ISSN: 2664-9640

Overview of automatic target recognition technology for aerial remote sensing

Jiefu Li [1,2, a,*], Mingtao Liu [1,2], Lijia Cao [1,2]

[1]School of Automation and Engineering, Sichuan University of Science & Engineering, Zigong, 643000, China;

[2]Artificial Intelligence Key Laboratory of Sichuan Province, Zigong, 643000, China;

[a]3228214943@qq.com

Abstract

Automatic target recognition has become a hot topic in the field of computer vision. It is widely used in robot navigation, intelligent video surveillance, aerospace and other fields. In this paper, research on automatic target recognition for aerial remote sensing is described, and traditional target recognition technology is briefly introduced. Then, the current popular target recognition technology based on deep convolution neural network is summarized.Development process of convolutional neural network are introduced, then the representatives of two kinds of algorithms are described: Faster R-CNN is the representative of target recognition method based on region suggestion, and the representative target recognition method based on regression is YOLOv3. Finally, according to the trend of more efficient development of current target recognition algorithms, the future research hotspots of unsupervised and unknown object recognition are prospected.

Keywords

Target recognition; deep learning; feature extraction; convolution neural network; aerial remote sensing.

1. Introduction

Space remote sensing, represented by satellite remote sensing, has a wide coverage and can realize monitoring and analysising in a large range or a region, but the acquisition cycle is long and the image resolution is low.Aerial remote sensing (mainly refers to Unmanned Aerial Vehicle(UAV) remote sensing) as a new remote sensing method, because it's efficient, flexible, fast, low-cost with high-resolution image, etc. has shown a good momentum in recent years[1-5]. As an indispensable supplementary means of satellite remote sensing, it can make up for the high-cost satellite remote sensing, greatly affected by weather conditions, with long working-cycle and other defects[6] .With features of flexibility and maneuverability， it can completely realize unmanned operation and avoid the safety risks brought by flight accidents to the pilots to the maximum extent. At the same time, UAV can easily reach traffic congestion areas or Inaccessible areas.

Compared with natural images, remote sensing images have characteristics of complex background, diverse scales, small and dense targets, etc[7-10]. Due to the low flying altitude of UAV, limited by many factors such as position, direction and space size of photography, especially when monitoring or photographing long-distance targets, tilt angle of the sensor is relatively large, coupled with environmental factors (weather interference, light changes, etc.) , the remote sensing image has more distortion and information loss compared with the general image

ISSN: 2664-9640

Target recognition is a hot topic in the field of computer vision, which is widely used in robot navigation, intelligent video surveillance, aerospace and other fields. Traditional target recognition technology mainly relies on manual designed features to build the model, and the quality of the model depends on the designer's prior knowledge, so the recognition accuracy of this kind of algorithm is not high. In addition, we need to design many models for different types of targets. In one word, the generalization ability of this kind of algorithm is insufficient.

Deep learning algorithm is one of the research hotspots in the field of machine learning. Deep learning algorithm has brought revolutionary progress to machine vision. As early as 1986, the term "deep learning" was introduced into the field of machine learning, and then applied to artificial neural network in 2000. The deep learning algorithm is composed of multiple network layers to complete the learning of data features with multiple abstract layers. In recent years, target recognition technology based on deep learning has made great progress and been widely used in various fields. Because of the strong ability of data learning, the recognition accuracy is significantly improved. In particular, the training model is related to the input data, and there is no need to model the various objectives so the generalization ability is better. In conclusion, it is of great significance to study the key technologies of automatic target recognition based on deep learning in the field of aerial remote sensing.

2. Traditional target recognition technology for aerial remote sensing

The process of traditional target recognition is shown in

Figure 1 process of traditional target recognition: firstly, some candidate regions are selected from the given image, then the features (such as Scale Invariant Feature Transform(SIFT[11]), Histogram of Oriented Gradient(HOG[12]), etc.) are extracted from these regions, and finally, the trained classifiers (such as support vector machines(SVM[13]), AdaBoost[14], etc.) are used for classification.

Figure 1 process of traditional target recognition

Traditional target recognition technology is usually based on manual designed features, which can be divided into methods based on template matching, prior knowledge, and combination of manual feature extraction and machine learning . In order to enhance the generalization ability of target recognition algorithm, only a variety of feature calculation methods based on prior knowledge can be designed to improve the expression ability of manual designed features.

A scale invariant feature transform (SIFT) feature calculation method based on scale-space preserving invariance of image scaling and rotation was proposed in 1999 and further improved in 2004. SIFT algorithm is a method to extract feature invariants from images , and can be used to perform reliable matching between different objects or scene views. SIFT features are invariant to image scale and rotation, and the algorithm can provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise, and change in illumination. An important aspect of SIFT algorithm is that it generates a large number of features to cover the whole scale and position of the image intensively, for example, a 500 x 500 pixel image can produce about 2000 stable features. The number of features is very important for target recognition, for instance, small objects in cluttered background need to match at least three features to ensure reliable recognition.

Support vector machines (SVM) is a machine learning method proposed by Vapnik et al in 1995. This method originated from the statistical learning theory which developed rapidly in the

ISSN: 2664-9640

1970s. It can reflect the idea of structural risk minimization. Because of its perfect statistical theoretical basis and excellent performance in classification and regression, it can solve practical problems such as small-sample, nonlinear, high-dimensional data, over fitting, local minimum value and so on.

Chen Shanjing[15] et al propose a landslide disaster detection method based on multi-source remote sensing images, leveraging the fusion of spatial, temporal and spectral features. The experimental results show that this method can achieve 95% accuracy, which is much better than many common landslide detection methods.

Deng liang[16] proposes a remote sensing target detection method based on non-local self-similarity HOG feature and joint sparse, compared with traditonal algorithums, accuracy is improved by more than 30%.

3. Target recognition algorithm based on deep learning for aerial remote sensing

After convolutional neural network achieved great success in ImageNet dataset in 2012[17], target recognition algorithms based on deep learning has attracted extensive attention[18-21]. Compared with traditional target recognition algorithms based on manual-designed features, target recognition algorithms based on deep learning have significantly improved the accuracy and speed. Especially, if high-quality region candidate frames are used, the localization ability and classification accuracy of target recognition algorithms based on deep learning can be further improved. In traditional target recognition algorithms, corresponding model for every kind of target needs to be designed according to the designer's prior knowledge, so traditional target recognition algorithms does not perform well in universality. However, target recognition algorithms based on deep learning can be applied to multi-class target recognition with only one trained model, which means that generalization ability is better.

Target recognition algorithms based on deep convolution neural network is the development direction in the field of target recognition. Since Girshick[22] proposed the R-CNN model in 2014, more and more fast and accurate target recognition methods based on convolutional neural network have emerge, such as Fast R-CNN[23], Faster R-CNN[24] , YOLO[25-28] (you only look once). These methods can be roughly divided into two categories: target recognition method based on region suggestion whose representative is Faster R-CNN, and target recognition method based on regression whose representative is YOLOv3（Faster R-CNN and YOLOv3 is shown in Figure 2 Target recognition method.

3.1. Faster R-CNN

Faster R-CNN is composed of a region proposal network (RPN) based on fully convolution and a region based target recognition network Fast R-CNN. They implement the tasks of acquiring regions of interest and extracting regional features respectively. The input image outputs the feature map through the shared convolution layer shared by RPN and Fast RCNN network which outputs the location information of multiple regions of interest through RPN network. The region of interest(ROI) and the feature map are input into Fast R-CNN network, and the feature vector of the corresponding region of interest is finally output through forward propagation.

After the image is input into Fast R-CNN network, it is first adjusted to a fixed size, and then a feature map is generated after several convolution layers. The RPN network first performs a 3x3 convolution operation on the input characteristic graph, and then divides it into two paths: one to judge the anchor category, the other to calculate the predicted value of bounding box.The proposed layer synthesizes the target according to the above input values, and eliminates the

ISSN: 2664-9640

target that is too small or beyond the boundary. The proposed layer unifies the size through the ROI pooling layer, and finally outputs the detection results through the full connection layer.

Dong Zhipeng[29] et al propose a convolutional neural network detection and recognition framework based on the scale feature of the high-resolution remote sensing target. Compared with Faster R-CNN, the accuracy of this method is improved by 8% for airplane and ship targets.

Wang jingyu[30] et al propose a multi-channel DNN to overcome relatively small size and weak visual characteristics of UAV, and simulation results show that the proposed DNN model can achieve good results with high robutness.

3.2. YOLOv3

YOLOv3 adopts a new network structure: Darknet-53, which is one of the most advanced feature extraction networks. It contains a large number of convolution layers using 3 × 3 and 1 × 1 convolution kernels. These convolution layers are obtained by integrating convolution layers from various mainstream network structures. They can make the network obtain better performance.

After the image is input into YOLOv3 network, it is first adjusted to a fixed size(for example 416×416), and three scale feature maps are generated in the three residual blocks through feature pyramid networks(FPN), such as 32×32, 16×16, 8×8. The image is divided into several grids according to the size of the feature map (for example, 8×8 grid for 8×8 feature map). Each grid is responsible for predicting the object type and confidence level of the center in this grid. K-means clustering Algorithm is used in YOLOv3 to get the size of priori frames. The prediction box meet the requirements of IOU (intersection over union) and confidence level is introduced to non maximum suppression to get final results.

Liu Bo[31] et al propose a combination of network model based on Darknet model and YOLOv3 al-gorithm to achieve the ship tracking and real-time detecting as well as identifying ship types. Compared with traditional and deep learning methods, the algorithm proposed do not only get better accuracy and faster speed, but also get better robustness to various environmental changes.

Dong Biao[32] et al propose an improved YOLOv3 algorithm to solve the problems of small building detection in remote sensing images. The improved YOLOv3 algorithm effectively solves the problem, mAP rise for 5.35%.

Yu Honggang[33] proposes a forest fire detection method based on Gaussian YOLOv3[34], and the recognition rate is 94% and the detection speed is 47fps on the ground server.

4. Datasets and evaluation indexes

Currently, popular data sets in general target recognition tasks include Pascal voc2007, CIFAR, COCO[35], Imagenet[36] and so on. However, these datasets are not adaptive to target recognition in aerial remote sensing for there are significant differences between these general pictures and aerial remote sensing pictures: resolutions, ranges, denseness and size of targets.

Common aerial remote sensing datasets include DOTA[37], UCAS-AOD[38], NWPU vhr-10[20, 39, 40], etc. DOTA dataset is a dataset jointly produced by State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing and School of Electronic Information and Communications of Huazhong University of Science and Technology. It contains 15 categories and 2806 remote sensing images. UCAS-AOD dataset is a dataset annotated by the pattern recognition and intelligent system development laboratory of the University of Chinese Academy of Sciences, only contains two types of targets: automobiles as well as aircrafts, and negative background samples. NWPU vhr-10 is a space remote sensing target detection dataset annotated by Northwestern Polytechnic University, with a total of 800 images, including 10

ISSN: 2664-9640

categories: aircraft, ship, oil tank, baseball field, tennis court, basketball court, trackfield, port, bridge and vehicle.

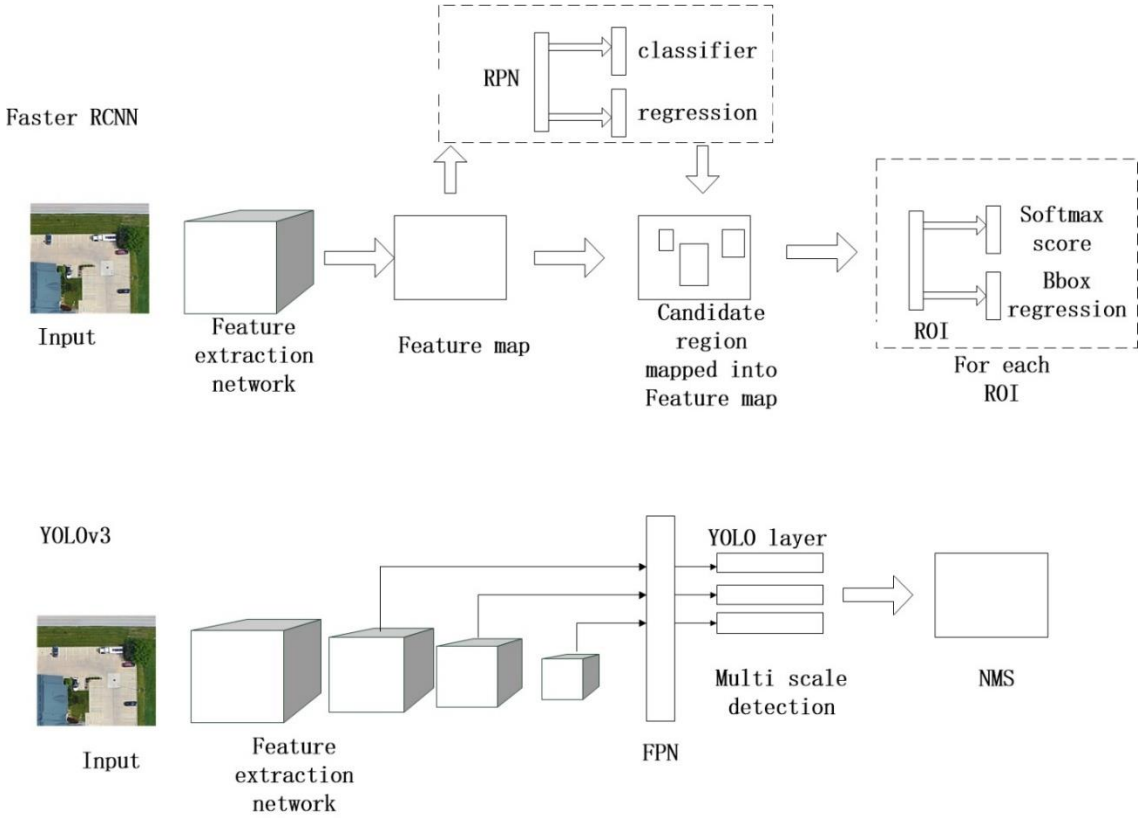

Figure 2 Target recognition method

The commonly used target recognition evaluation indexes are as follows:

Table 1 Confusion matrix

	labeled as positive	labeled as negative
predicted as positive	True Positive,TP	False Positive,FP
predicted as negative	False Negative,FN	True Negative,TN

In the confusion matrix table, the larger the value on the diagonal, the better. But the basic index statistics is just numbers, if the amount of data is large, it is difficult to directly measure the advantages and disadvantages of the model. Therefore, four secondary evaluation indexes and one third level index are extended on the basis of first-rank evaluation index.

Table 2 Evaluation index of second and third level

	formula	meaning
ACC	$ACC = \frac{TP + TN}{TP + TN + FP + FN}$	Proportion of all correctly predicted samples in total samples
PPV (precision)	$PPV = \frac{TP}{TP + FP}$	The proportion of samples predicted to be positive
TPR (recall)	$TPR = \frac{TP}{TP + FN}$	The proportion of predicted positive samples in the actual positive samples

ISSN: 2664-9640

TNR	$TNR = \frac{TN}{TN + FP}$	The proportion of predicted negative cases in the actual negative samples
F-score(third level index)	$F - score = \frac{(1+\beta^2)PPV * TPR}{\beta^2 PPV + TPR}$	reflects the weight relationship between PPV and TPR, where β is the balance coefficient

Over the past decades, the development of data and hardware equipment (especially GPU) and the investment of scientific research force make the classification and recognition based on convolutional neural network develop rapidly. Every year, relevant scientific research results are published in international top conferences and journals in the fields of computer vision, machine learning and artificial intelligence. The major international top conferences include International Conference on computer vision (ICCV), International Conference on machine learning (ICML), IEEE Conference on computer vision and pattern recognition, (CVPR), annual conference on neural information processing systems (NIPS), etc. The main international top journals are IEEE Transactions on pattern analysis and machine intelligence (tpami), IEEE Transactions on image processing (tip), International Journal of computer vision (IJCV), pattern recognition (PR), etc.

5. Summary

Target recognition is a very important research field and has a wide application prospect. In this paper, the emerging deep learning based target detection algorithms are divided into candidate region based and regression based methods. The development and improvement of these two algorithms are reviewed in detail. The popular data sets in the field of target recognition are introduced. Although the current target detection algorithms are widely used in real life, they still exist In many challenges, future target recognition algorithms are worthy of further study in the following aspects:

The first one is how to effectively combine the context information to solve the recognition of small target and occluded target in complex real scene;

The second is to explore a better feature extraction network or specially designed for recognition tasks, as well as a better detection frame selection method;

Thirdly, the current target recognition algorithms are all based on supervised learning, and there are a large number of unlabeled data in reality, so it is very valuable to study the target recognition algorithm based on weak supervised learning;

The last but not the least is to explore how to migrate from known categories of target detection, combined with effective semantic information, to unknown categories of target detection is also a worthy research direction.

Acknowledgements

This work was supported in part by Sichuan Science and Technology Program (No. 2020YJ0368), Zigong Science and Technology Program (No.2019YYJC03), Nature Science Foundation of Sichuan University of Science & Engineering (Nos. 2018RCL18,2017RCL52); Research Foundation of Department of Education of Sichuan Province (No. 17ZA0271); Foundation of Artificial Intelligence Key Laboratory of Sichuan Province (No.2017RZJ02).

References

[1] Fei J. Research on residents extraction of RS images based on texture features[D]. Zhengzhou:

Information Engineering University, 2013.
[2] Wang C. research on geometric correction and object recognition for remote sensing image[D]. Harbin: Harbin Institute of Technology, 2014.
[3] Dai W, Jin L, Li G, et al. Real-time airplane detection algorithm in remote-sensing images based on improved YOLOv3[J]. Opto-Electronic Engineering, 2018,45(12):84-92.
[4] Liu X, Chen J, Yang D, et al. Scene-Coupled Intelligent Multi-Task Detection Algorithm for Air-to-Ground Remote Sensing Image[J]. Acta Optica Sinica, 2018,38(12):262-270.
[5] Lu F, Li Y, Chen X, et al. Weak target detection for PM model based on Top-hat transform[J]. Systems Engineering and Electronics, 2018,40(07):1417-1422.
[6] Zhu S. Research on Target Recognition Method of UAV remote sensing image Based on Deep Learning[D]. Beijing: Beijing University of Civil Engineering and Architecture, 2018.
[7] Liu W, Qi K, Wu B, et al. High Resolution Remote Sensing Image Classification Using Multitask Joint Sparseand Low-rank Representation[J]. Geomatics and Information Science of Wuhan University, 2018,43(02):297-303.
[8] Wang X, Jiang H, Lin K. Remote sensing image ship detection based on modified YOLO algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020,06(46):1184-1191.
[9] Ge Y, Ma L, Jiang S, et al. The Combination and Pooling Based on High-level Feature Map for High-resolution Remote Sensing Image Retrieval[J]. Journal of Electronics and Information Technology, 2019,41(10):2487-2494.
[10] Cai B, Wang S, Wang L, et al. Extraction of Urban Impervious Surface from High-Resolution Remote Sensing Imagery based on Deep Learning[J]. Journal of Geo-Information Science, 2019,21 (09): 1420-1429.
[11] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004,60(2):91-110.
[12] N. D, B. T. Histograms of oriented gradients for human detection: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005[C].20-25 June 2005.
[13] Platt J. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines [R].1998.
[14] Freund Y, Schapire R E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting[J]. Journal of Computer and System Sciences, 1997,55(1):119-139.
[15] Chen S, Xiang C, Kang Q, et al. Multi-Source Remote Sensing Based Accurate Landslide Detection Leveraging Spatial-Temporal-Spectral Feature Fusion[J]. Journal of Computer Research and Development, 2020,57(09):1877-1887.
[16] Deng L. Based on non-local self-similarity HOG feature and joint sparse remote sensing target detection method[J]. Electronic Measurement Technology, 2020,43(06):128-133.
[17] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks: Advances in neural information processing systems, 2012[C].
[18] Sermanet P, Eigen D, Zhang X, et al. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks[J]. arXiv e-prints, 2013:1312-6229.
[19] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015[C].
[20] Cheng G, Han J. A survey on object detection in optical remote sensing images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016,117:11-28.
[21] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016[C].
[22] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014[C].
[23] Girshick R. Fast R-CNN: 2015 IEEE International Conference on Computer Vision (ICCV), 2015[C].
[24] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,39(6):1137-1149.
[25] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016[C].
[26] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger: 2017 IEEE Conference on Computer

ISSN: 2664-9640

Vision and Pattern Recognition (CVPR), 2017[C].
[27] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018:1804-2767.
[28] Bochkovskiy A, Wang C, Liao H M. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. arXiv preprint, 2020(arXiv:2004.10934).
[29] Zhiqi D Z W M. Object detection in remote sensing imagery based on convolutional neural networks with suitable scale features[J]. Acta Geodaetica et Cartographica Sinica, 2019,48(10):1285-1295.
[30] Wang J, Wang X, Zhang K, et al. Small UAV Target Detection Model Based on Deep Neural Network[J]. Journal of Northwestern Polytechnical University, 2018,36(02):258-263.
[31] Liu B, Wang S, Zhao J, et al. Ship tracking and recognition based on Darknet network and YOLOv3 algorithm[J]. Journal of Computer Applications, 2019,06(39):1663-1668.
[32] Dong B, Xiong F, Han X, et al. Research on Remote Sensing Building Detection Based on Improved Yolo v3 Algorithm[J]. Computer Engineering and Applications, 2020,56(18):209-213.
[33] Yu H. Research on Forest Fire Detection Method Based on Deep Learning[D]. Xi'an: Xi'an University of Technology, 2020.
[34] Choi J, Chun D, Kim H, et al. Gaussian YOLOv3: An Accurate and Fast Object Detector Using Localization Uncertainty for Autonomous Driving, 2019[C].October.
[35] Lin T, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context, Cham, 2014[C]. Springer International Publishing, 2014.
[36] J. D, W. D, R. S, et al. ImageNet: A large-scale hierarchical image database: IEEE Conference on Computer Vision and Pattern Recognition, 2009[C].20-25 June 2009.
[37] Xia G, Bai X, Ding J, et al. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images, 2018[C].June.
[38] Zhu H, Chen X, Dai W, et al. Orientation robust object detection in aerial images using deep convolutional neural network, 2015[C].
[39] Cheng G, Han J, Zhou P, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014,98:119-132.
[40] Cheng G, Zhou P, Han J. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016,54(12):7405-7415.