Revista Argentina de Clínica Psicológica
2021, Vol. XXX, N°2,293-304
DOI: 10.24205/03276716.2020.4028

A Remote Sensing Target Recognition Method based on Generative Adversarial Networks

Bin LIU [1, 2, 3, 4], Hui WANG [5 *], Yixuan ZHAO [5], Kun FU [1]

Abstract
The traditional target detection algorithm uses the method of multi feature combination for recognition, which has the problems of poor anti-noise ability, weak robustness and long time consuming. In order to solve these problems, we first design a conditional generation adversarial network (GAN), and then use PCA algorithm to reduce the dimension of the feature vector, so as to retain the most distinguishing features in the sample and complete the feature extraction of remote sensing target. This paper also proposes a recognition algorithm based on template matching, which matches the processed image segmentation results with the standard template to get the type of object. Finally, we test the proposed method in a remote sensing aircraft image data set, and the accuracy reaches 95.86%, which is higher than other comparison algorithms.
Keywords: generation adversarial network (GAN); target recognition; template matching

1 Introduction

With the rapid development of remote sensing technology, a variety of imaging methods and different spatial resolution remote sensing platforms are constantly emerging, resulting in a large number of remote sensing images. There are also a large number of ships, bridges, aircraft and other targets with large aspect ratio. In recent years, the target detection framework via deep learning has been improved continuously, and the problem of large aspect ratio target detection has gradually attracted researchers. For this kind of target detection problem, the current research mainly uses the detection frame with orientation information to replace the horizontal rectangular detection frame to describe the attitude of the large aspect ratio target, and orientation information extraction is an important part of the large aspect ratio target detection task. Researchers have designed many remote sensing large aspect ratio target detection frameworks.

Reference [1], a rotating ROI pooling layer is constructed to extract the target features, improve the non-maximum suppression process, and finally regress the ship target detection results. Reference [2], a ship target detection model based on R2CNN++ is constructed. By adding feature pyramid fusion space, attention mechanism and improved angle loss function, ship target rotation detection is realized. Reference [3], a two-stage angle regression method based on dense connected feature pyramid is proposed to solve the problem of ship rotation detection. This method generates rotation candidate frame with angle in the candidate frame extraction stage, and regresses the angle of the target in the final prediction stage, so the overall detection performance is better.

These methods add orientation information prediction module in the detection network terminal, which is juxtaposed with the target scale and category prediction, and is obtained through a forward propagation of the network. The accuracy of individual orientation prediction needs to be improved. At the same time, remote sensing data has strong professional characteristics, mainly for target recognition needs rich professional knowledge and experience, so it is very difficult to obtain labeled Remote sensing data. At present, the lack of public

1. Aerospace Information Research Institute, Chinese Academy of Sciences, 100094, China
2.University of Chinese Academy of Sciences, 100094, China
3. School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Science, 100049, China
4. Key Laboratory of Network Information System Technology, School of Electronic Chinese Academy of Sciences, 100094, China
5. Bei Hang University, State Key Laboratory of Software Development Environment , 100191, China
*Correspondence author: Hui WANG. Email: huiwang_bh@163.com

data sets brings challenges to the remote sensing image target recognition algorithm. Many image target recognition algorithms of remote sensing use a small data set for training and testing, but in this case, the generalization ability of the algorithm is weak and can not be used in the actual scene. Therefore, we choose unsupervised learning to solve the task of remote sensing image recognition, and propose a method based on template matching to complete the classification of target objects.

2 Background

2.1 Key points detection algorithm base on CNN

Remote sensing target has relatively fixed shape information. We can use key points to describe it, as shown in Fig.1.For aircraft target, we define its key points as the four intersection points of nose, tail, both ends of wing and fuselage. By connecting different key points in sequence, we can obtain rough semantic results of remote sensing targets. Although the result is rough, it still provides the basic shape, position, docking angle and other information of the aircraft. These data are sufficient for the subsequent conditions to GAN, which can better simulate the remote sensing target and extract the target features.

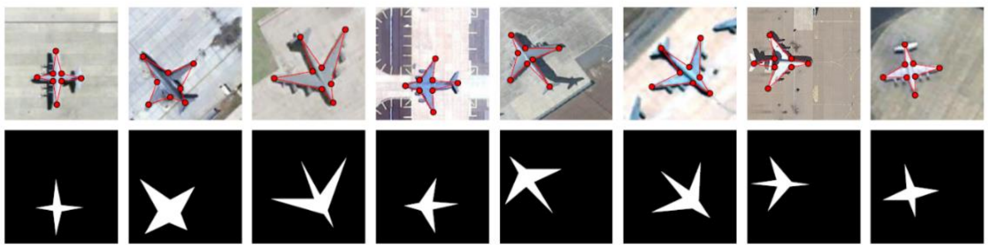

Fig.1 data set of aircraft key points detection

Some key points[4] detection algorithms transform detection into regression problem. That is to say, the depth convolution neural network is used to directly regress the key points of the target. Limited by the convolution network down-sampling strategy and the fully connected layer as the regression coordinate, the two-dimensional structure information of the image and the target has a large loss, which affects the prediction accuracy of the key points. In the method [5,6], they can only detect the target on the image of 40×40.

We detect the key points of 256×256 image. If the image is scaled to 40 × 40, and then the positioning result is mapped to the original image, the error will be introduced. If the direct regression method is used to locate the key points on the 256×256 image, the positioning error will be caused by the above problems. Therefore, we use the heat map to detect the key points.

Heat map is usually used for human pose estimation in natural images. In this method, convolution neural network model is used to generate probability map with the same size as the original image. Each value in the probability map represents the probability that the position belongs to the key point to be detected. This method avoids the limitation of fixed length output of regression method, and can deal with the situation of multi-target in the image. However, when this method is applied to remote sensing images, it is difficult to detect the symmetry points because the remote sensing images have obvious symmetry and rotation, and the remote sensing objects may dock at any position of the image at any angle. Some heat map detection methods detect each key point separately, and then combine them. In the application of remote sensing image detection, a key point is often located in a symmetrical position, resulting in large detection error. In this case, we design a symmetric point method. The specific method will be introduced in the next chapter.

2.2 Feature extraction method based on GAN

Most of the existing methods[7,8,9,10] of target recognition using GAN change the feature map learned by the discriminant network into feature vector directly, and then use the classifier for recognition. Although, benefiting from the algorithm designed in the previous paper, the discriminant network has better overall extraction ability, the noise generated by background information will still interfere with the representation ability of features. Therefore, on this basis, we design a ROI pooling feature extraction method to further simplify the feature dimension, eliminate the influence of background information, and further enhance the ability of feature representation. The specific flow of the algorithm is shown in Fig.2.

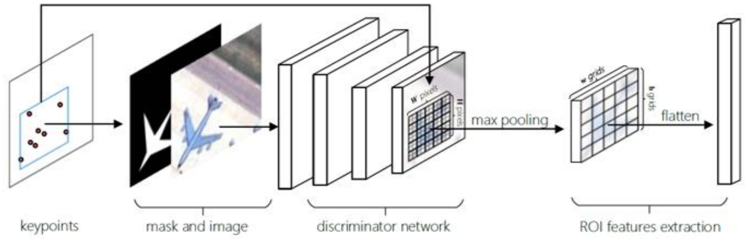

Fig.2 ROI pooling algorithm

First, we still draw the minimum external rectangle of the target according to the detected key points, and map it to the feature map as ROI region according to the scale. Then, we will divide ROI area in the feature map into a fixed number of grids, namely, $W \times H$, and extract the maximum pixel value as the feature value in each grid, then a feature graph will obtain and they will form a feature vector. Then for a network, we will get a feature vector with length of $C \times W \times H$, where C is the number of eigengraph channels.

In addition, in order to further enhance the representation ability of feature graph, we will extract features of different scales of the target. Specifically, we will build multiple discriminant networks, one of which will use the original image as input, and the other network will use the image after sampling according to a certain proportion as input. After the training, we will use ROI pooling to extract features from multiple discriminant networks. Because different discrimination networks focus on the characteristics of targets at different scales, they have good complementarity and can further enhance the representation ability of features.

3 Design

3.1 Simultaneous detection of symmetry points

In view of the shortcomings of traditional neural network, we use the basic network structure called hourglass. The network structure model is shown in Fig.3. The advantage of hourglass structure is that it can map the network from the features learned from the bottom layer of the network to the high layer, and the bottom features can better retain the two-dimensional structure information of the image and the target. Therefore, the bottom structure features and high-level semantic features can be fully considered when the network finally predicts the heat map, and the key points can be predicted more accurately.

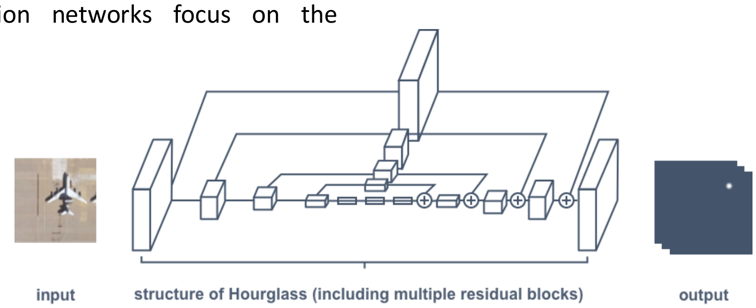

Fig.3 structure of aircraft key point detection network model

If the method of heat map is used, the symmetrical points in remote sensing objects are often detected by mistake. Therefore, the specific process of the joint detection method of symmetry points designed by us is shown in Fig.4.

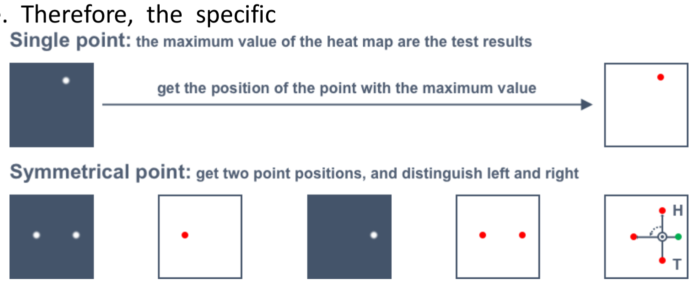

Fig.4 aircraft symmetry point joint detection algorithm

For a single key point, such as the nose and tail, we still use a separate heat map to predict, while for symmetric key points, such as the two wings, symmetric connection points, we will predict on the same heat map. Therefore, for a typical aircraft target, we will produce five prediction results, which are: nose, tail, wings, upper node (near nose) and lower node (near tail).

When we get the position of the point, we will get the maximum position of the heat graph for a single point, and for a symmetric point heat graph, we will first find the maximum point of the whole graph, and then set the point in a certain range near the value to zero, and then find the next global maximum.

Next, we will distinguish the symmetry points. First of all, assuming that the detection results of the nose and tail are correct, we can construct a vector from the tail to the head, and then connect two symmetrical points to form another vector. And we compute the outer product of the two vectors. According to the principle of mathematics, if the symmetrical point vector points from the left to the right, then the two vectors form a right-handed system, and the product is positive. On the contrary, the product is negative. In this way, we can easily judge the left-right relationship of symmetric points.

3.2 Design of GAN

According to the discussion in the previous chapter, only when the quality of the generated samples is close enough to the real data, the feature of the discrimination network will be more effective. Therefore, we need to choose the network structure with strong generation ability. Reference [11], the author proposes a multilevel generation model, which can guarantee the global similarity and detail of the generated samples at the same time. So, we use this network as the generation network model, and its structure is shown in Fig.5.

The generating network structure mainly consists of two parts. One part is called global network (G1), which down-samples the input image and sends it to the network structure with 8 residuals to learn global features. The other part is called local network (G2), which sends the original input image to the network to learn detailed features. The learning result is fused with that of G1 network and then sent to the back Continue the network to simulate the real sample. Because the network combines global and local features, the quality of the generated image is closer to the real sample.

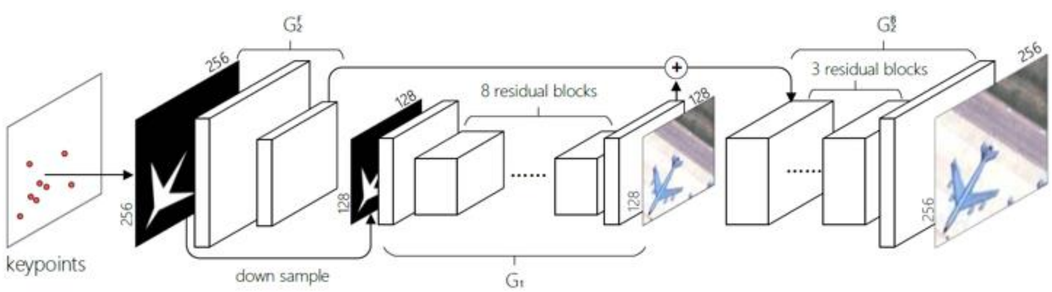

Fig.5 Generate network structure

For the discriminator, in order to facilitate the feature extraction and avoid over fitting, we choose a more common convolutional network and use the loss function defined by LSGAN [12] as the basic loss function.

（1）**input data**

Usually, the detailed semantic segmentation results of real images are used as condition inputs to learn the features of real samples. However, for remote sensing images, the amount of target data is too large, and it is difficult to achieve fine annotation of a large number of samples in a short time, so we use the mask image derived from the key points as the input of the generation network.

The specific steps are as follows: for a remote sensing target, we will use the algorithm designed in the previous section to extract the position of the key points of the target, and then connect these key points to generate the mask of the target. According to the position of the key points, the mask will be placed in an empty image as large as the original image (all pixel values are set to zero), and then the internal pixels of the mask will be set to 1. The single channel binary image generated by this method is used as the condition input of the condition GAN to participate in the training of the whole model points.

Although the generated mask image is relatively rough, it contains the key information of remote sensing target location, size, shape, angle and so on, which can ensure the quality of the generated samples. The experimental results also verify the effectiveness of the conditional input.

（2）ROI weighted loss function

According to the above, the generation of confrontation network will learn the features of the whole image, so a lot of background information will be extracted. And this part of information will seriously interfere with the information related to remote sensing targets, resulting in information redundancy, reducing the effectiveness of features, and even making subsequent classification algorithms tend to classify the samples with the same background into one category.

In order to make the GAN pay more attention to the target in the image, we design a loss function called ROI weighted, and its principle is shown in Fig.6.

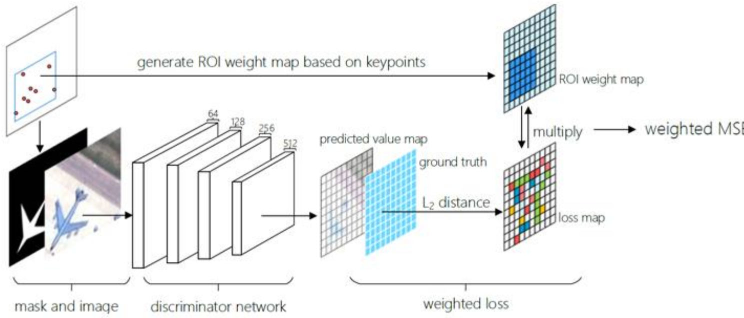

Fig.6 ROI weighted loss function

Specifically, for each input image, we will draw the circumscribed rectangle of the remote sensing target in the image according to the key points detected, and then the position of the circumscribed rectangle will be mapped to the loss layer called ROI region according to the scaling ratio of the feature map. We use the loss function defined in LSGAN as the basic loss function, which is the relationship between the result map and the true value Minimum mean square error. On this basis, we will increase the loss weight of ROI area. In order to increase the network's attention to the remote sensing target itself. The experimental results show that our loss function effectively increases the ability of feature representation and improves the accuracy of target recognition.

3.3 Remote sensing target recognition algorithm

We use the remote sensing target key detection algorithm to generate the rough semantic segmentation result of the image（mask）. The mask generated will be used as a condition input to train a generation of anti-network model, which can be used to extract the feature information of remote sensing target. In order to make the extracted information more refined and expressive, this paper designs ROI weight and ROI pooling feature methods to enhance the correlation between features and targets, and remove redundant information.

After the previous algorithm, the obtained features have strong representation ability, and can effectively distinguish and identify different types of targets. However, the feature dimensions obtained by the previous algorithm are still high, which contains a lot of redundant information, and will increase the time consumption of the model. Therefore, we will use PCA to reduce the dimension[13] of feature vector, in order to retain the most distinguishing features in the sample.

After that, we will use the classifier to recognize the samples. Because of the strong discrimination of the feature itself, we choose the simple linear kernel SVM [14]as the classifier to recognize the model of the target. Result show that our method can effectively improve the accuracy of target recognition. The flow of the algorithm is shown in Fig.7.

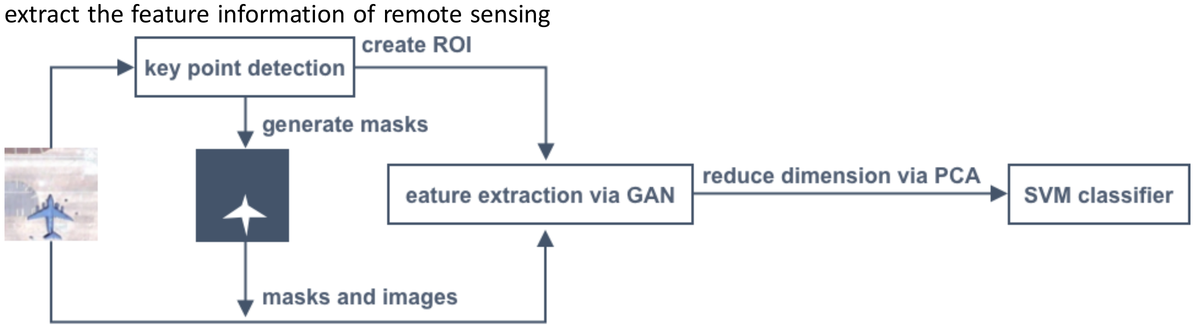

Fig.7 remote sensing target recognition algorithm

3.4 Aircraft model recognition algorithm based on template matching

Based on the results of aircraft key point detection and aircraft segmentation, we design the basic algorithm of aircraft type recognition. The basic flow of the algorithm is shown in Fig.8.

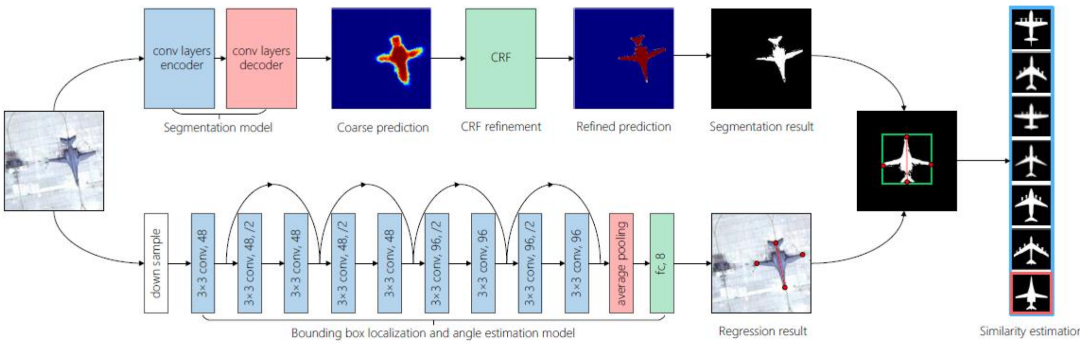

Fig.8 aircraft model recognition algorithm

For each aircraft slice to be identified, we first input the segmentation network to get the fine segmentation results of the aircraft target, then input the key point detection model to get the position of four key points of the aircraft, namely, the head, tail and two wings of the aircraft. Through these four key points, we can get the minimum external rectangle of the aircraft and calculate the aircraft docking angle. According to the docking angle and the minimum external rectangle, we can correct the aircraft segmentation results and cut off the background area. Finally, we can match the processed segmentation results with the standard template to get the aircraft model.

Choosing an appropriate method to measure the similarity between the segmentation results and each template is the key of aircraft recognition algorithm. And we use the index of IOU. The definition of IOU is shown in formula (1) The IOU index fully considers the size and shape of the aircraft, which can ensure that the segmentation results are fully used to distinguish the aircraft categories.

$$\mathrm{IOU} = \frac{TruePositive}{True\ positive + FalsePositive + FalseNegative} \quad (1)$$

The aircraft model recognition algorithm designed in this paper uses the key point detection and aircraft segmentation algorithm to obtain the aircraft standard template. In the training process of these two deep learning network models, there is no need to label the specific aircraft model, so there is no need to consider the uneven distribution of each category of aircraft. Thanks to the precision of aircraft segmentation results and the accuracy of aircraft key point detection results, the aircraft model recognition achieves a high accuracy.

4 Evaluation

4.1 Remote sensing target recognition experiment and result analysis

（1）data set

The data set of this experiment is divided into two part. The first part of the data set is 40000 aircraft slice images with key information. Each image has and has only one aircraft, but the aircraft may be parked at any angle. This part of the data set is used to train key detection networks and generate anti network models, of which 30000 images are used for training and 10000 images for testing. We extend the training set by rotating mirror. The second part is 21000 aircraft slice images with type annotation information, which contains 7 different aircraft samples. Similarly, each image has and has only one aircraft. 14000 images are used for training and 7000 images are used for testing.

（2）result and analysis

The target key detection module is used to extract the target key points, and the generated key points will be used to build input of the generated anti network, ROI weighted loss function, and ROI pooling feature. Therefore, the accuracy of key point detection will directly affect the performance of the whole model. This section mainly evaluates the performance of the key point detection module.

In this paper, we use hourglass structure to construct the key point detection network ,and design a joint detection method to correct the detection of symmetry point errors. In this paper, for example, in order to balance the error between large target and small target key point, we use an evaluation index called mean error, which formula is as follows:

$$mean\ error = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\sqrt{\left(p_x^{i*}-p_x^{i}\right)^2+\left(p_y^{i*}-p_y^{i}\right)^2}}{l_i^*}\right) \quad (2)$$

p_x^{i*}, p_y^{i*} represents the true value of this key point in the image, represents the detection value, and l_i^* represents the aircraft target fuselage length. Therefore, for larger aircraft with longer fuselage length, we allow the key point error to be larger, while for small aircraft targets, the key point detection standard is more stringent. In this way, our test results can be more close to the needs of practical application.

In order to evaluate the effectiveness of the proposed method, firstly, we compare the proposed model with the heat map model which detects 8 different key points separately, and the results are shown in table 1. It can be seen from the results that this method has made great progress compared with the original method, and the overall performance of key points has been improved, especially for the detection of symmetrical points, such as the wing. Due to the problem of false detection of symmetrical points solved by this method, the accuracy of wing points is greatly improved. In addition, the detection performance of other key points is similar, so the overall performance is improved.

Table 1 Comparison of the results of the joint detection of symmetrical points and the independent detection (unit:%)

Method	head	Intersection 1	left wing	intersection 2	tail	intersection 3	right wing	intersection 4	average
Our work	3.60	3.05	3.74	2.79	4.21	2.86	3.50	2.96	3.34
Separate detection of 8 points	3.96	2.92	4.14	2.73	4.50	2.80	4.36	2.94	3.54

In addition, we also compare this method with the method of directly using convolution network to regress the position of key points. The results are shown in table 2 in which we use ResNet-18 for regression network. Experimental results show that our method is better than direct regression. The main reason is that our network can integrate the low-level structural features and high-level semantic features, and can more accurately locate the key position.

The results are shown in table 2. Because the method [5] can only detect the key points on the image, we will zoom the image to, and then use the method in paper to detect, and then map it back to the original image. The experimental results show that the method[5] will introduce inevitable errors due to the need of image scaling operation, resulting in the decline of detection accuracy.

Table 2 Comparison of the effect of different key detection algorithms (unit:%)

Method	head	intersection 1	left wing	intersection 2	tail	intersection 3	right wing	intersection 4	average
Ourwork	3.60	3.05	3.74	2.79	4.21	2.86	3.50	2.96	3.34
ResNet	4.05	3.52	4.22	3.49	4.66	3.43	3.99	3.39	3.84
Method[5]	4.81	4.00	5.10	3.76	5.23	3.80	4.86	3.86	4.43

Finally, we randomly select some samples from the test sample set to show the actual effect of key point detection. As shown in Fig.9, we use different colors to display the detection results of different key points, in which yellow is the head of the aircraft, pink is the tail of the aircraft, the left point is represented by blue, and the right point is represented by green. It can be seen from the figure that this method can accurately locate each key point of the aircraft, and can effectively distinguish the left wing and right wing of the aircraft.

To sum up, this section describes the accuracy of the key point detection algorithm. In addition, we will build other experiments based on the experimental results in this section.

（3）effect display of GAN generated samples

As shown in Fig.10, this section shows some aircraft samples generated by GAN. It can be seen from the figure that the aircraft samples generated by GAN are close to the actual data, and contain rich details. GAN not only generates aircraft samples, but also learns the essential characteristics of the samples. Therefore, the high-quality generated samples also reflect the effectiveness of the features extracted by the discriminant network to a certain extent.

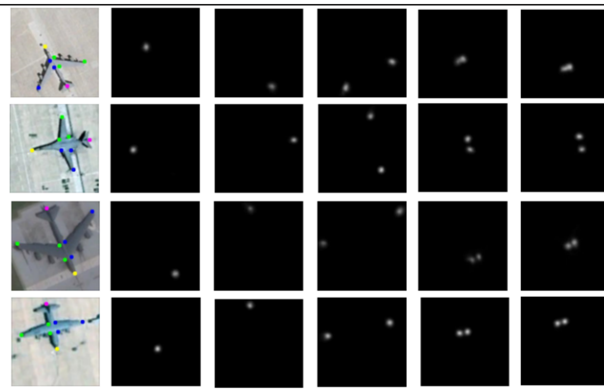

Fig.9 key point detection

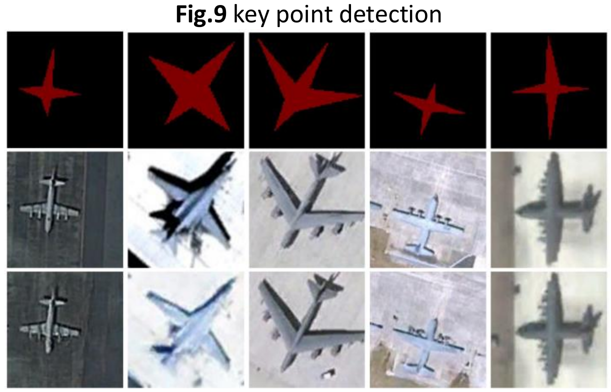

Fig.10 GANs generated image example. The first behavior is input label, the second behavior is real sample, and the third behavior is generated sample

(4) Target recognition results using ROI Pooling feature

In order to improve the representation ability of features on the target, and to remove the influence of background information. We designed ROI pooling characteristics. The specific method is to draw the minimum external moment of the target as ROI region on each image according to the key point detection results. Then, ROI area is divided into a certain number of grids ($C \times W \times H$) in the feature map, and the maximum value in each grid is extracted to form feature vector.

In the experiment, we extract the feature on the last feature layer of the discriminant network, and construct the feature vector. Because the feature vector is too long, there is more redundant information, which will slow down the speed of the algorithm. Therefore, we use PCA to reduce the dimension of the feature vector to 4096. Finally, we use a simple linear kernel SVM classifier for target recognition.

In this section, in order to verify the effectiveness of the feature extraction algorithm, we design the following comparative experiments

1. On the same feature graph, all the feature values are retained and formed into a feature vector. PCA is used to reduce the feature dimension to 4096, and the same classifier is used for recognition experiment.

2. On the same feature map, the whole feature map is pooled to extract features, PCA is also used to reduce the dimension to 4096, and classifier is used for target recognition.

The experimental results are shown in table 3 In the results, the results of feature extraction using pooling on the whole image are better than the results of retaining all the features, which indicates that retaining all the features will retain a lot of redundant and invalid information, and these information will submerge the effective information, which will cause difficulties to the subsequent algorithm The pooling method can extract more distinguishing features from the feature graph, filter out some useless information, and improve the representation ability and robustness of features. However, compared with the global pooling scheme, the proposed method is more effective, and the recognition rate is improved by about 5%. This is mainly because the algorithm in this paper extracts features directly from the target related areas, which can filter a lot of information only related to the background, reduce the impact of background information in the recognition process, so greatly improve the recognition accuracy.

Table 3 ROI Pooling feature effect assessment (unit:%)

method	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	average
All features	81.50	85.40	90.20	73.70	69.50	88.10	86.20	82.09
Global Pooling	83.30	87.60	92.40	76.50	70.20	89.60	88.70	84.04
ROI Pooling	84.70	97.80	97.60	78.30	76.80	96.00	96.60	89.68

To sum up, ROI pooling method can enhance the ability of feature representation and improve the accuracy of recognition algorithm. In the follow-up experiments, we use ROI pooling method for feature extraction.

（5）Target recognition results using multi-scale features

In order to further enhance the representation ability of features and enhance the representation ability of features for multi-scale information, we train multiple discriminant networks and input images of different sizes. In this way, multiple networks can learn features on different scales.

In the process of feature extraction, we will use ROI pooling method to extract fixed dimension features from three different networks, and then splice the three features to form the final feature. In the actual operation process $C_1 = W_1 = 5, C_2 = W_2 = 3, C_3 = W_3 = 2$, the sizes of the three network input images are $256 \times 256, 128 \times 128, 64 \times 64$. In the process of feature extraction. Finally, all feature vectors are reduced to 4096 by PCA and classified by linear kernel SVM.

Table 4 effectiveness evaluation of multi-scale features (unit:%)

method	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	average
256×256	84.70	97.80	97.60	78.30	76.80	96.00	96.60	89.68
128×128	83.50	95.10	97.00	72.20	71.10	75.70	75.60	84.31
64×64	83.40	89.70	95.50	74.60	73.50	90.80	80.50	84.00
Multi-Scale Features	92.40	98.80	99.40	88.90	86.60	96.00	94.90	93.80

The experimental results are shown in table 4. We test the effects of different discrimination model features separately. The results show that the recognition accuracy is similar by using the characteristics of a single discrimination network, and the recognition accuracy is similar, and the recognition accuracy has been greatly improved by integrating multiple features for model recognition.

To sum up, the use of multi-scale special diagnosis can effectively enhance the ability of feature representation and discrimination, and enhance the accuracy of target recognition. In the follow-up experiments, we use multi-scale features for target recognition.

（6）The results of target recognition using ROI weighted

In order to enhance the representation ability of generating features learned by the network, we designed the ROI weighted loss function, which makes the network model pay more attention to the regions related to remote sensing targets and ignore the information of background areas. At the same time, the combination of ROI weighted loss function and ROI pooling feature extraction can achieve better results.

Table 5 Effectiveness evaluation of weighted loss function (unit:%)

method	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	average
non weighted loss	92.40	98.80	99.40	88.90	86.60	96.00	94.90	93.80
weighted loss	95.30	99.20	99.50	92.60	83.80	95.10	99.80	95.04

In order to verify the effectiveness of the method, we compare the model trained by ROI weighted loss function with the model trained by ordinary loss function, extract features respectively, and carry out target recognition experiments, in which the weight is set to 8. The experimental results are shown in table 5. The experimental results show that using ROI weighted loss function can further enhance the ability of feature recognition and increase the recognition accuracy.

（7）Overall assessment

In order to evaluate the overall performance of the model, we compare the proposed method with the end-to-end neural network algorithm. We select ResNet-18 as the contrast neural network and carry out end-to-end recognition. Results are shown in table 6. The experimental results show that this method is better than the end-to-end neural network, mainly because this method can use a large number of samples without class labeling information for training, and can learn more essential and stronger performance characteristics. At the same time, the experimental results also show that the proposed method is more suitable for remote sensing target recognition problems with large amount of data and difficult category labeling.

Table 6 comparison of different aircraft recognition algorithms (unit:%)

method	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	average
Ourwork	95.30	99.20	99.50	92.60	83.80	95.10	99.80	95.04
ResNet-18	86.40	99.10	99.50	76.50	77.40	91.30	100.00	90.00

Because key points detection is the key step of the target recognition algorithm, in order to measure the influence of the key point detection results on the accuracy of aircraft target recognition, we use the true value of key points and the detection value of key points respectively to carry out aircraft model identification experiments. The experimental results are shown in table 7. From the data in the table, it can be seen that, thanks to the excellent effect of the key point detection algorithm proposed in this paper, the key point detection results have little influence on the accuracy of the overall aircraft model recognition algorithm. Compared with using the key point truth value, the recognition accuracy rate is only reduced by 0.39%.

Table 7 Influence of key point detection results on aircraft identification accuracy (unit:%)

method	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7	average
detection value	95.30	99.20	99.50	92.60	83.80	95.10	99.80	95.04
truthvalue	94.50	99.00	99.70	93.00	84.30	97.80	99.70	95.43

Finally, in order to measure the recognition effect of the aircraft model recognition algorithm in each category, we draw the recognition confusion matrix as shown in Fig.11.

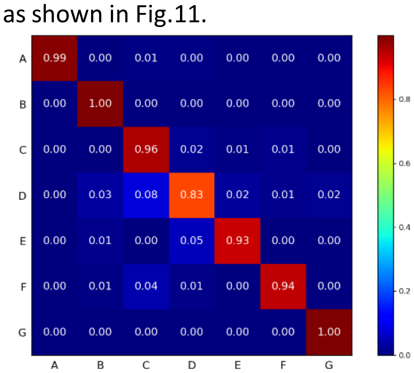

Fig.11 Confusion matrix of aircraft type identification

4.2 The experiment and result analysis of aircraft type identification

（1）data set

Because the aircraft recognition task needs to train the aircraft key point detection and aircraft segmentation depth learning model, the data used in this part is the same as the above training data, and only eight aircraft target key point positions are marked, and all 12000 images are used for training the model. In order to verify the effect of aircraft recognition, as shown in the Fig.12, we selected seven kinds of common aircraft, selected 100 images of each aircraft to build a test set, and made a standard template for each image.

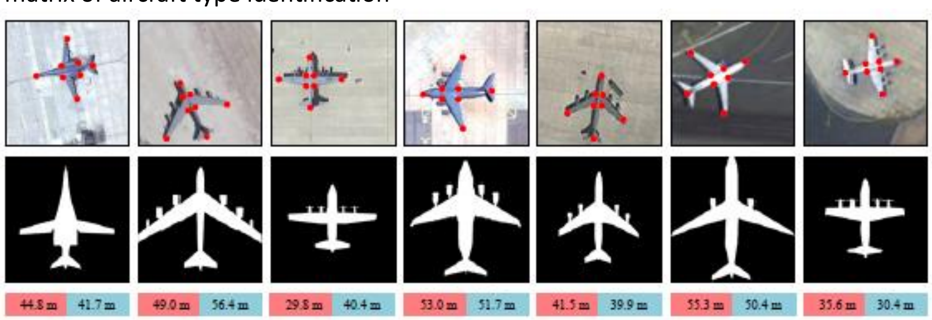

Fig.12 sample of aircraft target type identification test data set

(2) result and analysis

Firstly, we calculate the IOU values between the segmentation results of each category and the standard template. The results are shown in Fig.13. It can be seen from the figure that the IOU values between different categories are very distinguishable.

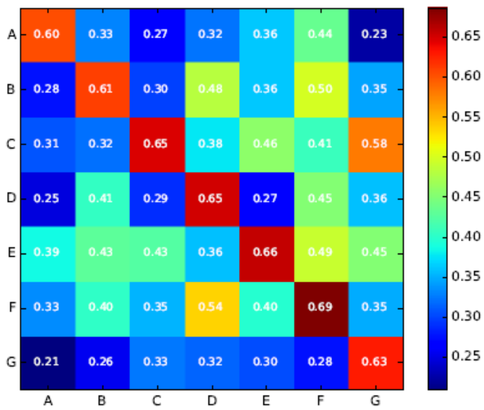

Fig.13 predicts the IOU value between the segmentation result and the standard template. The horizontal coordinate is the standard template, and the vertical coordinate is the segmentation result obtained by the experiment.

Next, we draw the confusion matrix of aircraft type identification, as shown in Fig.14. It can be seen from the confusion matrix that the classification accuracy of some types of aircraft reaches 100%.

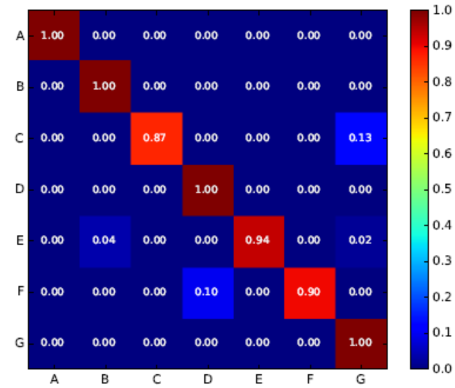

Fig.14. Confusion matrix of aircraft type identification. The abscissa is the true value and the ordinate is the prediction result.

We compare the aircraft model recognition algorithm designed in this paper with other existing methods. The comparison methods include both traditional template matching method and direct classification method. The comparison results are shown in table 8. It can be seen from the comparison results that this method effectively improves the accuracy of aircraft type recognition.

Table 8 accuracy of different aircraft recognition algorithms

method	Our work	DBN	traditional template matching algorithm	key point matching algorithm
accuracy	95.86%	91.14%	90.13%	93.29%

5 Summary

There are a lot of redundant background around the target concerned by remote sensing image. At present, the deep convolution neural network is often used in remote sensing target detection. The scale and center position of the target to be detected are recognized by a forward propagation method. The complex background information of remote sensing image will have a negative impact on the accurate prediction of target position. This paper designs a feature extraction method based on the generation of confrontation network, which reduces the cost of annotation and realizes the task of feature extraction of remote sensing image. At the same time, a template matching method is proposed to classify the target objects, and the method is applied to aircraft model recognition task, and good results are achieved.

Acknowledge

Foundation and Enhancement Technology Fund (2020-JCJQ-JJ-517)

Reference

[1] Liu Z, Hu J, Weng L, et al. Rotated region based CNN for ship detection[C]. international conference on image processing, 2017: 900-904.

[2] Yang X, Yang J, Yan J, et al. SCRDet: Towards More Robust Detection for Small, Cluttered and Rotated Objects[C]. international conference on computer vision, 2019: 8232-8241.

[3] Yang X, Hao S, Kun F, et al. Automatic Ship Detection in Remote Sensing Images from Google Earth of Complex Scenes Based on Multiscale Rotation Dense Feature Pyramid Networks[J]. Remote Sensing, 2018, 10(1): 132-133.

[4] Lowe D G. Distinctive image features from scale-invariant key points[J]. International journal of computer vision. 2004, 60 (2): 91–110.

[5] Zhao A, Fu K, Wang S, et al. Aircraft recognition based on landmark detection in remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters. Aug 2017, 14 (8): 1413–1417. doi: 10.1109/LGRS.2017.2715858.

[6] Zuo J, Xu G, Fu K, et al. Aircraft type recognition based on segmentation with deep convolutional neural networks[J]. IEEE Geoscience and Remote Sensing Letters. Feb 2018, 15 (2): 282–286. doi: 10.1109/LGRS.2017.2786232.

[7] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. International Conference on Neural Information Processing Systems. [S.l.], 2012: 1097-1105.

[8] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. Computer Science. 2015.

[9] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.], June 2015: 1–9. doi: 10.1109/CVPR.2015.7298594.

[10] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.], June 2016: 770–778. doi: 10.1109/CVPR.2016.90.

[11] Wang T, Liu M, Zhu J, et al. High-resolution image synthesis and semantic manipulation with conditional gans[J/OL]. CoRR. 2017, abs/1711.11585. http://arxiv.org/abs/-1711.11585.

[12] Mao X, Li Q, Xie H, et al. Least squares generative adversarial networks[C]. 2017 IEEE International Conference on Computer Vision (ICCV). [S.l.], Oct 2017: 2813-2821. doi: 10.1109/ICCV.2017.304.

[13] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. science. 2006, 313 (5786): 504–507.

[14] Cortes C, Vapnik V. Support-vector networks[J]. Machine learning. 1995, 20 (3): 273–297.