TGRS.2004.831865

Redmon J (2018) YOLOv3: An Incremental Improvement. Arxiv. https://arxiv.org/abs/1804.02767. Accessed 8 April 2018.

Redmon J, Divvala S, Girshick R (2016) You Only Look once: unified, real-time object detection. In: IEEE International conference on computer vision (ICCV), pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.

Redmon J, Farhadi A (2017) YOLO9000: better, faster, stronger. IEEE Conf Comput Vision Pattern Recogn. https://doi.org/10.1109/CVPR.2017.690

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

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), pp. 658–666 https://doi.org/10.1109/CVPR.2019.00075.

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

Tan M, Pang R, Le Q (2020) EfficientDet: scalable and efficient object detection. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 10778–10787. https://doi.org/10.1109/CVPR.2020.01079.

Ultralytics (2020) yolov5. Github. https://github.com/ultralytics/yolov5. Accessed 18 May 2020.

Viola P, Jones MJ (2004) Robust real-time face detection. Int J Comput Vision 57:137–154. https://doi.org/10.1023/B:VISI.0000013087.49260.fb

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
worksheets (CVPRW), pp. 1571–1580. https://doi.org/10.1109/CVPRW50498.2020.00203.

Wei Z, Liu Y (2021) Construction of super-resolution model of remote sensing image based on deep convolutional neural network. Comput Commun 178:191–200. https://doi.org/10.1016/j.comcom.2021.06.022

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.

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

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

Yu J, Jiang Y, Wang Z, Cao Z, Huang T (2016) UnitBox: an advanced object detection network. ACM. https://doi.org/10.1145/2964284.2967274

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

Zhou Q (2021) 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

Zhou L, Yan H, Shan Y, Zheng C, Yang L, Zuo X, Qiao B, Li Y (2021) Aircraft detection for remote sensing images based on deep convolutional neural networks. J Electr Comput Eng 2021:4685644. https://doi.org/10.1155/2021/4685644

Zhou X, Wang D, Krahenbühl P (2019) Objects as points. CoRR. https://arxiv.org/abs/1904.07850.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.