A Small Object Detection Research Based on Dynamic Convolution Neural Network

Chenshuai Bai

Lanzhou Jiaotong University

Kaijun Wu (✉ wkj@mail.lzjtu.cn)

Lanzhou Jiaotong University <https://orcid.org/0000-0002-8939-1874>

Dicong Wang

Lanzhou Jiaotong University

Mingjun Yan

Lanzhou Jiaotong University

Research

Keywords: dynamic convolution neural network, optimization parameters adjustment, Average Precision, Log Average, Miss Rate

Posted Date: December 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1116930/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

A Small Object Detection Research Based on Dynamic Convolution Neural Network

Chenshuai Bai¹, Kaijun Wu^{1*}, Dicong Wang^{1,2}, Mingjun Yan¹

Abstract: In view of the fact that the detection effect of EfficientNet-YOLOv3 object detection algorithm is not very good, this paper proposes a small object detection research based on dynamic convolution neural network. Firstly, the dynamic convolutional neural network is introduced to replace the traditional, which makes the algorithm model more robust; secondly, the optimization parameters are continuously adjusted in the training process to further strengthen the model structure; finally, the Learning Rate and Batch Size parameters are modified during the training process in order to prevent overfitting. In order to verify the effectiveness of the proposed algorithm, RSOD and TGRS-HRRSD remote sensing image data sets are used to test the effect. The results of the proposed algorithm on RSOD remote sensing image data sets show that compared with the original EfficientNet-YOLOv3 algorithm, the mean Average Precision (mAP) value is increased by 1.93% and the mean Log Average Miss Rate (mLAMR) value is reduced by 0.0500; The results of the proposed algorithm on TGRS-HRRSD remote sensing image data set show that compared with the original EfficientNet-YOLOv3 algorithm, the mAP value is increased by 0.07% and the mLAMR value is reduced by 0.0007.

Keywords: dynamic convolution neural network; optimization parameters adjustment; Average Precision; Log Average Miss Rate

*Corresponding author : Kaijun Wu(E-mail: wkj@mail.lzjtu.cn)

1 Introduction

The emergence of deep learning technology promotes the development of computer vision. And object detection based on remote sensing image is one of the important directions in the field of computer vision, which can be applied to urban planning, traffic safety and environmental monitoring^[1-5]. The object detection algorithm of remote sensing image comes from the object detection algorithm in the general field, which is a series of algorithms combined with the characteristics of multi-scale, strong density and shape difference of remote sensing image. The traditional remote sensing image object detection algorithm mainly adopts multi-step strategy, such as preprocessing, image segmentation, extracting regions of interest and object detection^[6-7]. Each step is independent and is the corresponding algorithm designed for specific problems, which has the problems of process solidification, insufficient automation and intelligence^[8]. Therefore, the performance of object detection algorithm in remote sensing image is greatly improved by deep learning technology due to its strong feature representation ability and end-to-end learning ability.

Nowadays, remote sensing image is a data set with small object and complex background, and there are huge problems in the research of remote sensing image object detection. Although many object detection algorithms have relatively good performance on ordinary data sets, the detection accuracy of object detection in remote sensing images is not very good. Chen et al.^[9] proposed a detection algorithm based on multi-classification learning, which realizes the accurate detection of ships by finding the minimum circumscribed rectangle of the object; Ding et al.^[10] proposed the region of interest transformer (ROI trans) to solve the remote sensing object detection with significant directivity; Xu et al.^[11] proposed the gliding vertex to characterize the directed bounding box by first detecting the horizontal bounding box and then learning the offset of the four corners of the horizontal bounding box; Yang et al.^[12] proposed dense rotating small object detector: by designing sampling fusion network, multi-layer features are fused into effective anchor sampling to improve the sensitivity to small objects; at the same time, by suppressing noise and highlighting object characteristics, a supervised pixel attention network and channel attention network are designed; finally, the Intersection Over Union (IOU) factor is added to smooth L1 loss to solve the problem of inaccurate boundary of rotating bounding box; aiming at the problems of large aspect ratio and category imbalance of object instances in remote sensing images, Lin et al.^[13] proposed a fine tuned single-stage detector, which realizes feature reconstruction and alignment by designing feature refinement module, and realizes object detection by stepwise regression from coarse-grained to fine-grained; Yao et al.^[14] proposed an object detection algorithm of remote sensing image based on multi-resolution feature fusion; Wang et al.^[15] proposed the method of combining multi-level features for regional proposal; Yao et al.^[16] proposed a multi-scale convolutional neural

network remote sensing object detection framework; Li et al. ^[17] proposed an object detection algorithm for visual perception of high-resolution remote sensing images; Chen et al. ^[18] proposed a fast ship object detection method in large scene remote sensing images based on cascaded convolutional neural network; in view of the problems such as poor object performance and slow inference speed of remote sensing images with dense permutation and significant direction detected by RoI Trans, Zhao et al. ^[19] adopted the method of progressively enhancing the positioning accuracy of rotating candidate frame and non-local enhancement of features to improve the object detection performance of remote sensing images; Cai et al. ^[20] proposed a new progressive shrinkage algorithm, a generalized pruning method, which reduces the model size in more dimensions (depth, width, core size and resolution) than pruning; Chen et al. ^[21] proposed an algorithm of dynamic convolution neural network; Liu et al. ^[22] proposed a new strategy to assemble multiple identical backbones through the composite connection between adjacent backbones to form a more powerful backbone, which is called composite backbone network; Hua et al. ^[23] proposed a cascade convolutional neural network real-time object detection framework combining visual perception and convolutional memory network reasoning; Li et al. ^[24] proposed an end-to-end depth network based on the shape of network structure, which detects salient objects in optical images in a pure data-driven manner; Zhang et al. ^[25] proposed a scale self-adaptive network scheme to improve the accuracy of multi-object detection; Zhu et al. ^[26] proposed a new object detection method based on multi-scale SELU-DenseNet and Dynamic Anchor Assignment (DAA) strategy; Yao et al. ^[27] proposed an online latent semantic scatter method to solve this problem; Pang ^[28] proposed a unified self-enhancement network R^2 -CNN: a convolutional neural network based on remote sensing region, which is composed of backbone microgrid, intermediate global attention block, final classifier and detector; Chen et al. ^[29] proposed a new object heat map network to solve these problems; Yao et al. ^[30] proposed a new calculation model for airport detection using optical Remote Sensing Image (RSI); Yu et al. ^[31] proposed a method Chan-Vese (CV) based on image decomposition and distance regularization model; Li et al. ^[32] proposed the remaining problems and future development trend of ship detection and classification methods based on optical remote sensing images; Youme et al. ^[33] proposed to divide the image into four regions, mark the image through the region of interest, and extract image features by single-stage detection algorithm SSD to improve the accuracy of small object detection.

This paper improves the EfficientNet-YOLOv3 algorithm ^[34] whose backbone feature extraction network is EfficientNet ^[35]. Firstly, based on the idea of dynamic convolution neural network proposed by Chen et al. ^[21], replace the two-dimensional convolution neural network of the backbone feature extraction network by the dynamic convolution neural network, in which dynamic convolution kernel dynamically aggregates multiple parallel convolution kernels related to input, improves the backbone feature extraction network and strengthens the feature extraction ability of the backbone network for remote sensing images; secondly, adjust and optimize the parameters to improve the feature extraction ability of small and medium-sized objects in remote sensing images; finally, modify the learning rate and batch size to prevent overfitting during the training. The validity of the proposed algorithm is verified by using RSOD and TGRS-HRRSD remote sensing image data sets.

2 Related works

2.1 EfficientNet backbone feature extraction network

EfficientNet backbone feature extraction network is a research under the same computing power constraints. The influence of network depth, width and resolution on the same operation type network is explored to find the optimal configuration proportion parameters. The algorithm process of backbone feature extraction network EfficientNet is as follows: firstly, the image needs to be transformed into the input dimension required by the Mobile Bottleneck Convolution (MBConv) module through the first Conv $3 * 3$ convolution layer, and the input remote sensing image needs to be extracted from the feature map through a series of MBConv modules; secondly, the parameters of each MBConv module are finely adjusted to adapt to the current use environment, thus, the combined scale optimization method can make the network obtain a better receptive field; finally, using the feature map self-adaptive connection method based on Full Convolutional Neural Network (FCNN), the Conv $1 * 1$ convolution network will be able to adapt to the feature maps of different sizes and unify them into the dimensions required by the algorithm. Then, the image classification, recognition and detection are completed through the output feature map. The structure of EfficientNet backbone feature network model is shown in Fig. 1:

Fig.1 EfficientNet model structure

2.2 YOLOv3 model structure

The model structure of YOLOv3 has no pooling layer and full connection layer. The size transformation of the tensor in the forward propagation process of YOLOv3 model is realized by changing the step size of the convolution kernel: the backbone feature extraction network reduces the output feature map to 1/32 of the input. YOLOv3 network structure is a full convolution network, which uses a large number of layer hopping connections of residuals rather than the pooling layer to reduce the negative gradient effect caused by pooling. And it uses the convolution step size 2 to realize downsampling. In order to enhance the accuracy of the algorithm for small object detection, YOLOv3 adopts the up sampling and fusion method similar to Feature Pyramid Networks (FPN): the three prediction branches in the YOLOv3 model structure adopt the full convolution structure, with a total of 9 anchor boxes and 3 outputs, and each output uses 3 anchors, so 3 boxes are predicted for each output position. The model structure of YOLOv3 is shown in Fig.2:

Fig. 2 YOLOv3 model structure

3 Methodology

3.1 research objective

Because there are some differences between remote sensing image and natural scene image, and their imaging methods are different, remote sensing image has its own unique characteristics, such as high resolution, small and densely arranged targets, complex background and so on. Therefore, it is necessary to further improve the migration method, so as to solve the difficult problems in remote sensing image detection, and overcoming these key detection difficulties has become the key research

direction of many scholars in this field. How to apply deep learning to the field of remote sensing technology and improve the performance of remote sensing detection has important research value and practical significance.

3.2 Optimization

Firstly, the dynamic convolutional neural network is introduced to replace the traditional, which makes the algorithm model more robust; secondly, the optimization parameters are continuously adjusted in the training process to further strengthen the model structure; finally, the Learning Rate and Batch Size parameters are modified during the training process in order to prevent overfitting.

3.3 Dynamic convolutional neural network

As shown in Fig.3, the goal of dynamic convolutional neural network is to provide a better balance between network performance and computational burden in the range of efficient neural network. Dynamic convolution neural network, which does not increase the depth or width of the network, but improves the performance of the model by paying attention to aggregating multiple convolution cores. It has convolution cores sharing the same kernel size, input and output dimensions and aggregates by using the attention weight π_k , after which the batch normalization and activation function are used to construct the dynamic convolution layer. Because the convolution kernel is very small, so the increase in the output dimension of each layer is within a controllable range. Therefore, traditional convolutional neural network can be easily replaced by dynamic convolutional neural network. In the traditional convolution neural network, the parameters of any input are fixed, while the dynamic convolution makes the convolution parameters of each layer change with the input during reasoning. For a feature map generated in the convolution process, first operate the feature map to generate K π_k parameters with a sum of 1. Then the K convolution kernel parameters are linearly summed. In this way, the convolution changes with the change of input while reasoning.

Fig. 3 Dynamic convolutional neural network

3.4 Improved YOLOv3 model structure

As shown in Fig.4, firstly, the input image size of the backbone feature extraction network EfficientNet is modified from 416*416*3 to 800*800*3, the residual network is convoluted once with a convolution kernel size of 3*3 and a step size of 2, and the width and height of the input feature layer are compressed to obtain a feature layer; secondly, carry out a 1*1 convolution and a 3*3 convolution on the feature layer, and then add this result to the feature layer to form a residual structure; finally, the network is greatly deepened through the superposition of 1*1 convolution, 3*3 convolution and residual edge. The two-dimensional convolution neural network Conv2D uses the Dynamic convolution neural network Dynamic_Conv2D.

Fig. 4 Improved YOLOv3 algorithm model

4 Experimental results and discussion

In order to verify the effectiveness of the improved EfficientNet-YOLOv3 algorithm, the RSOD remote sensing image data set labeled by Wuhan University and the TGRS-HRRSD remote sensing image data set labeled by Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences are used. The experiment uses five comparison algorithms: object detection algorithm based on Faster-RCNN, object detection algorithm based on RetinaNet, object detection algorithm based on SSD single stage, object detection algorithm based on YOLOv4 and object detection algorithm based on the original EfficientNet-YOLOv3 to objectively analyze the performance of the object detection algorithm proposed in this paper. Besides, use the mAP and mLAMR as the evaluation indexes: the larger the mAP value is, the better the performance is; the smaller the mLAMR value is, the better the performance is.

4.1 Experimental environment configuration

The object detection algorithm proposed in this paper uses the built experimental environment platform model: the computer is configured with i5-8250CPU, 8GRAM, 64 bit Windows10 operating system; the server is configured with BSCC-N22, and the GPU queue is configured as follows: each machine is configured with 8 GPUs with NVIDIA Tesla V100-SXM 232GB video memory, each GPU card is allocated 8CPU core and 36GB memory by default, and the memory ratio is 1GPU card,8CPU core and 36GB memory^[36].

4.2 Experimental results of RSOD remote sensing image data set

4.2.1 RSOD remote sensing image data set

As shown in Fig. 5, the RSOD remote sensing image data set is an open data set, which is applied to object detection of remote sensing images. The remote sensing image data set samples include: Fig. 5 (a) aircraft, Fig. 5 (b) oiltank, Fig. 5 (c) overpass and Fig. 5 (d) playground. The dataset uses PASCAL_VOC format and includes four files, each of which is an object. There are 4993 aircrafts in 446 images, 191 amusement parks in 189 images, 180 overpasses in 176 images and 1586 oil barrels in 165 images^[36].

(a) aircraft

(b) oiltank

Fig. 5 Samples of RSOD remote sensing image data

4.2.2 mAP value is used as evaluation index

Table 1 Experimental results of mAP value of RSOD remote sensing image data set

	EfficientNet-YOLOv3 method	Faster-RCNN method	RetinaNet method	SSD method	YOLOv4 method	the method of this paper
Aircraft (AP value)	0.85	0.33	0.72	0.70	0.97	0.95
Overpass (AP value)	0.93	0.83	1.00	0.96	0.75	0.92
Playground (AP value)	1.00	0.99	1.00	1.00	0.93	1.00
Oiltank (AP value)	1.00	0.86	1.00	0.99	1.00	0.99
mAP value	0.9461	0.7536	0.9304	0.9122	0.9233	0.9654

4.2.3 mLAMR value is used as evaluation index

Table 2 Experimental results of mLAMR value of RSOD remote sensing image data set

	EfficientNet-YOLOv3 method	Faster-RCNN method	RetinaNet method	SSD method	YOLOv4 method	the method of this paper
Aircraft (LAMR value)	0.52	0.88	0.51	0.72	0.16	0.25
Overpass (LAMR value)	0.11	0.37	0.00	0.04	0.32	0.17
Playground (LAMR value)	0.00	0.00	0.00	0.00	0.07	0.00
Oiltank (LAMR value)	0.00	0.20	0.01	0.13	0.00	0.01
mLAMR value	0.1575	0.3625	0.1300	0.2225	0.1375	0.1075

4.3 Experimental results of TGRS-HRRSD remote sensing image data set

4.3.1 TGRS-HRRSD remote sensing image data set

As shown in Fig. 6, TGRS-HRRSD remote sensing image data set is produced by the optical image analysis and learning center of Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences for studying high-resolution remote sensing image object detection. The format of the data set is made into PASCAL_VOC format for application, the remote sensing image data samples include: Fig. 6 (a) airplane, Fig. 6 (b) bridge, Fig. 6 (c) crossroad, Fig. 6 (d) ship, Fig. 6 (e) vehicle, Fig. 6 (f) harbor, Fig. 6 (g) ground track field, Fig. 6 (h) storage tank, Fig. 6 (i) basketball court, Fig. 6 (j) parking lot, Fig. 6 (k) tennis court, Fig. 6 (l) baseball diamond and Fig. 6 (m) T junction, which are 13 categories, 21000 diagrams and 40000 target objects in total^[36].

Fig. 6 TGRS-HRRSD remote sensing image data sample diagram

4.3.2 mAP value is used as evaluation index

Table 3 Experimental results of mAP value of TGRS-HRRSD remote sensing image data set

	EfficientNet-YOL Ov3 method	Faster-RCNN method	RetinaNet method	SSD method	YOLOv4 method	the method of this paper
Ground track field (AP value)	0.99	0.94	1.00	0.99	0.97	0.98
Crossroad (AP value)	0.98	0.87	1.00	0.98	0.93	0.98
Airplane (AP value)	1.00	0.87	0.98	0.99	0.98	0.99
Baseball diamond (AP value)	0.99	0.77	0.98	0.95	0.87	0.85
Harbor (AP value)	0.99	0.79	0.98	0.97	0.77	0.99
Storage tank (AP value)	0.98	0.86	0.98	0.95	0.60	0.98
T junction (AP value)	0.96	0.30	0.97	0.92	0.81	0.92
Bridge (AP value)	0.98	0.75	0.95	0.97	0.88	1.00
Tennis court (AP value)	0.97	0.61	0.95	0.93	0.97	0.95
Ship (AP value)	0.98	0.74	0.94	0.93	0.94	0.99

Basketball court (AP value)	0.99	0.17	0.85	0.75	0.74	0.85
Vehicle (AP value)	0.96	0.26	0.80	0.32	0.96	0.99
Parking lot (AP value)	0.85	0.38	0.77	0.70	0.71	0.85
mAP value	0.9569	0.6381	0.9340	0.8722	0.8571	0.9576

4.3.3 mLAMR value is used as evaluation index

Table 4 Experimental results of mLAMR value of TGRS-HRRSD remote sensing image data set

	EfficientNet-YOLO v3 method	Faster-RCNN method	RetinaNet method	SSD method	YOLOv4 method	the method of this paper
Ground track field (LAMR value)	0.01	0.10	0.01	0.02	0.03	0.02
Crossroad (LAMR value)	0.03	0.29	0.01	0.06	0.11	0.04
Airplane (LAMR value)	0.01	0.30	0.03	0.05	0.03	0.01
Baseball diamond (LAMR value)	0.03	0.48	0.04	0.43	0.24	0.03
Harbor (LAMR value)	0.02	0.52	0.09	0.14	0.31	0.04
Storage tank (AP value)	0.10	0.29	0.11	0.28	0.59	0.08
T junction (LAMR value)	0.12	0.88	0.09	0.21	0.38	0.23
Bridge (LAMR value)	0.02	0.38	0.07	0.05	0.23	0.00
Tennis court (LAMR value)	0.11	0.54	0.12	0.22	0.07	0.11
Ship (A LAMR value)	0.04	0.42	0.13	0.16	0.11	0.04
Basketball court (LAMR value)	0.39	0.89	0.28	0.43	0.40	0.39
Vehicle (LAMR value)	0.18	0.91	0.38	0.89	0.12	0.06
Parking lot (LAMR value)	0.41	0.80	0.49	0.61	0.53	0.41
mLAMR value	0.1130	0.5231	0.1423	0.2731	0.2423	0.1123

4.4 Discussion

As shown in Table 1, the experimental results AP values of numerical indexes of each class obtained by all algorithms in this paper are shown, and the last row is the mAP values of all classes in all methods. It can be seen from the mAP value in the last line that the mAP value obtained by the proposed algorithm is the highest compared with those of the other five compared algorithms. To be specific, the mAP value of the proposed algorithm is 1.93% higher than that of the original algorithm EfficientNet-YOLOv3, 21.18% higher than that of the Faster-RCNN algorithm, 3.50% higher than that of the RetinaNet algorithm, 5.32% higher than that of the SSD algorithm and 4.21% higher than that of the YOLOv4 algorithm, which shows that the algorithm proposed in this paper has achieved good results in the object detection process of remote sensing image.

Table 2 shows the experimental results of the numerical index LAMR of each class obtained by all algorithms in this paper. The last row is the mLAMR value obtained by adding and averaging the LAMR values of all classes in all algorithms. It can be seen from the mLAMR value that the mLAMR value obtained by the algorithm proposed in this paper is the lowest compared with those of the other five compared algorithms. To be specific, the mLAMR value of the algorithm proposed in this paper is 0.0500 lower than that of the original algorithm EfficientNet-YOLOv3, 0.2550 lower than that of Faster-RCNN algorithm, 0.0225 lower than that of RetinaNet algorithm, 0.1150 lower than that of SSD algorithm, 0.0300 lower than that of YOLOv4 algorithm, which shows that the algorithm proposed in this paper has achieved good results in the process of object detection.

As shown in Table 3, the numerical index AP values are obtained from the experimental results of all classes obtained by all algorithms in this paper. The last row shows that the mAP value is obtained by adding and averaging the AP values of all classes in all methods. It can be seen that the mAP value obtained by the algorithm in this paper is the highest compared with the other five compared algorithms. To be specific, the mAP value of the algorithm proposed in this paper is 0.07% higher than that of the original EfficientNet-YOLOv3 algorithm, 31.95% higher than that of Faster-RCNN algorithm, 2.36% higher than that of RetinaNet algorithm, 8.54% higher than that of SSD algorithm, and 10.05% higher than that of YOLOv4 algorithm, which further shows that the algorithm proposed in this paper has a good performance in the object detection process of remote sensing images.

Table 4 shows the experimental results of the numerical index LAMR of all classes obtained by all algorithms in this paper. The last row obtains the mLAMR value by adding and averaging the LAMR values of all classes in all algorithms. It can be seen from the mLAMR value that the mLAMR value obtained by the object detection algorithm proposed in this paper is the lowest compared with the other five compared algorithms. To be specific, the mLAMR value of the object detection algorithm proposed in this paper is 0.0007 lower than that of the original EfficientNet-YOLOv3 algorithm, 0.4108 lower than that of the Faster-RCNN algorithm, 0.0300 lower than that of the RetinaNet algorithm, 0.1608 lower than that of the SSD algorithm, and 0.1300 lower than that of the YOLOv4 algorithm, which shows again that the algorithm proposed in this paper obtains good

results in the process of object detection.

5 Conclusion and Prospect

This paper puts forward the research on small object detection based on dynamic convolution neural network, and compares the object detection algorithm with EfficientNet-YOLOv3 algorithm. Without adding any model parameters, it not only improves the mAP value of object detection in remote sensing image, but also reduces the mLAMR value of object detection in remote sensing image. In addition, it improves the accuracy of the detection process, reduces the missed detection rate of the algorithm, and enhances the stability of the algorithm. Although the object detection algorithm proposed in this paper has achieved good results in RSOD and TGRS-HRRSD remote sensing image data sets, the algorithm has two disadvantages in terms of object detection in remote sensing images. Firstly, the algorithm in this paper is modified on the basis of EfficientNet-YOLOv3 object detection algorithm, and uses dynamic convolutional neural network to replace the traditional convolutional neural network, which improves the complexity of the network model; secondly, the algorithm proposed in this paper is only for object detection in remote sensing images, which is not necessarily very good when using other data sets, that is, its universality is weak. It is hoped that in further research, the use of lightweight backbone feature extraction network will be considered to reduce the complexity of network model, minimize the model parameters, and improve the detection performance of object detection algorithm in remote sensing image.

Abbreviations

mAP:mean Average Precision ;mLAMR:mean Log Average Miss Rate ;ROI trans:region of interest transformer;IOU:Intersection Over Union;DAA:Dynamic Anchor Assignment ;RSI:Remote Sensing Image ;CV:Chan-Vese;SSD:Single-Stage Detection;YOLO: You Only Live Once;MBConv:Mobile inverted Bottleneck Convolution ;FCNN:Full Convolutional Neural Network ;FPN:Feature Pyramid Networks.

Acknowledgements

This work was supported by the Natural Science Fund Project of Gansu Province (No. 21JR7RA300) and the Open project of Gansu Provincial Research Center for Conservation of Dunhuang Cultural Heritage (No.GDW2021YB15).

Authors' Contributions

All authors have made significant contributions to the techniques or methods used, research concepts, data collection, experimental design and critical revision of the articles. The final draft read and approved by all authors.

Funding

Not applicable.

Availability of data and materials

The open datasets supporting the conclusions of this article are available at RSOD(<https://github.com/RSIA-LIESMARS-WHU/RSOD-Dataset>) and TGRS-HRRSD (<https://github.com/CrazyStoneonRoad/TGRS-HRRSD-Dataset/tree/master/OPT2017>). For other datasets, please contact author for data requests.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Electronic and Information Engineering of Lanzhou Jiaotong University, China. ²College of Intelligence and Computing of Tianjin University, China.

References

- [1] Han J W, Zhang D W, Cheng G, et al. Object detection in optical remote sensing images founded by weakly supervised learning and high-level function learning[J]. IEEE Transactions on Geoscience and Remote Sensing,2015, 53(6):3325-3337.
- [2] Cheng G, Han J W. A survey on object detection in optical remote sensing images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 117:11-28.
- [3] Cheng G, Han J W, Zhou P C, et al. Multi-class geospatial object detection and geographic image classification founded by collection of part detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 98:119-132.
- [4] Deng Z P, Sun H, Zhou S L, et al. Multi-scale object detection in remote sensing imagery with convolutional neural networks[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 145:3-22.
- [5] Zhong Y F, Han X B, Zhang L P. Multi-class geospatial object detection founded by a position sensitive balancing framework for high spatial resolution remote sensing imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 138:281-294.
- [6] Liu J K, Li J B, Ma L, et al. Radar target detection method of aircraft wake vortices founded by matrix information geometry[J]. Journal of Radars, 2017, 6(6): 699-708. doi: 10.12000/JR17058.
- [7] Chen X L, Guan J, Huang Y, et al. Radar refined processing and its applications for low-observable moving target[J]. Science & Technology Review, 2017, 35(20): 19-27.
- [8] Su N Y, Chen X L, Guan J, et al. Detection and classification of maritime target with micromotion founded by CNNs[J]. Journal of Radars, 2018, 7(5):565-574. doi:10.12000/JR18077.
- [9] Chen L Q, Shi W X, Fan C E, et al. Ship detection in optical remote sensing images founded by multi-class learning[J]. Journal of Huazhong University of science and Technology, 2019, 47(5):62-67.
- [10] Ding J, Nan X, Yang L, et al. Learning RoI trans-former for oriented object detection in aerial images [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Angeles:IEEE, 2019:2849-2858.
- [11] Xu Y, Fu M, Wang Q, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. IEEE Transactions on Pattern Analysis and

Machine Intelligence, 2020, 41(4):1452-1459.

- [12] Yang X, Yang J, Yan J, et al. Serdet:Towards more robust detection for small, cluttered and rotated objects[C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul: IEEE, 2019:8232-8241 .
- [13] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice:IEEE, 2017:2980-2988 .
- [14]Yao Y Q, Cheng G, Xie X X, et al. Optical remote sensing image object detection founded by multi-resolution function fusion[J]. Journal of remote sensing, 2021, 25(05):1124-1137.
- [15]Wang L W, Feng Y Q, Zhang Y B. Target detection method for optical remote sensing imagery[J]. Systems engineering and electronic technology, 2019, 41(10):2163-2169.
- [16]Yao Q L, Hu X, Lei H. Object Detection in Remote Sensing Images Using Multiscale Convolutional Neural Networks[J]. Journal of Optics, 2019, 39(11):346-353.
- [17] Li C, Zhang Y C, Lan T, et al. An Object Detection Algorithm with Visual Perception for High-Resolution Remote Sensing Images[J]. Journal of Xi'an Jiaotong University, 2018, 52(06):9-16.
- [18] Chen H Y, Liu Z Y, Guo W W, et al. Fast Detection of Ship Targets for Large-scale Remote Sensing Image Founded by a Cascade Convolutional Neural Network[J]. Journal of radar, 2019, 8(03):413-424.
- [19] Zhao T Z, Yang C W, Liu W. Remote sensing image object detection method founded by non-local function enhancement[J]. Journal of Huazhong University of Science and Technology, 2021:1-6.
- [20] Cai H, Gan C, Wang T Z, et al. Once-for-all: Train one network an specialize it for efficient deployment[J]. Computer Science, 2020, pp:1-15.
- [21] Chen Y P, Dai X Y, Liu M C, et al. Dynamic Convolution: Notice over Convolution Kernels[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp:11030-11039.
- [22] Liu, Y, Wang, Y, Wang, S, et al. CBNNet: A Novel Composite Backbone Network Architecture for Object Detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(07), 11653-11660.
- [23] Hua X, Wang X Q, Rui T, et al. A fast self-notice cascaded network for object detection in large scene remote sensing images[J]. Applied Soft Computing Journal,2020, pp:1-20.
- [24] Li C, Cong R, Hou J, et al. Nested Network With Two-Stream Pyramid for Salient Object Detection in Optical Remote Sensing Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(11), pp: 9156-9166.
- [25] Zhang S, He G, Chen H, et al. Scale Adaptive Proposal Network for Object Detection in Remote Sensing Images[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(6),pp.:864-868.
- [26] Zhu H M, Zhang P, Wang L B, et al. A multiscale object detection approach for remote sensing images founded by MSE-DenseNet and the dynamic anchor assignment[J]. Remote Sensing Letters, 2019, 10(10), pp:959-967.
- [27] Yao T, Wang G, Yan L S, et al. Online latent semantic hashing for cross-media retrieval[J]. Pattern Recognition, 2019, 89, pp:1-11.
- [28] Pang J, Li C, Shi J, et al. ,R^2 -CNN: Fast Tiny Object Detection in Large-Scale Remote Sensing Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(8), pp:5512-5524.
- [29] Chen H, Zhang L B, Ma J, et sl. Target heat-map network: An end-to-end deep network for target detection in remote sensing images[J]. Neuro Computing, 2019, 331, pp:375-387.
- [30] Yao X W, Han J W, Guo L, et al. A coarse-to-fine partten for airport detection from remote sensing images using target-oriented visual saliency and CRF[J]. Neuro Computing, 2015, 164, pp:162-172.
- [31] Song Y, Wu Y Q, Dai Y M. Automatic river target detection from remote sensing images founded by image decomposition and distance regularized CV partten[J]. Computers and Electrical Engineering, 2016, pp:1-11.
- [32] Li B, Xie X Y, Wei X X, et al. Ship detection and classification from optical remote sensing images: A survey[J]. 2021, 26(3), pp:145-163.
- [33] Youme O, Bayet T, Dembele J M, et al. Deep learning and remote sensing : detection of dumping waste using UAV[J]. Procedia Computer Science, 2021, 185, pp:361-369.
- [34] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J]. Computer Science, 2018, arXiv preprint arXiv:1804.02767, 2018,1-6.
- [35] Tian M X, Le Q. EfficientNet: Rethinking Partten Scaling for Convolutional Neural Networks[C]// Proceedings of the 36th International Conference on Machine Learning, 2019, 6105-6114.
- [36] Wu K J, Bai C S, Wang D C,et al. Improved Object Detection Algorithm of YOLOv3 Remote Sensing Image[J], in IEEE Access, 2021, vol. 9, pp:113889-113900, 2021, doi: 10.1109/ACCESS.2021.3103522.