JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Aircraft Recognition from Remote Sensing Images Based on Machine Vision

Lu Chen*, Liming Zhou*, and Jinming Liu**

Abstract

Due to the poor evaluation indexes such as detection accuracy and recall rate when Yolov3 network detects aircraft in remote sensing images, in this paper, we propose a remote sensing image aircraft detection method based on machine vision. In order to improve the target detection effect, the Inception module was introduced into the Yolov3 network structure, and then the data set was cluster analyzed using the k-means algorithm. In order to obtain the best aircraft detection model, on the basis of our proposed method, we adjusted the network parameters in the pre-training model and improved the resolution of the input image. Finally, our method adopted multi-scale training model. In this paper, we used remote sensing aircraft dataset of RSOD-Dataset to do experiments, and finally proved that our method improved some evaluation indicators. The experiment of this paper proves that our method also has good detection and recognition ability in other ground objects.

Keywords

Aircraft Recognition, Inception Module, Remote Sensing Images

1. Introduction

In recent years, with the development of science and technology, aerospace, remote sensing, sensors, and other related technologies are also improving day by day. Therefore, the data and information in remote sensing images are becoming more detailed and focused. Remote sensing image processing includes many aspects and target detection, and, as an integral part, it plays an important role in various fields [1]. This includes ship detection in the sea area or building detection in satellite images [2], and so on.

Aircraft are indispensable tools in civil life and military operations. It reflects not only a country's economic strength, but also its military capability. In the remote sensing image, the external features such as the shape scale and edge architecture of the aircraft were acquired from deep learning, which improved the number and accurate positioning of the aircraft in the detection of complex scenes. Due to the actual needs of social development, some early object detection methods are constantly updated and replaced according to different needs, and many new detection algorithms emerge in the academic circle and are used in the research of various fields. Compared with other detection models, Yolo (you only look once) has better generalization effect in some fields.

To solve the problem of poor evaluation indexes such as recall rate during Yolov3 [3] detection, this paper proposes a remote sensing image aircraft recognition method based on machine vision to improve the object detection effect.

The following points are the main content of this paper:

- The model trained on the PASCAL-VOC2012 dataset was used as the pre-training model for the training aircraft data to avoid overfitting, and the aircraft dataset was used for fine-tuning on the pre-training model to greatly shorten the time, obtain high detection accuracy and guarantee the robustness of the model.
- According to the k-means algorithm, the aircraft data set was re-clustered, which improved the aircraft detection results in the airport scene.
- The Inc-v1-Yolov3 network is proposed. Based on Yolov3 network, Inception module and multi-scale prediction are introduced to optimize the network structure. Training experiments were conducted on remote sensing aircraft data in RSOD-Dataset to verify the changes of various indexes such as Recall, F1-score and IOU (intersection-over-union).
- The algorithm proposed in this paper is extended to the recognition and detection of other ground objects to reflect its practicability and popularization value.

2. Related Works

With the development of target detection technology, there are two main methods of object recognition and detection. One method is region-based object recognition, such as region convolutional neural network (R-CNN), Fast R-CNN, Faster R-CNN, Mask R-CNN, etc. This method is effective to detect small targets, but it takes a long time. Another method is object recognition based on regression, such as Yolo, SSD (single shot multibox detector), Yolov2, Yolov3, etc. Among them, the object recognition method based on regression adopts the end-to-end object detection and recognition, which has the advantage of speed while ensuring the detection results.

The Yolo series models are widely used in many fields. Sindhu Ramachandran et al. [4] proposed the DetectNet architecture based on the Yolo model to detect nodules in CT scans of the lungs. Chang et al. [5] used Yolov2 for architecture modeling and model training, and proposed a deep learning method based on enhanced GPU to detect ships from SAR images. Aiming at the problem of oil tank detection in remote sensing images, Qian et al. [6] proposed an image detection method based on improved Yolov3 algorithm. In the research task of remote sensing image aircraft detection, the R-CNN series algorithm, SSD algorithm and Yolo series algorithm are controversial about the detection effect. Therefore, we select the Fast R-CNN algorithm with high evaluation in the R-CNN series and the latest version of Yolov3 algorithm in the Yolo series, and compare them with SSD algorithm for aircraft detection. The results of the experiment in [7] show that Yolov3 has better detection effect than the other two algorithms, whether it is aircraft data in RSOD-Dataset or aircraft data in NWPU VHR-10 dataset.

Since the Yolo network architecture was proposed, it has attracted extensive attention due to its fast recognition. Its accuracy of object detection, however, is still relatively poor. In recent years, Yolo has experienced three improvements by continuously optimizing itself and absorbing the advantages of other object detection models. The latest Yolov3 has significantly improved the speed and accuracy of identification and has been applied in many fields. Yolov3 using similar feature pyramid networks (FPN)

pyramid structure of the scale of the network, the original image according to the size of the scale of the characteristic figure is divided into S×S cells of equal size. Detection is performed on three scales with the size of the feature map of 13×13, 26×26, 52×52, and the feature map is transmitted on two adjacent scales using two times up-sampling. Each cell is predicted by three anchor boxes. The Yolov3 network structure is shown in Fig. 1.
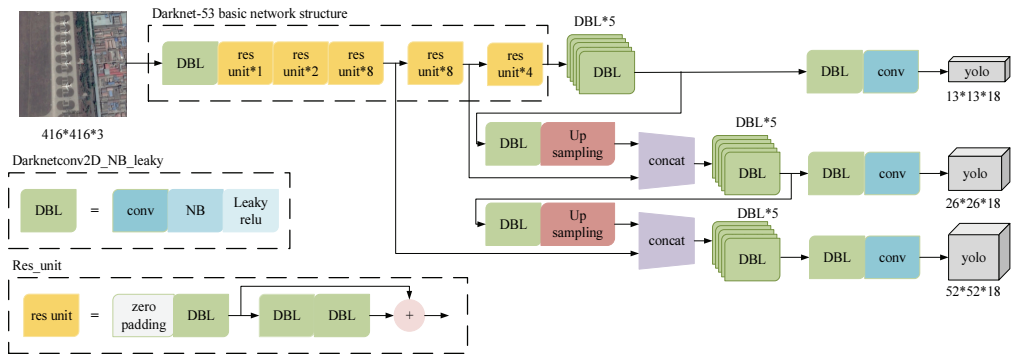
Fig. 1. Yolov3 network structure.

3. Proposed Method

This paper mainly studies aircraft detection in remote sensing images. In the conventional image, the background is relatively simple, the main object is usually located in the center of the image, occupying the main part of the image, so it is easy to distinguish the background from the main object. However, a high-resolution remote sensing image contains more objects than a conventional image, has more shape and texture information, and the objects may be scattered throughout the image. In addition, the detected object (aircraft) is relatively small and close to the background (airport scene). If we shrink the remote sensing image to a relatively small size to observe, we will lose a lot of details, and the detected object may hardly see it. As a result, anchor boxes of the original network are defined to detect conventional image target for the research object of this article is not applicable, need to change. At the same time, a new network structure is proposed to improve the target detection effect of remote sensing image in complex background. Flow chart of aircraft detection in airport scene is shown in Fig. 2.

(1) The "forward propagation" operation was first carried out, and the PASCAL-VOC2012 data set training samples were sent into the initial network structure for feedforward network calculation, and then the predicted results were obtained and input into the subsequent stage of "back propagation". In "back propagation", the error between the predicted output value and the actual label is calculated, the network weight is updated according to the error, and then we continue to return to the feedforward network through the error until the iterations is complete.

(2) According to the transfer learning [8], input remote sensing aircraft training samples. Pre-trained on PASCAL-VOC2012 dataset to obtain the model parameter darknet53.conv.74, and use this parameter as the initial weight value. Using the k-means clustering algorithm to cluster analysis of data set goals, adjust the value of the anchor boxes, and then modify the network structure (Inc-Yolov3 model), finally predict on three scales. In the process of training, parameters are adjusted constantly to optimize the model in order to achieve better results.
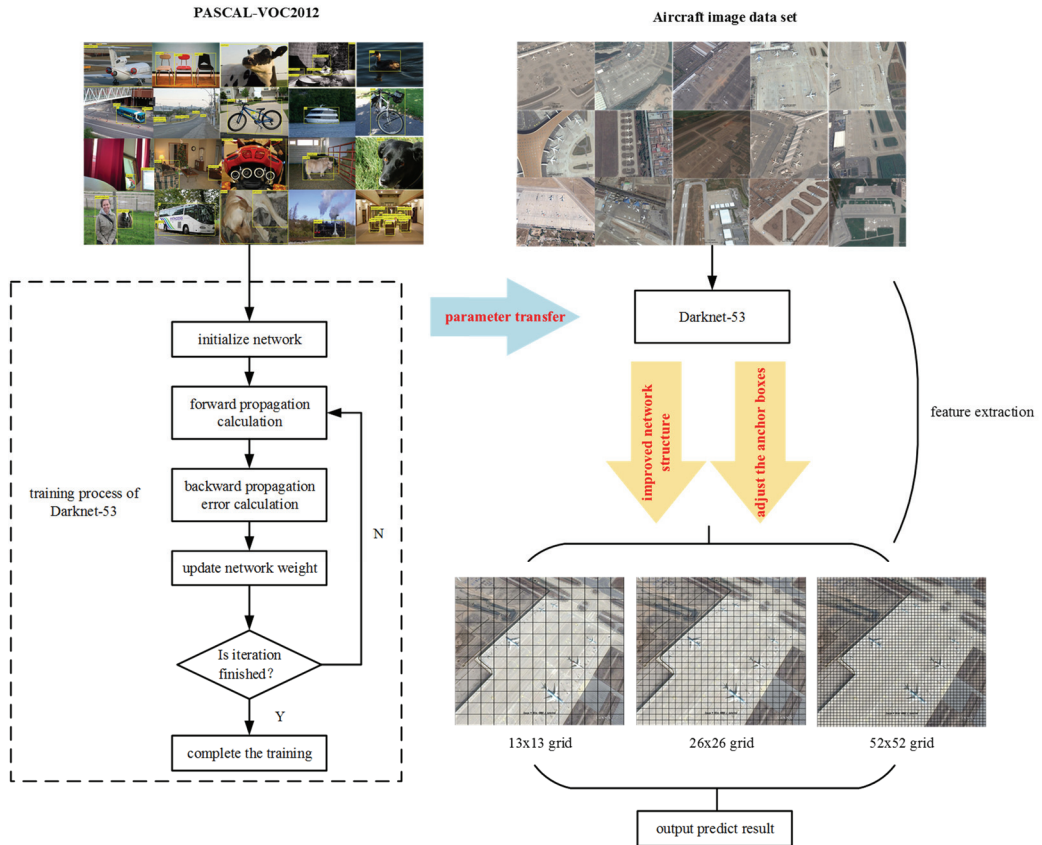
Fig. 2. Flow chart of aircraft detection in airport scene.

3.1 Data Clustering Analysis

Yolov3 uses anchor boxes as prior boxes to detect all targets in the image. Yolov3 uses k-means algorithm to cluster the target frame size of data set. In k-means algorithm, Euclidean metric, Manhattan distance, and Chebyshev distance are usually used as distance measurement to calculate the distance between two points. IOU is a standard for measuring the accuracy of corresponding objects in a specific data set in object detection. The main purpose of setting the prior boxes is to make the prediction box and the ground truth IOU better; thus, leading to the use of these common distance on our sample does not produce good results, such as using the Euclidean metric will make big bounding boxes than small bounding boxes to produce more errors, and we hope that we can through the anchor boxes to obtain good IOU, and IOU is independent of the size of the box. Therefore, the distance formula used in this paper is:

$$d(box, centroid) = 1 - IOU(box, centroid) \tag{1}$$

In the formula, $centroid$ represents the center of cluster, box represents the sample, and $IOU(box, centroid)$ represents the Intersection-over-Union of cluster center box and cluster box. The Avg IOU objective function of clustering is as follows:

$$argmax \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_k} IOU(box, centroid)}{n} \tag{2}$$

In the formula, n_k represents the number of samples in the k-th clustering center, n represents the total number of samples, and k represents the number of clusters.

The authors of Yolov3 cluster VOC data sets to get specific values (10,13), (16,30), (33,23), (30,61), (62,45), (59,119), (116,90), (156,198), (373,326), Among the 20 kinds of targets in VOC dataset, the size of targets ranges from bicycles and buses to birds and cats. Some anchors are unreasonable in this data set. To improve the bounding box's detection rate, we should change the value of anchor.

In this paper, k=1–20 is selected to cluster the samples in the data set, and the relationship between k and Avg IOU is obtained as shown in Fig. 3 With the increase of k value, the change of objective function becomes more and more stable. When k value is greater than 9, the curve starts to become smooth, so selecting 9 anchor boxes can not only accelerate the convergence of loss function, but also reduce the error brought by candidate boxes. In the experiment of this paper, the corresponding anchor boxes of the aircraft sample set is (13,14), (20,19), (22,27), (28,26), (31,33), (40,36), (35,45), (48,48), (62,66).
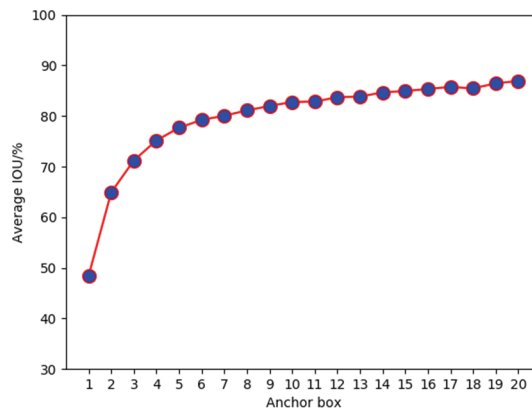
Fig. 3. The relationship between the number of Anchor box and the Average IOU.

3.2 Inc-Yolov3 Detection Model

In order to optimize the indicators and realize efficient remote sensing aircraft detection, this paper proposes Inc-Yolov3 to identify aircraft. Embedding Inception modules in Yolov3 to get the Inc-Yolov3, the following will describe the multi-scale prediction of images by the model and how to combine Inception modules with it.

Fig. 4 shows the multi-scale prediction method of Yolov3. The Inc-Yolov3 model also adopts the multi-scale prediction method and introduces FPN structure. Meanwhile, the features of different layers were fused through up-sampling, and three scales were used for prediction. When inputting images in the training process, the random multi-scale input method is used to train the detection network model so that the model can be robust to remote sensing image detection at different scales.

The Inception module is introduced between the convolutional set and 3×3 convolution in Fig. 4 to form the network Inc-Yolov3. The introduction method is shown in Fig. 5. Inception structure has two main functions: one is to use 1×1 convolution to increase and decrease dimension; the other is to use convolution reassembly on multiple sizes at the same time. Inception structure is introduced to stack more convolution in the same size sensor field to extract richer features, which is very beneficial for remote sensing image object detection. At the same time, the convolution of 1×1 is used to reduce the dimension,

which reduces the computational complexity. Inception makes four branches of input, convoluting or pooling with filters of different sizes, and finally splicing features together. This structure convolutes simultaneously on multiple scales intuitively, and can extract features of different scales. This structure uses the principle that sparse matrix is decomposed into dense matrix to accelerate the convergence rate.

We use the new anchor boxes generated by clustering in Inc-Yolov3 and improve the resolution of the input network. Finally, we get the optimized version of Inc-v1-Yolov3. This method can quickly converge and extract more detailed features before detection, and improve the classification accuracy.
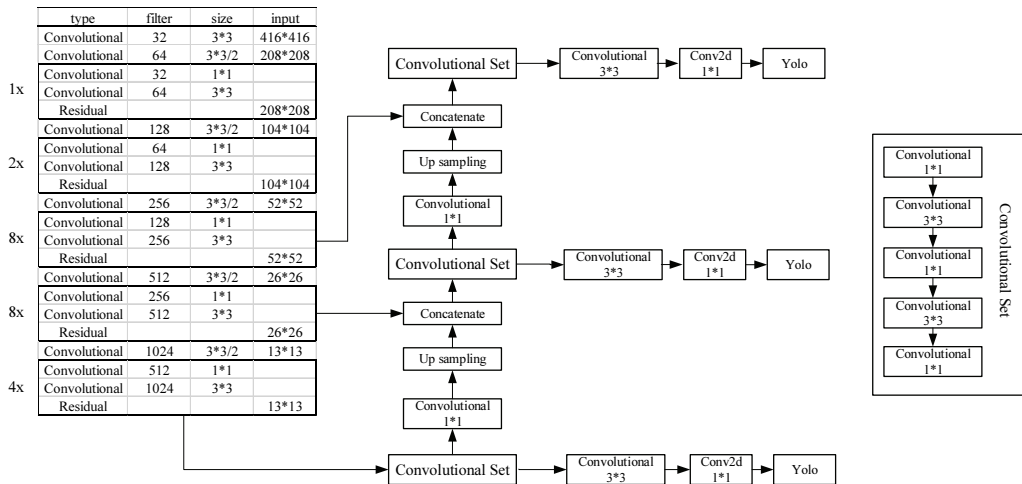

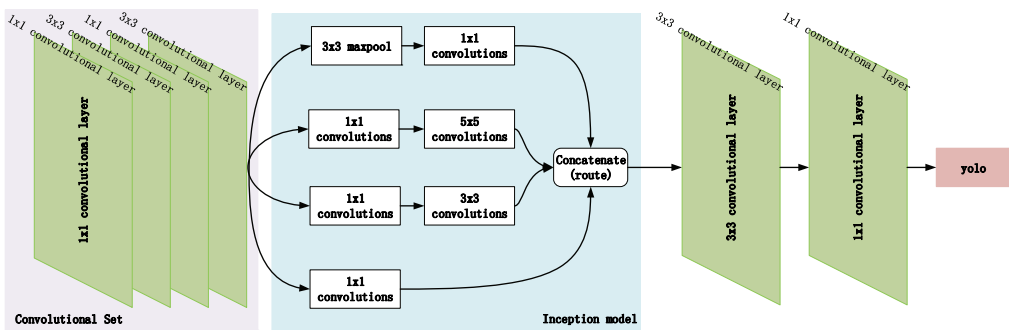**Fig. 4.** Yolov3 multi-scale prediction.

Fig. 5. Schematic diagram of Inception module introduction.

4. Experimental Design

4.1 Datasets

The data of this paper were collected from RSOD-Dataset, which was annotated by the team of Wuhan University, including four types of targets: aircraft, playground, overpass, and oil barrel. The annotation format of the dataset is consistent with that of PASCAL VOC2012 of the public dataset. The annotation information of the corresponding image is stored in an XML file. The public download address of RSOD-

Dataset is https://github.com/RSIA-LIESMARS-WHU/RSOD-Dataset-. The 446 remote sensing images of airports in different weather conditions and different regions were taken from RSOD-Dataset. Each image was cut by Google Earth software, and all aircraft objects were manually marked. The image size was about 1100×900 pixels. For each aircraft object, use the smallest circumscribed rectangle. Each boundary box is represented by the coordinates in its upper left and lower right corners (x1, y1, x2, y2), marking a total of 4,993 aircraft targets. The data set was divided into a ratio of 4:1, and 356 images were randomly selected as the training set and 90 images as the test set.

4.2 Experimental Setting

In this experiment, TensorFlow framework was used. In the training stage, the momentum adopted was 0.9, and use the asynchronous mini-batch gradient descent (MBGD) method to make the nearest gradient weight higher and alleviate the problem of ill curvature. The initial learning rate was 0.001, and the attenuation coefficient was 0.0005. When the number of training iterations was 30,000 and 35,000, the learning rate was reduced to 0.0001 and 0.00001. This method of reducing the learning rate by a certain proportion after the fixed number of iterations is a discrete learning rate change strategy, which makes the model converge quickly.

In order to make the model stronger, different scale training methods are adopted in this paper. After 64 batches of training, a new image size is randomly selected for training, effectively improving memory utilization. The image size ranges from 320 × 320 to 608 × 608, and the sampling interval is 32. At the same time, adjust saturation, exposure and tone to enhance the training sample.

4.3 Evaluation Methods

The Loss curve, IOU curve, precision-recall (P-R) curve, mean average precision (mAP) and F1-score values in the experimental results in this paper are used to evaluate the object detection model.

The change curve of loss can reflect the error between the predicted result and the actual result. Therefore, the faster the loss curve decreases, the smaller the loss value, which means the model training is better.

IOU is a standard for measuring the accuracy of corresponding objects in a specific data set during object detection. The IOU represents the degree of overlap between the generated candidate box and the original marker box, which is the ratio of intersection and union. The calculation formula is as follows:

$$IOU = \frac{DetectionResult \cap GroundTruth}{DetectionResult \cup GroundTruth} \tag{3}$$

$DetectionResult$ means the result obtained through neural network, $GroundTruth$ means the result of the original tag box.

In deep learning, if the actual situation and predicted results are positive with TP, it is called true positive. If the actual situation is positive and the prediction result is negative, it is called false negative and represented by FN. If the actual situation is negative and the prediction result is positive, it is called false positive and represented by FP. If the actual situation and predicted result are both negative, it is called true negative and represented by TN.

P-R curve is often used to measure the performance of the classifier. The calculation formulas of $Precision$ and $Recall$ are as follows:

$$Precision = \frac{TP}{TP+FP} \times 100\% \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{5}$$

F1-score is one of the indicators to evaluate the performance of the model that balances and integrates recall rate and accuracy. The calculation formulas of $F1\text{-}score$ is as follows:

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision+Recall} \times 100\% \tag{6}$$

5. Experimental Results

5.1 Experimental Results and Analysis

Fig. 6 shows the Loss value convergence curves of Yolov3, Inc-Yolov3 and Inc-v1-Yolov3 of the central plains network in the training process. The horizontal axis represents the number of training iterations, the maximum iteration is 40,200, and the vertical axis represents the loss value. As can be seen from Fig. 6(a) when the iteration exceeds 30,000 times, the values of each parameter are basically stable. Compared with the original network Yolov3, it was found that from 2,000 to 3,000 times, Inc-Yolov3 and Inc-v1-Yolov3 converged rapidly, and the final loss value decreased to about 0.18. Fig. 6(b) shows a comparison chart of loss curves obtained by narrowing the scope of loss and enlarging step. It can be clearly seen that the decrease trend of loss of Inc-Yolov3 is superior to that of Yolov3, while that of Inc-v1-Yolov3 is superior to that of Inc-Yolov3. From the convergence of the parameters, the training results of Inc-v1-Yolov3 network are relatively ideal.


(a)   (b)
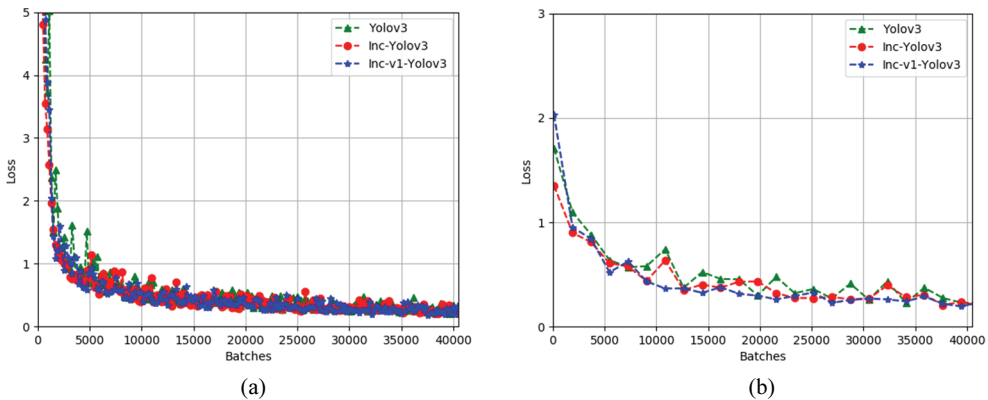
Fig. 6. Loss curve of different methods on the dataset: (a) loss 0–5, step=80 and (b) loss 0–3, step=700.

The IOU curve of Inc-Yolov3 is shown in Fig. 7. It can be seen from the graph that IOU gradually approaches 1 and finally stabilizes above 0.8.

Fig. 8 shows the results of the Inc-v1-Yolov3 boxed test on the test set. In each image, the blue rectangle represents the real object labeled by the dataset, and the purple rectangle represents the final detection result of our algorithm. As can be seen from Fig. 8, most aircraft targets can be detected correctly. It is

worth noting that our algorithm can detect aircraft that are not marked (artificially omitted) on the data set, as shown by the green arrow in Fig. 8, which includes half of the aircraft in Fig. 8(c) and the aircraft in Fig. 8(d) under strong light conditions.
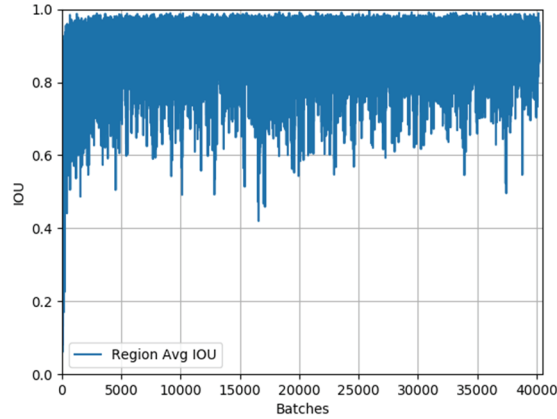
Fig. 7. IOU graph of Inc-Yolov3.


(a)   (b)   (c)   (d)

Fig. 8. Experimental results of Inc-v1-Yolov3 on test set: (a, b) smaller aircraft detected, (c) half-price aircraft detected, and (d) aircraft detected in high light.

Fig. 9 shows the P-R curves of different methods on the dataset in this paper. Compare the area enclosed by different curves and the horizontal and vertical axes, Inc-Yolov3 obtained the best performance. Compared with Yolov3, the detection effect of Inc-Yolov3 is obviously improved, which indicates that the improved network structure in this paper is effective. Comparing with Inc-Yolov3 and Inc-v1-Yolov3, Inc-Yolov3 can improve recall rate with guaranteed accuracy, while Inc-v1-Yolov3 can improve recall rate with guaranteed recall rate. Both curves are better than the original model Yolov3, but the key point is that Inc-v1-Yolov3 provides a better balance.

Fig. 10 is a bar statistical chart comparing the various evaluation indicators. In terms of recall rate, Inc-v1-Yolov3 is 4.57% higher than that of Inc-Yolov3 algorithm, and in terms of accuracy rate, it is 0.31% lower. In this case, F1-score is needed for overall measurement. Experimental results show that the F1-score of Inc-Yolov3 after Inception module is added into the network structure in this paper is 0.24% higher than that of Yolov3, and the F1-score of Inc-v1-Yolov3 after k-means clustering algorithm and improved input resolution is 2.50% higher than that of Yolov3. Therefore, Inc-v1-Yolov3 algorithm in this paper is obviously superior to Yolov3 algorithm.

Yao [9] proposes to reduce the difficulty of target border regression based on clustering idea, and cluster FPN to make its detection performance better. Zhou [10] improves the object location network of

CNN model based on Fast R-CNN, and proposes soft non-maximum suppression (Soft-NMS), to replace the original NMS, which is used to solve the problem of aircraft missed detection when the object is close to each other or occluded with each other. Table 1 shows that mAP of Inc-v1-Yolov3 proposed in this paper are higher than the above two methods.
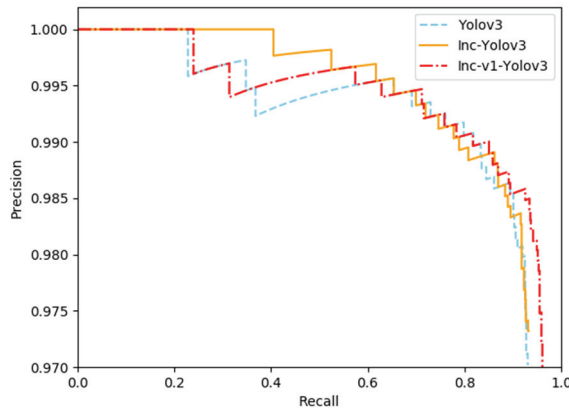
Fig. 9. Precision-Recall curves on data sets by different methods.
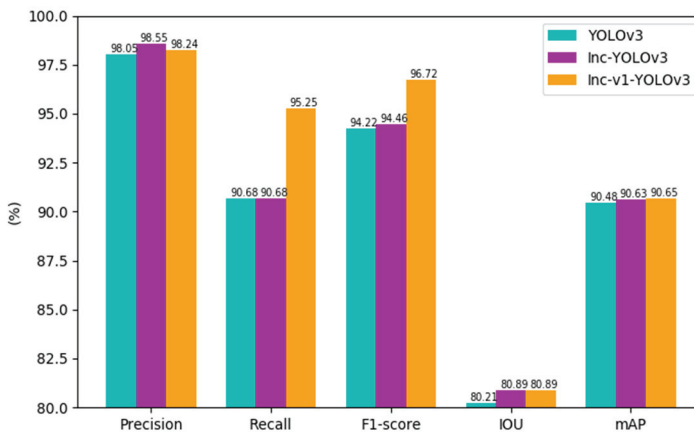
Fig. 10. Bar comparison graphs of Precision, Recall, F1-score, IOU and mAP of different methods on the data set.

Table 1. Comparison of mAP values of different methods on RSOD-Dataset

Method	mAP (%)
FPN+K-means++ [9]	89.30
Soft-NMS [10]	89.72
Inc-v1-Yolov3	90.65

   In this paper, 90 high quality remote sensing images were tested, and it was found that the accuracy of the aircraft in high quality remote sensing images was very high, and it also had a good recognition effect in the dense areas of the aircraft. The detection effect and detection accuracy were shown in Fig. 11, and the contrast areas were marked with yellow boxes. From Fig. 11(a) and 11(b), we can find that Inc-v1-

Yolov3 detected more and smaller targets that were not detected by the original network. It was found that the Inc-v1-Yolov3 also performed well in detecting incomplete targets. For example, the half-price aircraft was detected in Fig. 11(c), and the incomplete target whose detection rate was 91% was increased to 100%. Similarly, compared with Fig. 11(d) box selection, the detection accuracy of a single aircraft has been improved from 57% to 100%, which clearly shows that the aircraft detection accuracy of Inc-v1-Yolov3 is higher than that of the original network, and the situation of missed detection has been improved.

Dai et al. [11] proposes a region-based, fully convoluted R-FCN network for accurate and efficient object detection. In order to locate the object accurately, Long et al. [12] propose a new object localization framework named CNN-based, which can be divided into three processes: region proposal, classification, and accurate object localization process. Table 2 shows that both precision and recall of Inc-v1-Yolov3 proposed in this paper are higher than the above two methods.
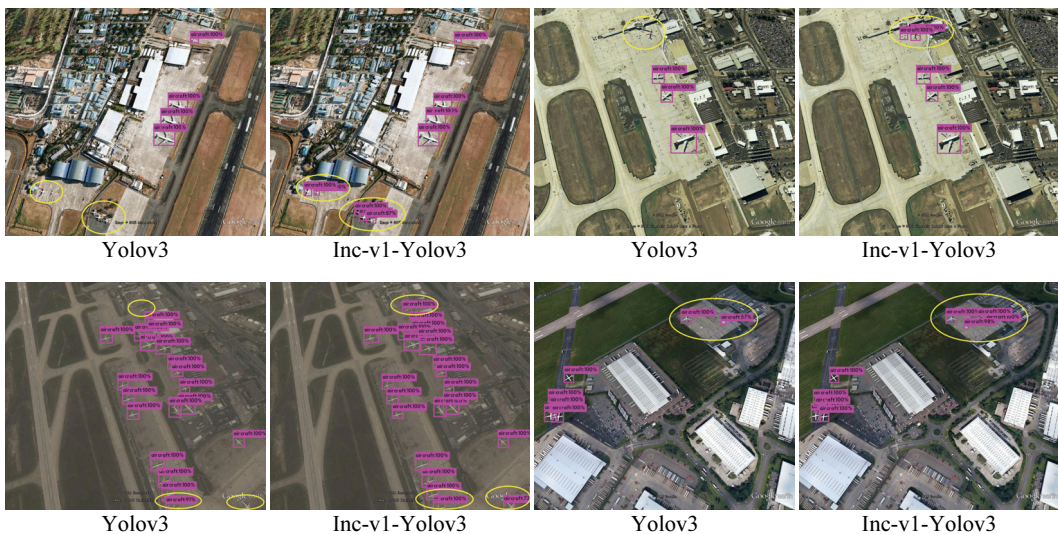
Fig. 11. Comparison of object detection effect between Yolov3 and Inc-v1-Yolov3.

Table 2. Network performance comparison of different methods

| Method	Precision (%)	Recall (%)
R-FCN [11]	94.37	93.44
CNN-based [12]	96.88	94.99
Inc-v1-Yolov3	98.24	95.25

Inc-Yolov3 has a high accuracy and recall rate, which indicates that the model in this paper has a strong generalization ability and there is no overfitting of the model. Because when training a deep neural network, all parameters of the model are not randomly initialized, but the model is first trained on PASCAL-VOC2012 data set according to the transfer learning to obtain the pre-training parameters, and then the parameters are taken as the initial value of the model. Since the pre-training model has convergent on large data sets and has good performance, fine-tuning it with aircraft training data can not only ensure the robustness and generalization of the model, but also improve the adaptability of the model to aircraft, thus showing high recall rate and accuracy.

5.2 Extend the Experiment

Since the structure of the model is modified in this paper, it is not only suitable for the aircraft target data set, so we can extend the model to other ground object recognition and detection. We applied the Inc-v1-Yolov3 method to other two targets in the RSOD-Dataset, playground and oil tank.

The size of the dataset of oil tank targets is generally similar to that of aircraft targets, and some of them are larger than aircraft targets. Through the experimental test, we found that the Inc-v1-Yolov3 also showed a good effect in detecting the oil tank dataset of RSOD-Dataset. After 40,000 iterations, the recall rate was 97.96%, the map value was 90.91%, and the F1-score was as high as 97.30%. The test results are shown in Fig. 12(a). Different from the data sets of aircraft targets and oil tank, the playground is a relatively large target in the whole remote sensing image, with a small number of targets in each image.

Meanwhile, most of the targets are also in the center of the image, which are easier to be detected. Through experimental tests, we found that on the playground dataset of RSOD-Dataset, after 40,000 iterations, the recall rate and map value had reached 100% satisfactory effect, and the F1-score was also as high as 98.41%.The test results are shown in Fig. 12(b).

(a)

(b)
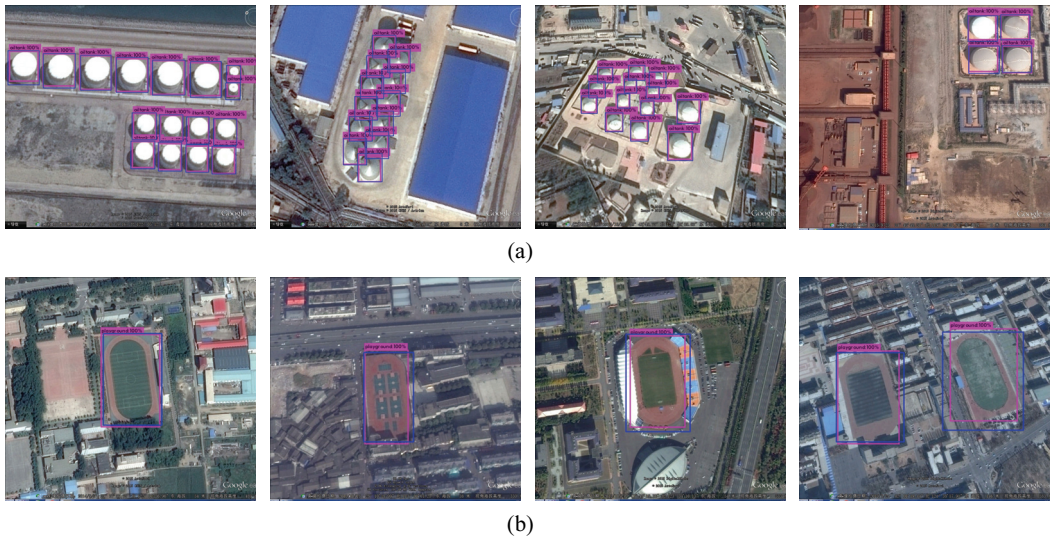
Fig. 12. Experimental results of Inc-v1-Yolov3 on other test sets: Inc-v1-Yolov3 object detection results on the RSOD-Dataset oil tank test set (a) and playground test set (b). The blue rectangle represents the real object marked by the dataset, and the purple rectangle represents the final detection result of Inc-v1-Yolov3 algorithm.

6. Conclusions

Aiming at the poor evaluation indexes such as precision and recall rate in Yolov3 detection of aircraft in remote sensing images, the Inc-Yolov3 network was proposed. On the basis of Yolov3 network, the network structure was optimized by adding Inception module and multi-scale prediction. According to the experimental analysis in this paper, the reasons for the performance improvement are as follows: (1) The pre-training model is used as the initial training value. Fine-tuning the aircraft training data set can not only ensure the robustness and generalization of the model, but also improve the adaptability of the

model to the aircraft, so as to show high recall rate and accuracy. (2) The original network structure and anchor box are only suitable for conventional targets, while the optimized network structure and re-clustering anchor box are more suitable for aircraft detection in the airport scene, which ultimately improves the detection accuracy of the model and reduces false detection. (3) In the optimized network structure, the number of network layers is deeper than that of the original network, which increases the network parameters and leads to a decrease in the detection speed, but does not affect its detection ability.

Although compared with the original method, the performance of this paper is improved greatly, but there are still some shortcomings. The size of the data volume is limited, and capturing more data will have a greater performance benefit. The following work can be optimized for such problems as less data, introduction of interference background into target labeling, and how to improve the detection speed under the premise of ensuring accuracy and Recall rate.

Acknowledgement

References

[1] M. Aamir, Y. F. Pu, Z. Rahman, W. A. Abro, H. Naeem, F. Ullah, and A. M. Badr, "A hybrid proposed framework for object detection and classification," *Journal of Information Processing Systems*, vol. 14, no. 5, pp. 1176-1194, 2018.

[2] M. Aamir, Y. F. Pu, Z. Rahman, M. Tahir, H. Naeem, and Q. Dai, "A framework for automatic building detection from low-contrast satellite images," *Symmetry*, vol. 11, article no. 3, 2019.

[3] J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," 2018 [Online]. Available: https://arxiv.org/abs/1804.02767.

[4] S. Sindhu Ramachandran, J. George, S. Skariam and V. V. Varun, "Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans," in *Proceedings of SPIE 10575: Medical Imaging 2018: Computer-Aided Diagnosis*. Bellingham, WA: International Society for Optics and Photonics; 2018.

[5] Y. L. Chang, A. Anagaw, L. Chang, Y. C. Wang, C. Y. Hsiao, and W. H. Lee, "Ship detection based on YOLOv2 for SAR imagery," *Remote Sensing*, vol. 11, article no. 786, 2019.

[6] X. Qian, S. Lin, G. Cheng, X. Yao, H. Ren, and W. Wang, "Object detection in remote sensing images based on improved bounding box regression and multi-level features fusion," *Remote Sensing*, vol. 12, article no. 143, 2020.

[7] Y. Zhang, H. Yang, and X. Liu, "Research on remote sensing image object detection method based on densely connected multi-scale features," *Journal of China Academy of Electronics and Information Technology*, vol. 14, no. 5, pp. 530-536, 2019.

[8] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Artificial Neural Networks and Machine Learning – ICANN 2018*. Cham: Springer, 2018, pp. 270-279.

[9] Z. Yao, "Research on the application of object detection technology based on deep learning algorithm," Ph.D. dissertation, Beijing University of Posts and Telecommunications, Beijing, China, 2019.

[10] T. Zhou, "Research on object detection based on deep convolutional neural network," Ph.D. dissertation, Harbin Institute of Technology, Harbin, China, 2019.

[11] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: object detection via region-based fully convolutional networks," *Advances in Neural Information Processing Systems*, vol. 29, pp. 379-387, 2016.

[12] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 5, pp. 2486-2498, 2017.

Lu Chen https://orcid.org/0000-0002-7330-9741

She received the B.E. degree in network engineering from Henan University, in 2017. She is currently pursuing the M.E. degree in Computer Application Technology from Henan University, Kaifeng, Henan, China. Her research interests include deep learning, object detection and image recognition.

Liming Zhou https://orcid.org/0000-0001-8741-0827

He received the Ph.D. degree in State Key Laboratory of Networking and Switch Technology from Beijing University of Posts and Telecommunications in 2015. He is an associate professor in the School of Computer and Information Engineering, Henan University, from 2015. His research interests include deep learning, artificial intelligence and information security.

Jinming Liu https://orcid.org/0000-0002-5650-8028

He received the B.E. degree in Computer science and technology from Henan University, in 2018. He is currently pursuing the M.E. degree in Computer Technology from Henan University, Kaifeng, Henan, China. His research interests include machine learning, deep learning and image target recognition.