Method research on ship detection in remote sensing image based on Yolo algorithm

Xi Li
College of Computer, National University of Defense Technology
Changsha, P.R.China
lxyl_1124@163.com

Kaiyu Cai
College of Computer, National University of Defense Technology
Changsha, P.R.China
kycai@nudt.edu.cns

***Abstract*—Small target recognition is a classic challenge in the field of target recognition. In this paper, an improved algorithm based on Yolo V3 is proposed for ship detection in remote sensing images. In order to solve the problem that the input satellite image size is too large, the window sliding segmentation technology is used to cut the large image into several small images; because the detection target only has a single kind of ship, K-means algorithm is used to re cluster anchor Box, to improve the detection accuracy of the algorithm; in view of the small ship target in the ocean background, the multi-level feature extraction method is used to highlight the feature vector of small target to improve the recognition ability of small target. Through the above-mentioned methods, using the data set of Airbus ship detection provided by kaggle, the ship target detection has achieved a better comprehensive accuracy rate, which is higher than the accuracy and recall rate of similar target detection algorithms.**

Keywords-component; Ship detection; remote sensing image; window sliding segmentation; Yolo V3; small target detection; FPN

I. INTRODUCTION

There are many hot spots and sensitive areas in China's waters, and foreign ships frequently violate and create conflicts. If they can detect and warn in advance, they will have more initiative in the disposal. With the increasing resolution of remote sensing satellites and the increasing transmission bandwidth of satellite-ground link, a large number of HD satellite image resources provide a good solution. Image recognition technology can be applied to remote sensing images, and timely detection of ships in hot spots can be carried out by taking advantage of technology and resources.

The application of image recognition to ship intelligent detection has the advantages of high efficiency, high reliability and all-weather operation. However, cloud and land background, complex sea conditions and small ship targets also restrict the effectiveness of the detection model. Faced with these problems, traditional detection methods such as SIFT, HOG[1] and SVM[2] can no longer achieve good detection results. With the rapid development of deep learning in the field of target detection in recent years, the target detection algorithm based on neural network shows good performance, image recognition algorithm based on neural network can be divided into two categories: 1. The two stage method, sample first generate candidate box, and then processed by convolution neural network, the method detection accuracy and better positioning accuracy, representative algorithms such as fast R-CNN[3], faster R-CNN[4]. 2. The method of one stage directly transforms the problem of target border positioning into a regression problem, which is then processed by the convolutional neural network. This method has a relatively fast operation speed and represents algorithms such as SSD[5], YOLO[6], YOLO V3[7].

The width range of remote sensing image is 10km-30km, faced with such a huge shooting range, the ship occupies only a few pixels in the image, to realize fast and accurate target recognition, it is necessary to consider the universality of the algorithm and detection speed. So this paper chooses YOLO V3 detection algorithm for the detection of ship targets, and according to the actual situation to make improvements to the algorithm.

II. YOLO V3 ALGORITHM OVERVIEW

YOLO is also known as the Only Look Once algorithm. Its core idea is to combine target area prediction and target category prediction as a regression problem. The neural network is used to directly predict the target boundary and category probability and realize end-to-end target detection, thus greatly improving the performance of the algorithm. YOLO algorithm creates the precedent of anchor free in the field of target detection. On this basis, the author successively put forward YOLO V2[8] and YOLO V3 algorithms.

The YOLO V3 detection process can be roughly divided into three steps: 1. Resize images and recommend 416*416 pixels[9]; 2. Input the image into Darknet-53 network, extract image features and output feature vectors[10]; 3. Perform non-maximal suppression optimization (NMS) on the feature vectors and output the target detection probability[11], as shown in Figure (1).

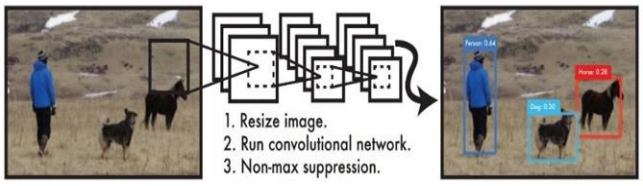

Figure (1) Workflow of YOLO V3 algorithm

A. Image Resize

YOLO V3 algorithm will detect the image size when the image is input and carry out corresponding Resize, mainly

978-1-7281-9668-8/20/$31.00 ©2020 IEEE
DOI 10.1109/ISPDS51347.2020.00029

Authorized licensed use limited to: University of Exeter. Downloaded on May 26,2021 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.

including compression and filling. For example, when a 1024*768 image is input, the image is first compressed to the size of 416*312 using the bicubic interpolation algorithm, and then the image is grayscale filling and expanded to 416*416. The process is shown in Figure (2).

Figure (2) Image Resize process

B. *Darknet-53 network structure*

YOLO V3 uses a series of convolutional layers to extract image feature vectors. From the 0th layer to the 74th layer, there are a total of 53 convolutional layers, each of which carries out BN operation on the input data. 32 convolution kernels are adopted for each convolutional layer, each convolution kernel has a size of 3*3 and a step of 1. There are 21 Res layers, and five different scales and depths are selected to conduct residual operation between outputs of different layers.

In order to achieve accurate projections for objects of different sizes, for a picture in three kinds of feature map to 32 times the sampling size of 13 x13 feature map is more suitable for the prediction of large target, 16 times the sampling size of 26 x26 feature map is more suitable for forecasting medium target, eight times the sampling size is 52 x52 feature map is more suitable for prediction of small target, the network structure as shown in figure (3).

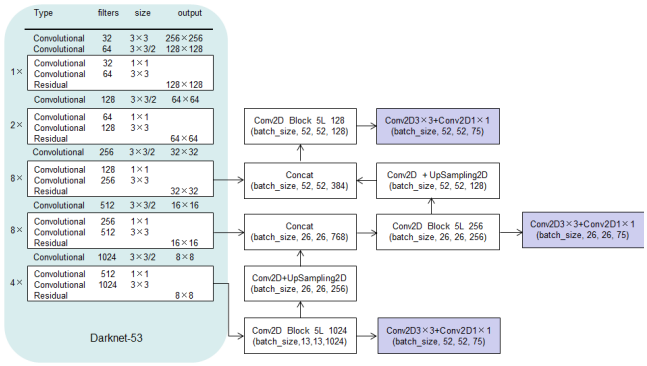

Figure (3) Darknet-53 network structure

C. *Boundary box prediction*

YOLO V3 uses k-means algorithm clustering to initialize Anchor Box. On the COCO data set, according to the size of the input image is 416*416, 9 kinds of clustering results are obtained. Detect network for input do 13 x 13, 26 x 26 and 52 x 52 characteristic of three dimension regression, using formula (1) it is concluded that the boundary Box attributes, including (t_x, t_y, t_w, t_h) is the model prediction output, (c_x, c_y) is the offset, the coordinates of the cell (p_w, p_h) is the presupposition of the Anchor Box of side length, (b_x, b_y, b_w, b_h) is the resulting predict the bounding Box center coordinates of high and wide.

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_w = p_w + e^{t_w} \\ b_h = p_h + e^{t_h} \end{cases}$$

Formula (1)

D. *Tag classification algorithm*

YOLO V3 improved the single label classification adopted by YOLO V2 to multi-label classification. The network structure replaced the Softmax classifier used for single label classification with the Logistic classifier used for multi-label classification, which solved the problem that a target may belong to multiple classes in complex scenes. In the training process, Binary cross-entropy Loss is used as a Loss Function to train the category to predict when the confidence of a bounding box category of an image after feature extraction is greater than 0.5 after sigmoid constraint, it means that the target for which the bounding box is responsible belongs to this category.

III. YOLO V3 ALGORITHM IMPROVEMENT

YOLO V3 algorithm, as one of the representatives in the field of target detection, has achieved good results in the detection of COCO data sets, including the classification of people, vehicles, animals and so on.

But COCO data set target species and in this paper, the detection of target difference is very big, for training the weights of COCO dataset file when the ship detection in remote sensing image, therefore, this article mainly in the following several aspects to the algorithm, modified to enhance the algorithm in this paper in the scene detection effect:

(1) Modify the image Resize mode, sliding partition of the input image window, so as to avoid big loss of pixel image compression;

(2) Use the k-means clustering algorithm to re-cluster the ship targets in the data set, and the Anchor box conforming to the ship targets in the scene in this paper is obtained.

(3) Optimize the feature pyramid network for multi-scale feature extraction to strengthen the recognition of small targets.:

A. *Image window slide segmentation*

Coverage satellite images are often very large, it is mostly in more than 10 km in length, and the size of the ship between 30 meters to 300 meters, this creates in a picture of tens of pixels, a large ship to possess several hundred pixels. If directly to the algorithms of image input, which can identify on the one hand, can cause the ship targets in the process of compression was excessive compression cannot identify, on the other hand can take up a large amount of video memory, affect the operation speed.

Authorized licensed use limited to: University of Exeter. Downloaded on May 26,2021 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.

A pattern needed to look for make very large satellite images can be reasonable divided into multiple small block diagram, the input image size of the algorithm meets the requirements, is divided the tiles covered the space of the independent of each other, in figure edge region and has a certain piece of coincidence degree, thus reducing image recognition at the edge of the ship is divided and lead to error.

This paper design a sliding window mechanism, segmentation window sliding in high resolution image segmentation, the window size for algorithm recommended 416 * 416 pixels, identification error due to considering the edge segmentation, according to the training data set tag box size and window size, the ratio between the setting window sliding is picked up from the 20% of repeat, thus reduce the identification error, as shown in figure (4).

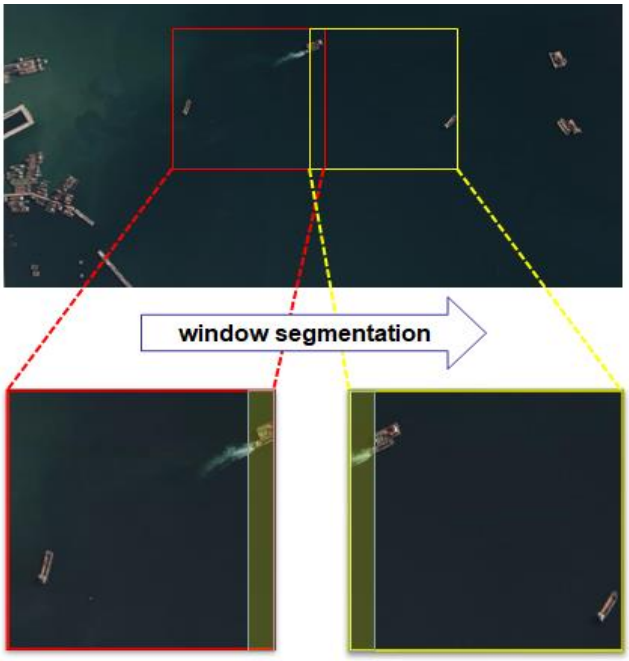

Figure (4) Sliding window segmentation

B. *Prior frame dimension adjustment*

In order to optimize the size of the target box more quickly, YOLO V3 USES k-means algorithm to cluster the annotation boxes in the COCO data set, and obtains 9 prior boxes, which are evenly divided into 3 scale feature maps, so as to obtain more accurate target edge information.

The nine anchor boxes in the COCO dataset are :(10x13), (16x30), (33x23), (30x61), (62x45), (59x119), (116x90), (156x198), (373x326).In terms of allocation, the largest Anchor box (116x90), (156x198) and (373x326) are applied in the smallest 13*13 feature map because of the largest receptor field, which is suitable for detecting larger targets. Medium Anchor Box (30x61), (62x45), (59x119) are used in the feature map of medium 26*26 due to its medium receptor field, which is suitable for detecting medium size targets. In the larger 52*52 feature map, the smallest Anchor box(10x13), (16x30) and (33x23) are applied because of their smaller receptor field, which is suitable for detecting smaller targets.

However, in the application scene of this paper, the Anchor box obtained through COCO data aggregation class is not suitable for the size of ships in remote sensing images. Therefore, in the scenario of remote sensing ship detection, k-means algorithm is used to carry out re-clustering analysis on the training data set. The clustering results are shown in figure 5, and the new anchor box is obtained, from small to large, which are respectively :(14,12), (34,31), (55,88), (84,88), (171,78), (171,176).26×26 feature graphs with lower resolution have larger receptive fields, so a larger prior frame (84,88), (171,78), (171,176) is used; a smaller prior frame (14,12), (34,31), (55,88) is used for feature graphs with higher resolution (52×52) have smaller receptive fields.

```
[0.01771369 0.01598836]
[0.07157069 0.11461432]
[0.10897506 0.05348007]
[0.04395838 0.04031668]
[0.22250873 0.10195847]
[0.22300858 0.22848978]
```

Figure (5) K-means clustering results

C. *Feature pyramid network modification*

In order to strengthen the detection of small targets, YOLO V3 refers to the Feature Pyramid Network and fuses the high-level Feature information with the shallow Feature information. The method of multi-scale fusion is adopted to predict the location and category on the multi-scale Feature map. However, the three-scale feature fusion method adopted by YOLO V3 network structure has a negative impact on the detection of small targets in remote sensing images. The semantic loss of 13×13 feature map is serious, which is easy to cause the loss of small targets.

Considering that the resolution of the feature will directly affect the detection and overall performance index of the small target, the resolution of 13×13, 26×26 and 52×52 of the three scales of the original feature map on the basis of Darknet-53 is modified to the resolution of 26×26 and 52×52 of the two larger scales, as shown in Figure 6.

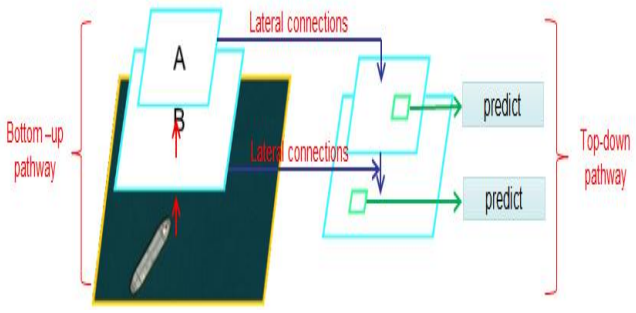

Figure (6) YOLO V3 feature pyramid network model

Authorized licensed use limited to: University of Exeter. Downloaded on May 26,2021 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.

The modified network model takes the resolution of the original network structure 26×26 characteristic map as the first scale; convolve the 61-layer result five times. Firstly, in order to improve the computational efficiency, dimensionality reduction is carried out by 1×1 convolution operation; and then up-sampling, then, it is fused with 36 layers; finally, after fusion, 3×3 convolution is adopted to check the fusion results for convolution, aimed to eliminate the aliasing effect of up-sampling, and thus a new feature graph 52×52 is obtained as the second scale feature, after modification, YOLO V3 network structure can maintain a higher resolution and a larger feature map in the deep network, and improve the speed of recognition.

IV. ANALYSIS OF EXPERIMENTAL RESULTS

A. Experimental platform and data set

The computer hardware configuration used in the experiment in this paper is as follows: CPU i5-4570, graphics card GTX1060, 8GB memory, operating system 64-bit win10, and development environment is python3.7, vs2015, opencv4.1, pycharm2019, CUDA 10.0 and CUDNN 10.0.

The data set used was remote sensing images provided by Airbus Ship Detection Challenge from Kaggle company. Images without ships in the training set were first eliminated, with 20,448 marked images remaining. Among them, 70% is the training set, 20% is the verification set, and 10% is the test set.

Since the annotation file provided by the data set is .CSV structure, the format of the data set is converted to .TXT format required by YOLO. Reconfigure the training files, since there is only one classification, and video card memory is lesser, set up the batch = 64, subdivisions = 32, classes = 1, filters = 18, and through the k-means to clustering of anchor box in the data to replace the original file COCO data set parameters, set training wheels for 20000, add some dummy samples in the late training improve robustness of the algorithm, the training after 2000 round loss curve tends to be stable, as shown in figure 7.

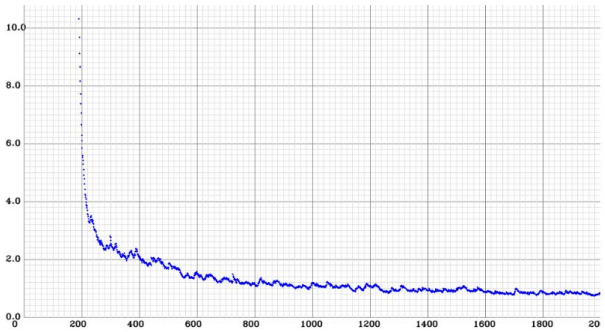

Figure (7) Loss curve

B. The experimental results

The network model of this paper is based on YOLO V3 algorithm, improved the algorithm by segmenting the input super-large image, using k-means algorithm to re-cluster the data set samples, optimizing the feature pyramid network and improving the detection method of small targets, various scenes of remote sensing images in the test set were selected for the test, the effect is shown in Figure (8).

When the input image is too large, the direct identification can only mark out the larger ship in the image. After adopting the sliding window segmentation method, the pixel loss in the image compression process is avoided, so that the small target can be better identified when the large image is input, the recognition rate is greatly improved, and the comparison of detection effects is shown in Figure (9).

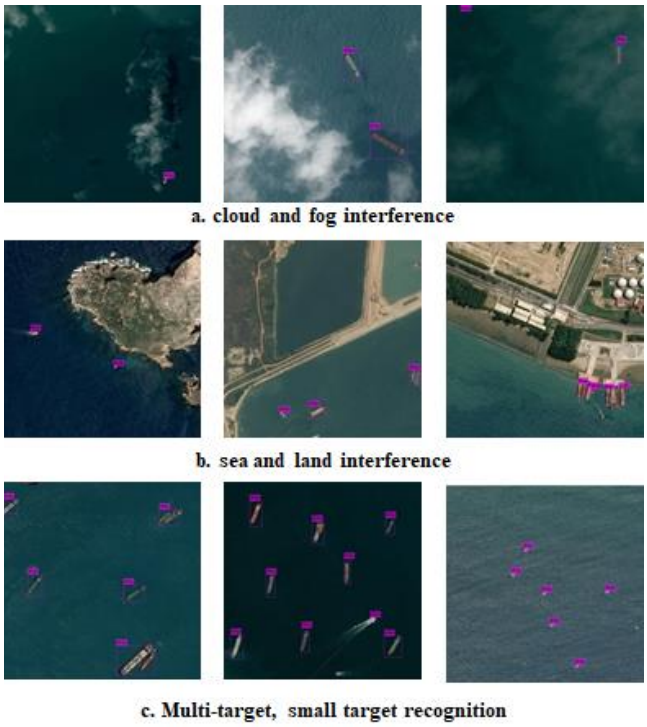

Figure (8) Detection effects in various scenarios

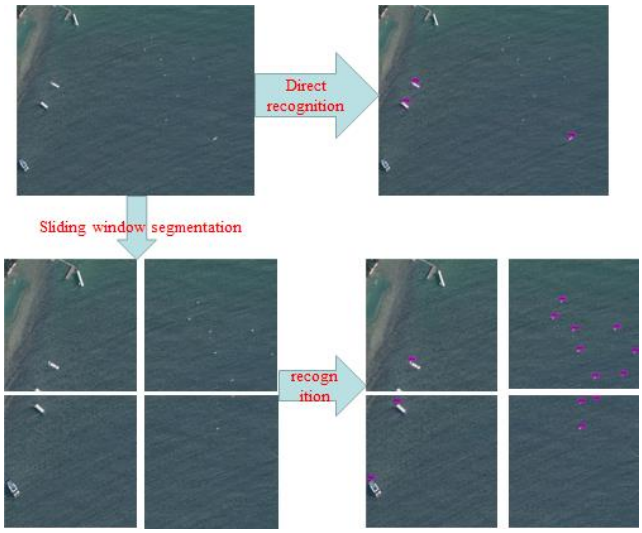

Figure (9) Comparison of detection effects of large images

Authorized licensed use limited to: University of Exeter. Downloaded on May 26,2021 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.

C. Analysis of experimental results

In order to estimate the detection effect of the proposed method on ships, mAP and Recall are used to measure the performance of the algorithm. The values of mAP and Recall are [0, 1]. The specific calculation methods are shown in formula (2), (3) and (4) respectively:

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{Formula (2)}$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{Formula (3)}$$

$$AP = \frac{1}{11}\sum_{recall \in [0,0.1,\cdots 1]} \text{Precision(Recall)} \quad \text{Formula (4)}$$

Among them, TP is the correct number of ship detection; FP is the number of errors detected by the ship; FN is the number of ships missed during detection; AP is the average precision of a class, while mAP is the average of all classes of AP.

YOLO V3, as a representative of the image recognition algorithm based on regression, was modified according to the characteristics of the data set. In order to verify the effect of the improved network algorithm, it was compared and analyzed with SSD, Faster R-CNN and other mainstream algorithms. The test experimental design was as follows: more than 4,000 pieces of Kaggle Airbus test set data were selected for testing, and efficiency comparisons were made by comparing the improved YOLO V3 algorithm, the original YOLO V3 algorithm, SSD algorithm and Faster RCNN algorithm in terms of average detection accuracy, average time per detection and recall rate. The specific parameters are shown in Table 1.

TABLE I. ALGORITHM EFFECT

Algorithm	mAP(%)	FPS	Recall(%)
YOLO V3	*85.9*	28	76.6
SSD	*79.5*	30	72.1
Faster R-CNN	*87.4*	0.7	73.6
Improved YOLO V3	*89.6*	34	85.3

As can be seen from the experimental results, in terms of average detection accuracy, the improved YOLO V3 algorithm for specific data sets has a good recognition effect. The improved algorithm improves the recognition accuracy by 3.7%, and has certain advantages compared with other image recognition algorithms. In terms of detection time (time), the improved YOLO V3 algorithm showed the best recognition speed, reaching 34 frames per second, showing the strong performance of the algorithm in real-time; In terms of recall, the improved YOLO V3 algorithm has improved 8.7%, effectively improving the situation of missing detection and effectively eliminating the interference of clouds, sea and land.

V. CONCLUSION

Aiming at the difficult problem of ship identification in remote sensing image, this paper proposes the YOLO V3 based identification method.

The algorithm was optimized through training sample pre-screening, Anchor box recombination class, feature pyramid network optimization, sliding segmentation of input large image window and other methods, so that the neural network could better extract the ship feature vector, realize accurate identification of ship target in remote sensing image, and reduce the situation of missed detection or false detection.

For complex situations such as cloud cover, sea and land interference, and ocean background interference, negative sample enhancement learning method is adopted to add cloud, land background, diverse ocean background and other images into the training set as negative sample enhancement algorithm's robustness. The experimental results show that the proposed method not only realizes the ship identification in remote sensing images, but also effectively reduces the problems of ship missing detection and false detection caused by the unfavorable factors such as cloud cover, and effectively improves the algorithm performance.

REFERENCES

[1] Moranduzzo T, Melgani F. A SIFT-SVM method for detecting cars in UAV images[C]. 2012 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2012: 6868-6871.

[2] X.-B. Li, W.-F. Sun, L. Li. Ocean moving ship detection method for remote sensing satellite in geostationary orbit[J]. Journal of Electronics & Information Technology, 2015.

[3] GIRSHICK R. Fast r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 1440‑1448.

[4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 39(6): 1137.

[5] REDMON J, DIVYALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 779‑788.

[6] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[J]. ar Xiv preprint ar Xiv: 1612. 08242, 2016.

[7] Redmon J , Farhadi A . YOLOv3: An Incremental Improvement[J]. 2018

[8] Sermanet P , Eigen D , Zhang X , et al. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks[J]. eprint arxiv, 2013.

[9] H. Zhu, X. Chen, W. Dai, K. Fu, Q. Ye, and J. Jiao. Orientation robust object detection in aerial images using deep convolutional neural network[J]. IEEE International Conference on Image Processing (ICIP), 2015: 3735-3739.

[10] Van Etten A . You Only Look Twice: Rapid Multi-Scale Object Detection In Satellite Imagery[J]. 2018.

[11] International Conference on computer vision & Pattern Recognition (CVPR'05). IEEE Computer Society, 2005: 886--893. Dalal N, Triggs B. Histograms of oriented gradients for human detection[C].

Authorized licensed use limited to: University of Exeter. Downloaded on May 26,2021 at 21:15:31 UTC from IEEE Xplore. Restrictions apply.