

Improved Faster R-CNN Algorithm for Sea Object Detection Under Complex Sea Conditions

Liu Yabin

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, Shaanxi, China
E-mail: 1135616154@qq.com

Hu Zhiyi

Engineering Design Institute
Army Research Laboratory
Beijing, 100000, China
E-mail: 763757335@qq.com

Yu Jun

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, Shaanxi, China
E-mail: yujun@xatu.edu.cn

Abstract—In the process of sea surface object detection by remote sensing images, this paper proposes an improved Faster R-CNN sea surface object detection algorithm for the problems of low accuracy and long detection time using deep neural networks. Firstly, the size of the bounding box in the region proposal networks is analyzed, and then the object in the image is clustered using the K-Means algorithm, and the clustering results are input into region proposal networks, thereby realizing the region proposal networks improvement of. Secondly, the Soft-NMS algorithm is used to screen the target candidate frame to obtain the detected sea objects. The experimental results show that the algorithm of this paper can detect the sea surface object of remote sensing images under complex sea conditions, and its mAP can reach 87.25%, which is an average increase of 3.75% compared with the commonly used detection methods.

Keywords—Sea Surface Object Detection; Faster R-CNN Algorithm; K-Means Clustering Algorithm; Soft-NMS

I. INTRODUCTION

As an important field of computer vision research, marine object recognition is aimed at achieving automatic recognition of marine ships. Realizing the automatic identification of ships can help the construction of unmanned ports, ship search and rescue, monitoring and combating illegal activities such as illegal fishing, smuggling and piracy. At the same time, it also has certain application value in the military[1-2], such as supervising our country's territorial waters, monitoring illegal intrusion ships, and assisting in analyzing the deployment of the enemy's key ports and military warships. Therefore, it is of great practical

significance to improve the detection accuracy of sea surface objects.

In recent years, with the sharp increase in the resolution of satellite remote sensing images in China, many trainable remote sensing image samples have been provided for deep learning. How to effectively use massive high-resolution remote sensing images is one of the main problems faced by current object detection[3-6]. Reference[7] studied the object recognition of remote sensing images based on Faster R-CNN deep network, and proved that the deep learning method can realize the rapid and accurate recognition of remote sensing image objects. Reference[8] discusses that when satellites take remote sensing images of the sea, they are greatly affected by weather conditions. Remote sensing images may have problems such as relatively small ship size, cloud cover, and background interference on land. Reference[9] proposed a ship detection algorithm based on the Mask R-CNN framework, but this algorithm will have problems such as missed detection and false detection for closely arranged multi-scale ships. Reference[10] cites the hollow convolution and global attention modules to extract more feature information, and finally constrains the salient features obscured by clouds. However, when there are small objects, cloud cover and land background interference, the detection accuracy of the above method is not high. In this paper, the Faster R-CNN[11-13] object detection algorithm is improved. The K-Means clustering algorithm[14] is

introduced to perform cluster analysis on the size of the object in the image, and the clustering results are directly input into the area recommendation network to achieve the improvement of the area recommendation network. Using Soft -The NMS algorithm replaces the NMS algorithm to reduce the miss detection probability of small objects. Finally, the average accuracy of object detection is improved.

II. IMPROVEMENT OF FASTER R-CNN

The object detection algorithm studied in this paper is mainly the Faster R-CNN object detection network model, and it has been improved.

A. Principle of Faster R-CNN

Faster R-CNN is a object detection network model proposed by Shaoqing Ren. It introduces RPN (Region Proposal Networks) to extract features from object candidate regions, and uses RPN to replace the original selective search algorithm. The detection process of Faster R-CNN model can be shown in Figure 1. First, input images of any size and corresponding annotation files into the network model. Second, they will go through the convolution layer to extract the features of the input image. Third, use RPN for region prediction, and use the ROI mapping operation to map the predicted candidate frame to the feature map. Finally, identify the target category and locate the bounding box. It can be seen that in the Faster R-CNN detection process, the recommendation of target candidate regions, feature extraction, and object detection process are all integrated into the network model.

In the Faster R-CNN network model, the main function of RPN is to generate target candidate boxes. The input of the RPN is a $n \times n$ size dimensional feature map, and the output is a series of target candidate boxes. Before entering the area generation network, a small network will choose sliding windows on the final convolutional feature map. For each window obtained, k target candidate regions are predicted, which are called anchors. Each anchor corresponds to a different scale, size, and center point on the convolutional feature map.

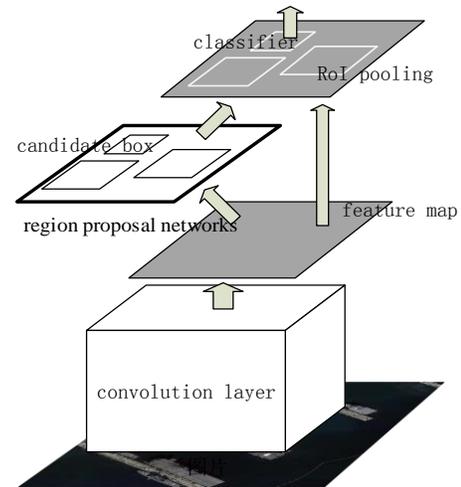


Figure 1. Fast R-CNN model framework

The Region Proposal Networks is shown in Figure 2. A $w \times h$ size convolutional feature map, corresponding to $w \times h$ anchors of different sizes, each window is mapped into a low-dimensional vector, and then this feature vector is input into two subnetworks: border classification network, border Back to the network. The function of the border regression network is to modify the size of the anchor to obtain an accurate candidate area. Its output is the translation zoom value of each anchor, and each window will have $4k$ outputs. The function of the boundary classification network is to determine whether the anchor belongs to the background or the object, and its output is the probability that each anchor belongs to the background or object, and each window corresponds to $2k$ values.

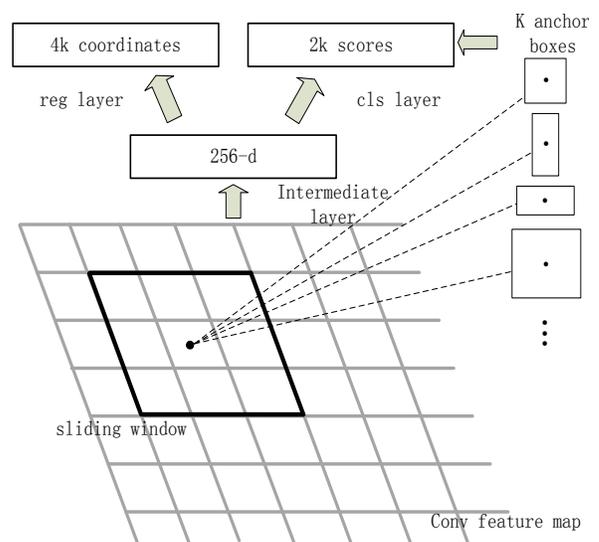


Figure 2. Region Proposal Networks

The Region Proposal Networks training process is end-to-end. The optimization methods used are stochastic gradient descent method and back propagation method. The loss function of the network model is a combined function of classification error and regression error. The formula is shown in formula (1).

$$L(\{P_i\} + \{t_i\}) = \frac{1}{N_{cls}} \cdot \sum_i L_{cls}(p_i, p_i^*) + \frac{\lambda}{N_{reg}} \cdot \sum_i p_i^* \cdot L_{reg}(t_i, t_i^*) \quad (1)$$

Among them, i in the formula represents the i -th anchor point; p_i represents the probability that the candidate box is the object, if $p_i=1$, it means that the i -th anchor point is a positive sample; t^* represents the deviation between the candidate box and the real box. During training, if the IoU value of the candidate area and any real frame is greater than 0.7, it is expressed as a positive sample; if the IoU value of the candidate area and any real frame is less than 0.3, it is expressed as a negative sample; if it does not belong to the above two cases, It means that the candidate area is not used in the training.

B. Faster R-CNN improved method

Although the Faster R-CNN network model integrates the RPN network and the Fast R-CNN network into a network model, a object detection algorithm framework based on deep convolutional networks that uses end-to-end training in a true sense is constructed. But Faster R-CNN also has shortcomings, mainly including:

a) *The accuracy of Faster R-CNN on remote sensing data set is low.* Especially when there is cloud and fog blocking and shore-based interference, the problem of missed detection of objects is prone to occur.

b) *The threshold of Non-Maximum Suppression (NMS) is difficult to determine.* If the threshold is set too small, it will cause the problem of missed detection of the network model; if the threshold is set too large, it will cause the model to misdetect.

Based on the above problems, this paper proposes an improved detection model of Faster R-CNN algorithm, which mainly includes:

a) *Use the K-means algorithm to readjust the size and number of RATIOS.*

b) *Replace the non-maximum suppression (NMS) algorithm with the Soft-NMS algorithm.*

Each will be described below. The following will be described in detail.

1) *K-means algorithm readjusts the size and number of RATIOS*

Because the existing public data sets (VOC, COCO) do not contain data related to remote sensing sea surface object images, if the original parameters of Faster R-CNN are used for training, it will have a certain impact on the training speed and detection effect. Therefore, it is necessary to re-cluster the labels of the remote sensing data sets to obtain the size and number of RATIOS that are favorable for detecting the remote sensing data sets.

TABLE I. ANCHOR_RATIOS FOR DIFFERENT K VALUES

number	K=1	K=2	K=3	K=4	K=5	K=6	K=7
	0.6	1.3	0.5	0.7	0.6	0.5	0.5
	-	1.4	1.2	1.2	1.2	0.7	0.6
	-	-	1.6	1.3	1.3	1.2	1.1
results	-	-		2.9	1.7	1.3	1.2
	-	-	-	-	2.6	2.4	1.8
	-	-	-	-	-	3	2.1
	-	-	-	-	-	-	2.9
mAP(%)	83.84	84.3	84.32	84.98	84.29	84.57	84.34
Time(s)	503	510	515	518	524	515	517

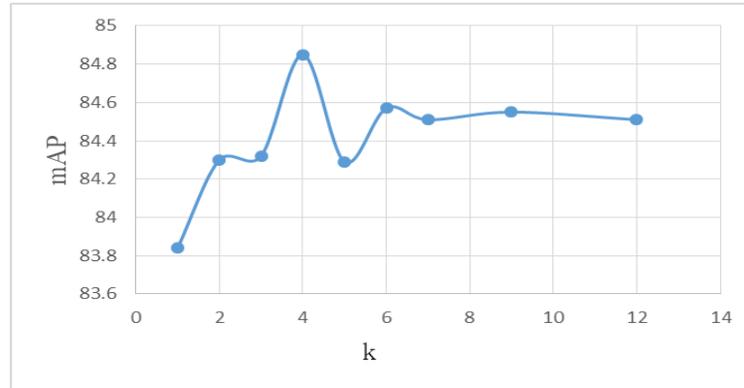


Figure 3. mAP of different K values

The K-means algorithm is used to perform dimensional clustering analysis on the labels of the remote sensing data set. With different values of K, the values of RATIOS are shown in Table 1. In the case where other parameters are unchanged, the resulting average accuracy change is shown in Figure 3.

It can be seen from Figure 3 that when K=4, the average accuracy rate is the highest; as the K value increases, the average accuracy rate tends to be stable; combined with Table 1, it can be seen that with the increase of the K value, the detection time It is also increasing; therefore, the clustering result of K=4 is taken as the improved parameter of RATIOS, that is, RATIOS =[0.7, 1.2, 1.3, 2.9].

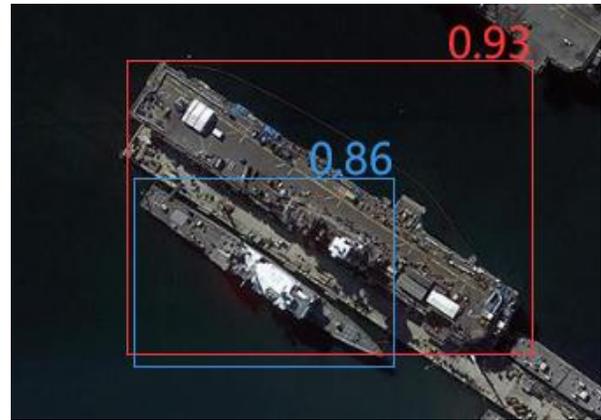
2) *Replace NMS algorithm with Soft-NMS algorithm*

Non-maximum suppression (NMS) is an important part of object detection process. First, it sorted based on the object's confidence, and selected the detection box M with the highest score. Then, if the overlap area between other detection boxes and M is greater than a certain proportion, the other detection boxes will be deleted. This process is recursively applied to the remaining check boxes. The formula of NMS algorithm is shown in Equation (2).

$$s_i = \begin{cases} s_i, iou(M, b_i) < N_t \\ 0, iou(M, b_i) \geq N_t \end{cases} \quad (2)$$

Where, S_i represents the confidence score of all detection boxes; N_t is an artificially set threshold; M is the detection box with the highest confidence score in each iteration; B_i is the detection box to be processed; $Iou(M, B_i)$ is the overlapping area between M and B_i . According to formula (2): when the overlapping area

$Iou(M, B_i)$ is greater than the artificially set threshold N_t , B_i will be deleted. Otherwise, retain b_i .



(a) Schematic diagram of the omission situation



(b) Schematic diagram of false detection

Figure 4. Is a schematic diagram of the shortcomings of The NMS algorithm

The disadvantage of the NMS algorithm is that the artificially set threshold N_t is difficult to determine. If the threshold is set too high, many objects with low confidence scores will be deleted directly; if the

threshold is set too low, many false objects will be detected, resulting in false detections. In addition, when b_i is very close to the selected detection frame M and contains an object, the formula (2) still cannot achieve the effect of accurate detection. As shown in Figure 4(a), the detection results of the current image are marked with red boxes and blue boxes, respectively, and the scores of the two are 0.93 and 0.86, respectively. If the traditional non-maximum suppression method is adopted, the red box with the highest score will be preferentially selected, and the green box that is too large in the overlapping area will be deleted. This caused a missed inspection. Similarly, if the threshold is set too low, it will cause false detection as shown in Figure 4(b).

In order to improve the accuracy of object detection, we replaced the NMS algorithm with Soft-NMS algorithm. The Soft-NMS algorithm reduces the score of the adjacent detection frame through a function related to the overlap of the detection frame M , but it will not be completely deleted. Although the score has dropped, the adjacent detection frames are still in the sequence of the target detector. Therefore, the Soft-NMS algorithm effectively solves the problem of missed detection of objects. The core formula of its algorithm is the Gaussian penalty function, as shown in equation (3).

$$s_i = s_i \cdot e^{-\frac{\text{IoU}^2(M, b_i)}{\sigma}} \quad (3)$$

Among them, s_i represents the confidence score of all detection frames; M represents the detection frame with the highest confidence score; b_i is the unprocessed detection frame. If the IoU values of b_i and M are larger, the value of confidence s_i decreases more. The formula will be applied to each subsequent iteration and update the confidence of all remaining detection frames.

III. PREPARE EXPERIMENTAL AND ANALYSIS

The improved Faster R-CNN was used to conduct remote sensing image sea surface object detection experiment, and its detection process was shown in Figure 5. First, enter an image of any size and enter the ResNet101 convolutional network layer after certain

adjustments. Secondly, the ResNet101 network layer outputs the feature map corresponding to the image. Again, RPN extracts candidate box feature blocks from the convolutional feature map. Then, the ROI pooling layer extracts the candidate block feature block and the feature map output by the convolution layer. Finally, the above results are sent to the fully connected layer, and the final object detection result is output after passing through the Soft-NMS.

C. Preparation of data sets

Experimental data were captured from Kaggle competition platform and Google Earth. Among them, the training set has a total of 192,556 images, including 150,000 images (77.9%) with no ship empty, and 42,556 images (22.1%) with ship images. Since there are many negative samples in the data set, it is necessary to select an effective graph from the training set.

In the data set, there are mostly cargo ships and passenger ships, and the number of container ships, warships, and aircraft carriers is relatively small, and most of the remote sensing data sets are overhead shots, and the orientation of the ships changes greatly. In order to make the trained model extensive, it is necessary to amplify the data set. There are four main amplification methods: flip horizontally or vertically, add random noise, rotate the image at different angles, and blur the image by the specified amount. The amplified data set is divided into three categories: training set, validation set and test set, and the division ratio is 65%, 15% and 20% respectively. After that, annotate the data set.

D. Experimental Environment

The experiment in this article is conducted under Linux system. In order to reduce the training time of the deep neural network model and improve the calculation speed, the Nvidia TITAN XP 12G graphics card is used, and CUDA9.0 and cuDNN7.0 are configured to call the GPU for acceleration. Use Tensorflow deep learning framework.

E. Training model

The main parameters of the detection network are shown in Table 2. Set up to save a model every 5000 iteration batches, and finally, use the obtained model to carry out the detection experiment.

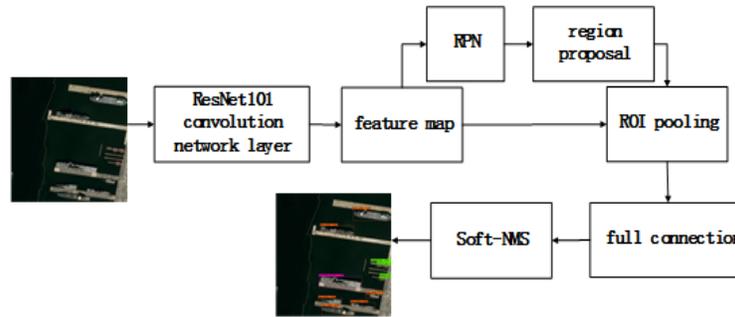


Figure 5. Improved object detection network

TABLE II. VALUES OF MAIN NETWORK PARAMETERS

Parameters	Values	Parameters	Values
LEARNING_RATE	1e-3	Batch_size	256
Anchor_Scales	[8,16,32]	Anchor_RATIOS	[0.7, 1.2, 1.3, 2.9]
ITERS	85000	num_classes	6
SOFT_NMS	1		

In Table 2, if the value of LEARNING_RATE is set to a small value, the rate of model convergence will become slower, and it is easy to fall into the local optimal value; if the setting is too large, the difficulty of model convergence will increase, the network parameters will change in a large range, and oscillate serious. After many experiments, set LEARNING_RATE to 1e-3.

In general, the larger the Batch_size, the better the model. However, in the actual training process, if the Batch_size is too large, the memory of the computer is insufficient, and if the Batch_size is too small, the trained network is difficult to converge. In this experiment, the Batch_size is set to 256. During the model training process, after each epoch training is completed, the network will output the total loss function value, classification loss function and loss function of the positioning frame of each epoch. If the

total loss the function value and other loss function values are continuously decreasing, indicating that the network is converging and the Batch_size value is appropriate.

F. Experimental results and analysis

The experimental results include three parts: the detection results of the VGG-16 structure, the detection results of the ResNet101 structure, and the detection results of the algorithm in this paper. The experimental results are shown in Table 3. Table 3 Comparison of detection results of three network structures.

It can be seen from Table 3 that the detection results of the algorithm in this paper is best. Compared with the VGG-16 and ResNet101 structure methods, mAP has increased by 4.29% and 3.21%, respectively. The average accuracy of Passenger_ship has increased by 16% and 12.8%, respectively.

TABLE III. COMPARISON OF DETECTION RESULTS OF THREE NETWORK STRUCTURES

Detection method	Passenger_ship	Cargo_ship	Container_ship	Aircraft_ship	War_ship	mAP
VGG-16 structure	50.9%	80.6%	92.4%	98.1%	92.8%	82.96%
ResNet101 structure	54.1%	81.0%	92.8%	99.3%	93.0%	84.04%
improved ResNet101 structure	66.9%	82.3%	93.65%	99.51%	93.9%	87.25%

It can be seen from Figure 6 that under the background of complex sea surface (such as: shoreline interference, cloud interference, the presence of small ships), the ships in the remote sensing image can be

detected by the algorithm of this paper and marked in each picture. The box shows the predicted confidence score.

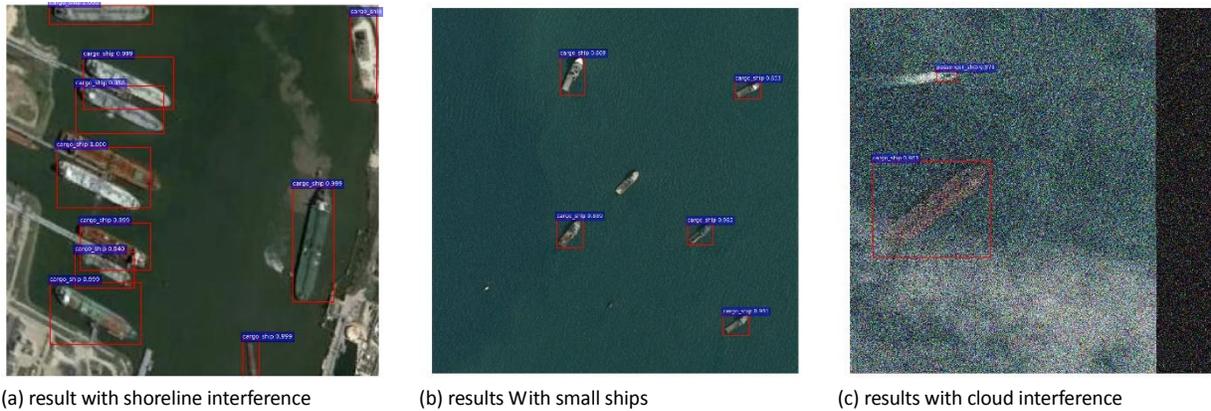


Figure 6. Test results

IV. CONCLUSION

In this paper, an improved Faster R-CNN algorithm for sea surface object detection is proposed, and the training and testing of the object detection network are carried out on the remote sensing image sea surface ship data set. The experimental results show that the improved Faster R-CNN algorithm improves the average accuracy of sea object, and solves the problems of low average accuracy and easy missing of small objects on the detection process of sea objects in complex sea environment.

REFERENCES

- [1] Xu Fang. Research on key Technologies for Automatic Detection of Surface Objects in Visible light remote sensing images [D]. University of Chinese Academy of Sciences (Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences), 2018.
- [2] Zhang Zemiao, Huo Huan, Zhao Fengyu. Review of Object detection algorithms for deep convolutional neural networks [J]. Minicomputers, 2019, 40(09): 1825-1831.
- [3] Yao Qunli, HU Xian, LEI Hong. Research Progress of Deep convolutional neural network in Object Detection [J]. Computer Engineering and Applications, 2016,54(17):1-9.
- [4] Xu Junfeng, Zhang Baoming, Guo Haitao. A multi-feature fusion object-oriented multi-source remote sensing image change detection method [J]. Journal of surveying and mapping science and technology, 2015, 32(05):505-509.
- [5] Sun Hao, Sun Xian, Wang Hongqi. Research on a high-resolution remote sensing image ship detection method [J]. Surveying and mapping science, 2013, 38(5):112-116.
- [6] Wang Tengfei. Research on high-resolution Remote Sensing Image Deep Learning Ship Detection Technology [D]. Harbin Institute of Technology, 2017.
- [7] Wang Jinchuan, TAN Xicheng, WANG Zhaohai, ZHONG Yanfei, Dong Huaping, Zhou Songtao, Cheng Boyi. Research on remote sensing image Object recognition Method based on Faster R-CNN Deep Network [J]. Journal of Earth Information Science, 2016, 20(10):1500-1508.
- [8] Yao Hongge, Wang Cheng, Yu Jun, Bai Xiaojun, Li Wei. Small Object ship identification in complex satellite images [J]. Journal of remote sensing, 2020, 24(02): 116-125.
- [9] Wu Jinliang, WANG Gang, Liang Shuo, Chen Jinyong, Gao Feng. Research on Ship Object Detection based on Mask R-CNN [J]. Radio Engineering, 2008, 48(11):947-952.
- [10] He Yubo, LIU Kun. Surface Significance Object Detection based on convolutional neural network [J/OL]. Computer Engineering and Application:1-10[2020-06-09].
- [11] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. Computer Science, 2015(99):1.
- [12] Zhang Y, Sohn K, Villegas R, et al. Improving object detection with deep convolutional networks via Bayesian optimization and structured prediction[C]. Computer Vision and Pattern Recognition. IEEE, 2015:249-258
- [13] Girshick R. Fast R-CNN[C]. the IEEE International Conference on Computer Vision, December 7-13, 2015, Chile. New Jersey, IEEE Press, 2015:1440-1448.
- [14] Engineering; Study Data from Northeast Electric Power University Provide New Insights into Engineering (Fingerprint location algorithm based on K-means for spatial farthest access point in Wi-Fi environment)[J]. Journal of Engineering, 2020.