Hindawi

Research Article

High-Precision Target Detection of Remote Sensing Image Based on Feature Enhancement with 6G Technology

Haiyan Chen (ID)

Beijing Polytechnic, Beijing 100176, China

Correspondence should be addressed to Haiyan Chen; gracechen128@sina.com

We provide remote sensing image enhancement technology based on 6G nursing urine technology to solve the main problems and current challenges of target detection in remote sensing imaging. High-resolution images: to get a good image, first look at the remote control image, and objective findings on the remote control image play an important role in this process. Military and civilian use: at present, many advanced object detection algorithms have achieved success in natural imaging, but their development is limited by two factors: large variation in object size and low detection accuracy, the scale, rotation direction, and distribution density of remote sensing images. On the other hand, the first remote sensing image is usually a high-resolution, large-scale image, which requires more time than blank image detection. The model is fine-tuned on the ICDAR2015 training package for 60,000 repetitions. The same model experimentally reduced the image to 1280×768 and displayed the results in one dimension. Therefore, in order to solve the problem of low detection due to changes in azimuth, rotation direction, and velocity distribution, it is necessary to further study methods for detecting large-capacity remote sensing objects using abundant high-precision remote sensing technologies, sensing imaging service on remote sensing maps.

1. Introduction

From the launch of the world's first satellite in the 1950s to the launch of China's first satellite Dongfanghong-1 in the 1970s, remote sensing image research has been steadily advancing. Compared to synthetic diaphragm radar (SAR), it can take pictures in different weather conditions. Remote sensing images are more accurate and less susceptible to noise, and have more features. Therefore, object detection in remote sensing images is usually prone to noise. Real-time detection of long-range maps involves determining whether the aforementioned maps and satellites contain one or more objects of interest, and marking the location of each object on the map. Target detection based on high-precision remote sensing images has been widely used in military and civilian fields. For public use, high-resolution imaging can be used for research, urban planning, agriculture, and commercial use. In the military, remote electronics are critical to economic development, which is also important to national security. This is an important technology used in modern military reconnaissance and early warning. There are four challenges in target detection using remote sensing images: first, the appearance and size of remote sensing targets are different, the basic shape of targets is often too long or too large, and different kinds of targets have different scales, aspect ratios, and shapes; second, because the remote sensing image is obtained from the space platform, the target has multiple directions, which makes the detection task more difficult; third, there are complex background clutter around the target, such as wave interference and cloud occlusion; fourth, we need to consider the changes of target visual shape caused by illumination and shadow, as well as the distribution of complex scenes caused by different targets.

2. Literature Review

Li Zhen and others said that in recent years, the remote sensing optical imaging technology has developed rapidly, and the spatial accuracy of long-distance optical imaging has been continuously improved. Compared with other types of

remote sensing images such as synthetic aperture radar (SAR), high-resolution optical remote sensing images have similar characteristics [1]. The research and application of long-range optical imaging in the civilian and military fields have received extensive attention. You and others said that many countries such as the USA, Russia, and Israel are developing their own high-resolution remote sensing satellite technology. For example, Israel EROS-B provides remote sensing images with a resolution of 0.7 m, the image resolution of QuickBird satellite of the USA reaches 0.6 m, and KH-12 satellite launched by the National Investigation Bureau of the USA adopts automatic optical imaging, which can reach a resolution of 0.1 M [2]. Patil and Ram Singla and other commercial companies provide high-resolution remote sensing imaging services. For example, satellite remote sensing maps from Google Earth can have an accuracy of 0.27 m [3]. Wang et al. said that China has a wide variety of remote sensing satellites and a complete system. In recent years, my country has increased the use of high-precision remote sensing satellites, reaching the highest level in the world [4]. The panchromatic image resolution of "Gaogao-1" launched in April 2013 can reach 2 m and can also provide 8 m spatial resolution multispectral images with combined width better than 60 km or 16 m spatial resolution multispectral images with combined width better than 800 km. Pandit and Bhiwani said that in August 2014, the resolution of Gaogao-2 satellite improved to LMO. Jilin-1, which was put into use in October 2015, is China's first commercial remote sensing satellite, creating a precedent for the application of commercial satellites in China [5]. It is also said that the first four stars are optical stars, two intelligent photography stars and intelligent light stars. It supports images with a resolution of 0.72 m and high-resolution video with a resolution of 1.13 m. The Jilin-1 satellite is currently under construction. In January 2018, the number of satellites in orbit reached ten [6]. You et al. launched in 2016 the Gaojin-1 series which is my country's first 0.5-meter high-precision remote sensing satellite. They can display high-resolution images over distances over 60 km and have better panchromatic accuracy of 0.5 m or 2 m multispectral [7]. Li et al. stated that the highest-precision industrial remote sensing satellite developed by China Aerospace Science and Technology Corporation in December 2018 has a density of 0.3 meters, and its performance is the best commercial remote control satellite in the world today [8]. Zhou and others said that at the same time, with the increase of the number of orbit remote sensing satellites, more high-frequency revisit observations in the same region can be realized, which makes the time resolution of the current optical remote sensing image rapidly improve [9]. Yu and others said that at present, the time resolution of remote sensing images has achieved a revisit cycle of several hours, and the phase II construction goal of "Jilin No. 1" satellite constellation will have the ability to revisit any place in the world in 10 minutes [10]. High-resolution optical displays have the advantages of high resolution, high sensitivity, low noise, low distortion, and no electromagnetic interference. This is due to the high reliability of applications in many industries. Object detection is one of the most important

studies and applications of information in various remote sensing systems, including the ability to remotely control optical images. The goal is to study objects of interest from a distance, measure their functions, and distinguish them from their properties. The higher the accuracy, the larger the field of view brought by the image, the better the data, the easier it is to understand on the ground, and the more demanding the detection procedure, such as locating many objects, small objects in the background. Remote sensing-based target detection technology is widely used in service, environmental protection, rescue, disaster relief, traffic scheduling, urban management, agriculture, forestry, national defense, military, and other fields. For example, in urban planning and management, we can help build roads, bridges, and ports. By identifying urban vehicles and determining the distribution of products and land, it is easier and more convenient to detect public ships such as rivers and seas, ports, water control, fishing management, and emergency zone detection; in the defense and military fields, it has been used for high science and technology research missions to identify targets such as major enemy military bases, weapons, combat vehicles, operating airfields, military aircraft, and nuclear facilities. Efficiency is important. The establishment of military service monitoring is shown in Figure 1.

3. Method

The perceptron proposed by Frank Rosenblatt in 1957 is often regarded as the ancestor of neural network. In the 1980s, a neurocognitive machine with visual cognitive function (neocognitron) appeared. This multilayer neural network directly inspired the later convolutional neural network. In 1997, LeNet-5 proposed by Yan Lecun first appeared a multilayer cascade convolutional structure. It is an early convolutional neural network with real practical value, which can effectively recognize handwritten digits. AlexNet proposed in 2012 has a similar structure to LeNet, but it is deeper and wider. It is a deep convolutional neural network in the modern sense. It has successfully applied new technologies such as ReLU activation function, dropout, lRNC local response normalization layer, and GPU acceleration in the convolutional neural network, and won the champion of ImageNet large-scale visual recognition challenge (ILSVRC) with great advantages. The error rate of top 5 is 16.4070, which is much higher than that of top 5 in the second place by 26.2% [11]. The success of AlexNet inspired the subsequent technological innovation of convolutional neural network and greatly stimulated the research and upsurge of convolutional neural network [12]. Subsequently, CNN entered the "evolution" stage of rapid development. A large number of theories and applications emerge every year. With the continuous improvement of performance, the network structure is becoming more and more complex. On the whole, there are two trends in the structural development of convolutional neural networks in recent years: deepening and widening. The convolutional neural network with "widening" is represented by "network in network" Nin (network in network, 2013), Google perception V1, Google perception V2,
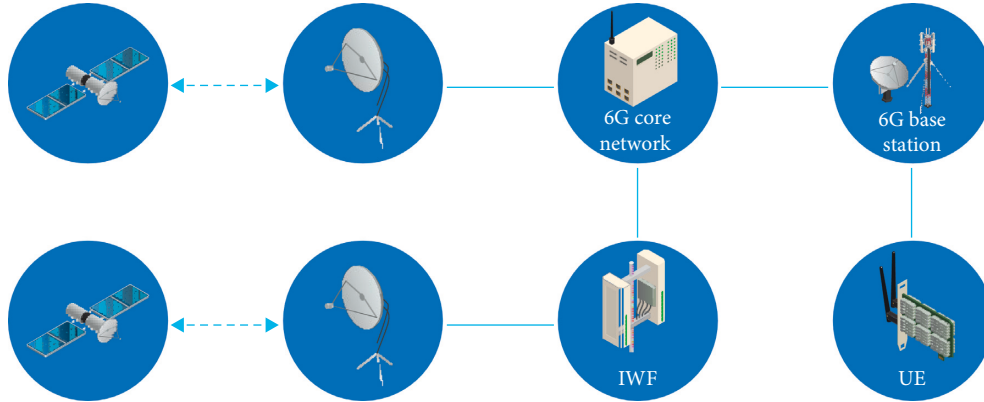
Figure 1: High-precision target detection of remote sensing image enhanced by 6G technical features.

Google perception V3, Google perception V4, and dense convolutional network DenseNet. The typical representatives of "deepening" networks include VGGNet, MARANet, residual network ResNet, ResNet V2, and ResNeXt. ResNet has reached 152 layers, and the error rate of top 5 is 3.57 070, winning the championship of ILCVRC 2015. In addition, there is a network that combines the advantages of two directions, such as inception ResNet V1/V2, which integrates inception module and ResNet residual learning module; and lightweight convolutional neural networks born to improve training and operation efficiency, such as SqueezeNet, MobileNet, ShuffleNet, and Xception. On the premise of ensuring certain performance, through the redesigned network structure rather than relying on the previous model compression means, it significantly reduces the storage space and speeds up the training and reasoning speed. For example, SqueezeNet has the same accuracy as AlexNet. However, the parameter is only 1/50 of the latter. The model parameters only need a small memory space and are more suitable for mobile phone and other energy constrained platforms. Several typical CNNs are introduced and analyzed below [13]. Google inception V1 was proposed by Google in 2014. It is mainly to greatly improve the two problems caused by the sharp increase of parameters after the network becomes deeper and wider: the sharp increase of calculation and easy overfitting. Inception module is creatively proposed to improve the utilization rate of parameters; in addition, the global draw pool layer is used to replace the last full connection layer, which greatly reduces the number of parameters. Achieve a higher accuracy of 1/12 of AlexNet, with a top 5 errors of 6.67070, winning ILC V RC 2014. The concept as n-modules pulls the idea of multiscale Gabor filters to make photos. Multiple measurements of core connections connected in series are equivalent to a diagram or map for each location. The structural model is shown in Figure 2.

ResNet's core component module is a one-dimensional component that takes information from the highway network design philosophy and converts the network output into a nonlinear superposition of the input. The goal of network training shifts from achieving results to different learning outcomes and strategies, i.e., balance. Assuming that the network input is x and the advantage is $h(x)$, after $H(x) = f(x) + x$ plus its own report, the target of network training will

change from $H(x)$. $f(x)$, and the remaining $H(x) - x$ reduces the learning difficulty [14]. Among them, the three-tier training class draws instructions from the 1×1 rotation on Nin and Google Minds, which can reduce negative numbers. "Jump Connections" send inputs directly to outputs, helping to reduce data loss. The residual element is simple and practical, does not bring additional parameters, and can be easily applied to other convolutional neural networks. In addition, ResNet V2 replaces the nonlinear activation function ReLU in the residual unit with an identity map and adds batch regularization and batch normalization n to each layer, which simplifies the training and enhances the generalization ability. The structural diagram of the residual learning unit of layer 2 and layer 3 is shown in Figure 3.

Dense connected revolutionary network (DenseNet) is composed of multiple dense blocks and translation layers alternately. Dense connection block is its most significant innovative structure. The internal "skip connections" of each dense connection block are no longer limited to the adjacent convolutional layer. The input to the current layer comes from all previous layers, and there is a direct connection between any of the layers. That is why it is called a "tight junction." For L-layer tomographic neural network, traditional network has l connections, while DenseNet has $L * (L+1)/2$ connections. The output of each layer combines the properties of all previous layers. For example, if the output of the current batch 1 is assumed to be XL, the output of all previous packs $[X0, X1, \ldots, X1-1]$ is given by

$$X_l = H_L\left([X_0, X_1, \ldots, X_{l-1}]\right). \tag{1}$$

In regional applications, the Fast R-CNN initiative in 2015 was the first to define end-to-end $(X_l = H_l([X_0, X_1, \ldots, X_{l-1}]))$ training and discovery. Chinese equipment has achieved remarkable results in terms of speed, precision, and usability. In that year, it achieved the best results in the target detection tasks of PascalVOC 2012, ILSVRC 2015, and MS COCO 2015, which attracted extensive attention, displaying its detection frame. These methods are constantly being improved and put into use due to the high-precision detection methods available in the region. For example, R-CNN mask can perform segmentation and object detection at the same time and represents potential and available office space, as shown in Figure 4.
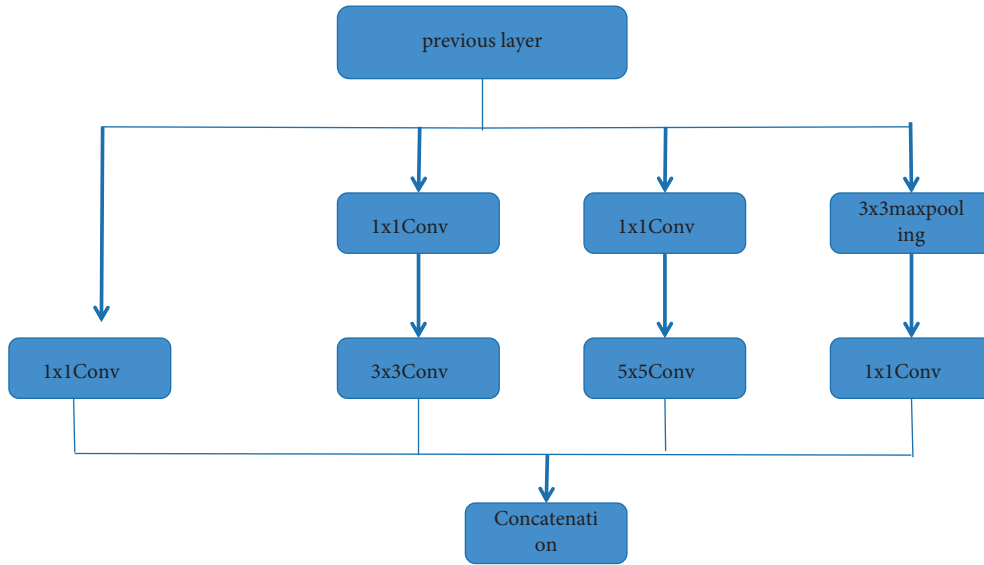
```
                              previous layer


         ┌─────────────────┬──────────────┬──────────────┐
         │                 │              │              │
                      1x1Conv         1x1Conv      3x3maxpool
                                                      ing


      1x1Conv          3x3Conv         5x5Conv       1x1Conv


         └───────────────┬──────────────┘              │
                         │                              │
                    Concatenati
                        on
```

FIGURE 2: Structure of inception module.

```
                                                    ┌──────────┐
                                                    │          │
      ┌──────────┐                         1x1,64   │
      │          │                                  │
   3x3,64        │                                  │
                 │                         3x3,64   │
                 │                                  │
   3x3,64        │                                  │
                 │                         1x1,256  │
                 │                                  │
       +  ◀──────┘                            +  ◀──┘
       │                                      │
       ▼                                      ▼
```

FIGURE 3: Residual block structure.

```
                                              ┌──────┐  ┌──────┐    classfier(soft
                                              │      │  │      │        max)
                          RPM  ─▶  regions    │ RolP │  │ PCLa
                           ▲                  │ oolin│  │ yers
                           │                  │  g   │  │
   Image ─▶ CNN ─▶ Fealture│map ─────────────▶│      │  │     ─▶ Boundary
                                              │      │  │         box(regressor)
                                              └──────┘  └──────┘
```
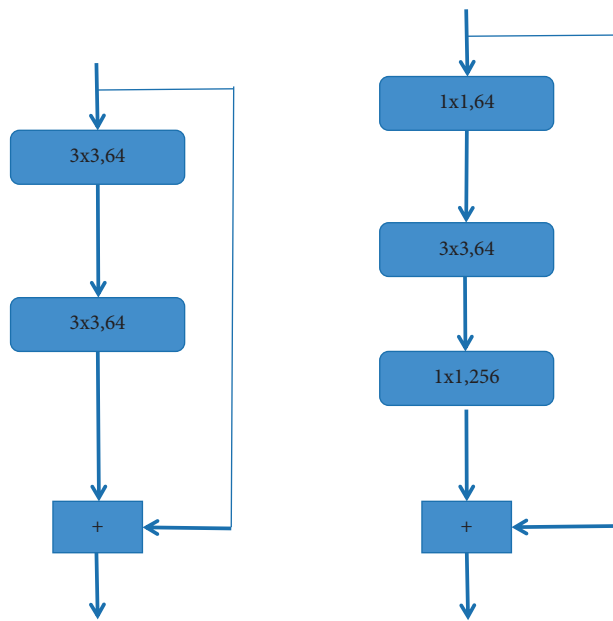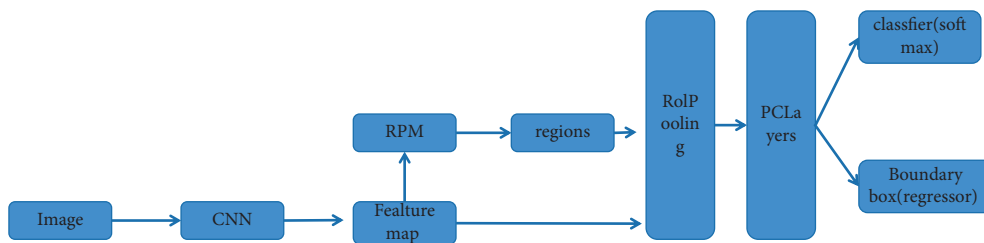
FIGURE 4: Structure diagram of the faster R-CNN model.

RPN is a small full capacity network. Its main structure is a 3×3 convolutional layer and two parallel 1×1 convolutions. Both 1×1 convolutions complete the predivision and position regression parts. A special map of the old image is obtained from the neural network connection and sent to the RPN to remove the candidate boxes. The main process is as follows: the rectangular anchor boxes with various scales are used to scan the feature map step by step, and K anchor frames are generated at each position. In the original contribution of Fast R-CNN, the author sets three scales $\{12822562, 5122\}$, three aspect ratios $\{1:1, 1:2, 2:1\}$, and $K = 9$ anchor frames at each position. For example, the 40×60 feature map will produce about 20000 ($40 * 60 * 9$) candidate frames. Because most anchor frames overlap each other and have high redundancy, in order to improve efficiency, screening needs to be done. First, the foreground/background classification is done through softmax. Here, the specific categories are not concerned, and the background anchor frames that cannot exist the target of interest are discarded; then, the highly coincident boxes are removed by nonmaximum suppression (NMS) method, and the top n with the highest score is obtained. Finally, only parts of the most likely target areas are reserved. The specific number of reserved areas is determined by the super-parameters set during the construction of the model, which is generally set to hundreds to 2000. At the same time, another network of RPN performs a bounding box regression to adjust the position and size of the anchor box to form a more accurate recommended area box [15]. In order to train RPN, two kinds of labels are designed for each anchor point: foreground (which may contain objects of interest) and background (which does not contain objects of interest). Two types of anchor points are marked as positive examples: the anchor point with the largest overlap with the dimension box and the anchor point with an overlap IOU of more than 0.7 with any dimension box. If the overlap with the dimension box is less than 0.3, it will be marked as a negative example. Nonpositive and non-negative anchors are directly discarded and do not participate in training. The multitask loss function of RPN is shown in

$$L(\{p_i\}, \{t_i\}) = L_{cls} + L_{box}$$
$$= \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i) + \lambda \frac{1}{N_{reg}} \sum_i p_i L_{reg}(t_i, t_i). \tag{2}$$

Among them, the first part is the loss of the target distribution class Lcls, the second part is the regression loss of the Lbox position box, and the third part is the unadjusted weight. I is the measured size of the anchor, P is the approximate value of the anchor of Ship II relative to the anchor, and P is the approximate value. This anchor is a bad example $p * i = 0$, and a good example is $p * i = 1$. Lcls (pi, $p * i$) is the binary logarithmic drop shown in

$$L_{cls}(p_i, p_i) = -\log[p_i p_i + (1 - p_i)(1 - p_i)]. \tag{3}$$

T is a four-dimensional parameterized coordinate vector of the approximate position location, $t = \{$tx.u. Tws. Th$\}$, and t^* is the parameterized coordinate vector corresponding to the real dimension box, as shown in

$$t_x = \frac{(x - x_a)}{w_a, t_y}$$
$$= \frac{(y - y_a)}{h_a, t_w}$$
$$= \log\left(\frac{w}{w_a}\right), \tag{4}$$
$$t_h = \log\frac{h}{h_a},$$

$$t_x = \frac{(x - x_a)}{w_a, t_y}$$
$$= \frac{(y - y_a)}{h_a, t_w}$$
$$= \log\left(\frac{w}{w_a}\right), \tag{5}$$
$$t_h = \log\left(\frac{h}{h_a}\right),$$

$$t_x^* = \frac{(x^* - x_a)}{w_a, t_y^*},$$
$$= \frac{(y^* - y_a)}{h_a, t_w^*},$$
$$= \log\left(\frac{w^*}{w_a}\right), \tag{6}$$
$$t_h^* = \log\left(\frac{h^*}{h_a}\right).$$

The regression loss is $L_{reg}(t_i - t_i)$, and R is the smooth L1 function, which is defined as follows. The function is differentiable at zero, grows linearly in the area with large error, has low sensitivity to outliers, and can better avoid gradient explosion, as shown in

$$\text{Smooth}_{L1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1, \\ |x| - 0.5, & \text{otherwise.} \end{cases} \tag{7}$$

After the candidate regions are obtained from the RPN, the candidate region classification and location border regression are carried out. The structure of this part inherits the Fast R-CNN model, mainly including ROI pooling layer, fully connected layers (FC), and two sub-fully connected layers used for classification and border regression, respectively. Candidate regions of different sizes generate fixed size (7×7) feature maps through the pooling operation of regions of interest and then use the full connection layer to map the feature map to the category space, that is, generate feature vectors of specific dimensions. The classification layer combines softmax to distinguish specific categories. If the class of the object to be detected is n, the process output vector is $n + 1$, including the background. At the same time, the regression process combines L1 drop to complete the

boundary regression function, obtain the correct position of the target, and generate a value vector with a length of 4 $(n + 1)$. Its top 5 bug is that only 5.25% of ImageNet packets achieve low cost and high distribution, suitable for remote touch processing applications; these models are traditional and easy to maintain. Its structure is shown in the table, including several categories of parts. A convolution with stride 2 is used to determine the function of the mixing process [16]. The global average pooling (GAP) realizes the function of partial full connection layer, which can realize the input of any image size. More importantly, it greatly reduces the amount of parameters brought by the full connection layer and reduces the problem of overfitting. The primary residual network supports 1000 distribution targets. In effect, the number of output nodes is converted to the number of clusters they need. This test will be done on 10 target files, so the number of devices published on the network and Softmax distribution is adjusted to 11, including a background group. There are many options for feature removal networks, such as AlexNet, GoogLeNet, VGGNet, and ResNet. The backbone of these networks is a combination of multiple convolutional layers and multiple pooling layers. Because the pool layer reduces the image size and resolution, the higher the level, the lower the image resolution; at the same time, the network structure of CNN determines that the higher the level, the larger the receptive field of the feature map, and is more inclined to detect large targets. This characteristic of convolutional network leads to the disadvantage of small target detection. Taking the classic VGG16 as an example, only the first five convolution blocks of VGG16 are used for Fast R-CNN shared feature extraction, including four maximum pooling layers. Each pooling operation is $2 * 2$ downsampling, and the image size is reduced by 50070. Finally, the size of the feature image conv53 is only 1/16 of the original input image. When the target scale in the original image is lower than 16 pixels, there is almost no corresponding positioning and classification information on the feature image, which is naturally difficult to detect. This deficiency can also be seen from the perspective of receptive field. The calculation of receptive field of multilayer CNN convolutional layer is shown in

$$\mathrm{RF}_k = \mathrm{RF}_{k-1} + (f_{k-1}) * \prod_{i=1}^{k-1} s_i. \qquad (8)$$

There are three ways to gather information in a pyramid system: bottom-up, top-down, and horizontal. The following is further processing of the recurrent neural network. The boxes on the left are the final output features of each convective neural network connection. The map feature accuracy of each rotated block is reduced by 1/2 from top to bottom, i.e., double the top map feature, and then special labels are applied during the horizontal join process. Use 1×1 turns on the assembly size, melt the dots one by one, then use a $3 * 3$ damage test to assemble each combination to eliminate the effect of the different names in the example, and create similar special statements. The first feature of this pack is the size of what is referred to here as the super-feature. The features of each layer of the feature pyramid are accumulated by the features of all layers on it, including features with different resolutions and semantic intensities, and the features of each layer from top to bottom are constantly enriched [17]. FPN has good generalization ability. Its application to some image processing methods based on deep neural network can bring significant performance improvement and even speed improvement, including target detection and instance segmentation. At the same time, because this method only adds several additional cross-layer connections and simple feature fusion on the basis of the original network, the additional time consumption and calculation amount are very small. Select the output feature map of the four groups of residual blocks after the basic network ResNet for feature fusion to enhance the detection ability of small targets. For the design of each group of sections, the results of Convl, Conv2x, Conv3x, Conv4x, and Conv5x are denoted as {C1, C2, C3, C4, and C5}. Using nta features {P2, P3, P4, P5}, to size C1, such that memory size and low semantic value are bad for the distribution target, and no features are used when creating super. The special operation is shown in Figure 5.

Here, a convolution kernel of 3×3 size is taken as an example to illustrate its operation details. The conventional block convolution operation mainly includes two steps: determining the regular region R to be sampled on the upper input characteristic graph, which is the receptive field of the block convolution kernel; the value of each sampling point of R is weighted and summed with the weight W of the corresponding position of the conventional convolution kernel. If there are multiple channels, the results are added as shown in

$$R = \{(-1, -1), (-1, 0), (-1, 1), \ldots, (0, 1), (1, 1)\}, \qquad (9)$$

or a point P0 in the output feature, and the convolution operation is defined as shown in

$$y(p_0) = \sum_{p_n \in R} W(p_n) \bullet X(p_0 + p_n). \qquad (10)$$

Deformed rotation structures have their own offset direction and are not limited to the inside of the block. The deformed rotated samples have larger scale changes and no rotation changes than regular rotations. Equation (10) follows (11) immediately after applying the bending rotation.

$$y(p_0) = \sum_{p_n \in R} W(p_n) \bullet X(p_0 + p_n + \Delta p_n). \qquad (11)$$

Since the offset Δp_n is usually not an integer, and the pixel position coordinates of the actual sampling points must be integers, direct simple and rough rounding will lead to obvious errors. Here, the only way to obtain STN is to obtain the pixel value of each point by bilinear interaction. The bilinear difference formula is shown in

$$X(p) = \sum_q G(q, p) \bullet X(q). \qquad (12)$$

The model needs to get the rotating rectangle in the region generation network stage, but how to reasonably express a rotating rectangle is a very difficult problem. Most
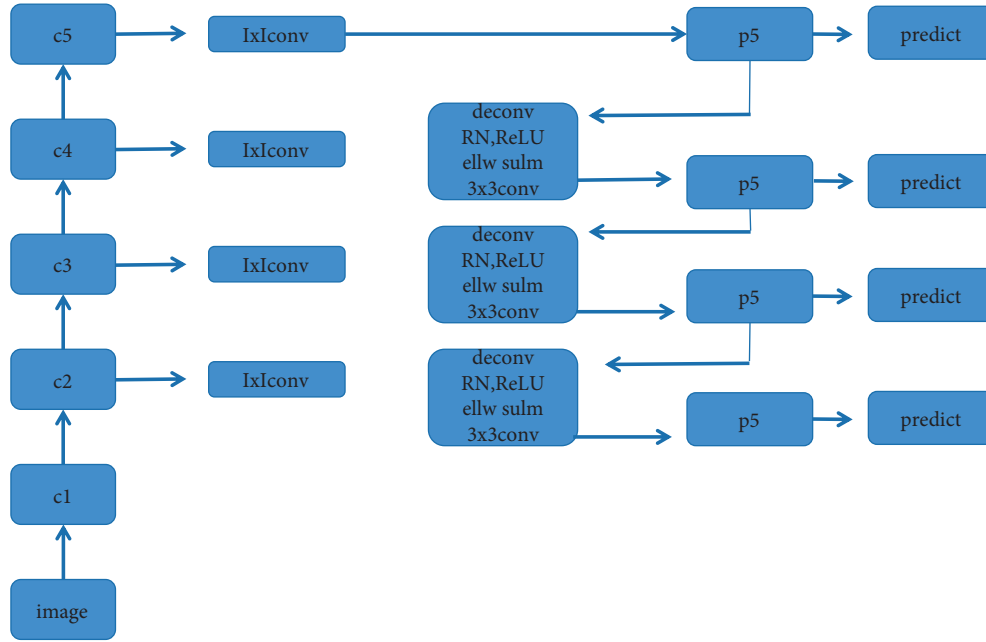
FIGURE 5: Combination mode of ResNet and FPN.

of the previous methods directly regress the angle of the rotating rectangular box. If the scale angle is too large, the efficiency of the system will decrease. To determine whether changes in the encoding time used are related to transition angle effects, this study provides control experiments: one model regresses directly to the optimal longitudinal angle and the other regresses to the nonadjustable period. Coding and evaluation of shaft quality in power grid area: The network only classifies objects into two categories (positive and negative samples). The model uses average accuracy (AP) as its evaluation standard. The comparison results are given in Table 1. It can be seen that the modification cycle has been achieved more effectively. It can be seen that after using the time-varying transformer, the region produced by the box is more accurate compared with the concept of direct angle, as given in Table 1.

To determine the validity of long-term comparisons of independence, a similar experiment was conducted in this study. The comparative experimental results are given in Table 2 [18]. Quick R-CNN means that only one R-CNN is used here. O cascade R-CNN means that R-CNN is used twice. The first model computes the overlap of box rotations through the interaction of two stages. In the second model, the overlap is calculated by the length of the independent intersections in the first R-CNN, and the overlap of the rectangles is calculated by the intersections in the second R-CNN. The total close price is set to 0.5. It can be seen that cascade R-CNN with length-independent interaction has better performance than cascade R-CNN. It can be seen that the length-independent interaction ratio improves the quality and recall of the output box. Length-independent interaction has positive samples for targets of any angle and scale, so the detector can better deal with objects of any length and scale, as given in Table 2.

TABLE 1: Experimental results on adjustable periodic coding (APE) during region generation on DOTA's verification set.

Methods	w/o APE	W/APE
AP	70.06	72.21

TABLE 2: Experiments on length-independent interaction ratio and interaction ratio on DOTA verification set.

Methods	Faster R-CNN	Cascade R-CNN	Faster R-CNN + LIIoU
mAP	71.41	72.75	73.87

In order to verify the performance of the length-independent interaction ratio, corresponding experiments are designed in this study. The results of the comparative experiments are given in Table 2. Fast R-CNN means that only one R-CNN is used here. O cascade R-CNN means that R-CNN is used twice. The first model calculates the overlap rate between rotated rectangles by the interaction rate of the two stages. In the second model, the overlap ratio is calculated by the length-independent interaction ratio in the first R-CNN, and the overlap ratio of the rotated rectangle is calculated by the interaction ratio in the second R-CNN. All closed values are set to 0.5. It can be seen that cascade R-CNN with length-independent interaction has better performance [19]. It can be seen that the length-independent interaction ratio improves the quality and recall of the output box. The length-independent interaction ratio has positive samples [6] for targets of any angle and scale, so the detector can better deal with objects of any length and scale [20]. The comparison with other schemes is given in Table 3.

This study compares the proposed method with other schemes in DOTA data set and cvpr2019-doai competition

data set, respectively. The results are given in Tables 3 and 4. In DOTA experiment, the model only uses a single scale in both training and testing models [22]. When the proposed model only uses the training set for training, it has well exceeded the other methods. When additional verification sets are added for training. The proposed model achieves better performance. For cvpr2019-doai competition, in order to obtain the best performance, the model additionally adopts image rotation, multiscale training and testing, and model fusion strategy. The model uses ResNeXt-101 (32×4) as the backbone of the network and integrates three models. Finally, the training set and verification set are combined to train the proposed model. It can be seen that the proposed scheme ranks first in the rotating frame target detection task, and the comparison with other schemes is given in Table 4.

Optical remote sensing target detection under rotating frame annotation is a very challenging task. By making full use of the periodicity of the angle, a method called adjustable periodic coding is proposed. This scheme can well regress the rotating rectangular box in remote sensing image. A vector with adjustable period can learn the periodicity of angle, which cannot be solved by a single dimension vector. The proposed method can be used in single-stage method or two-stage method. Other detectors can also directly call the proposed adjustable period coding module. In addition, this study also proposes a length-independent interaction ratio method, which improves the regression quality of R-CNN stage by setting more target boxes matching with long samples as positive samples. The proposed scheme won the first place in DOAI2019 rotating frame competition. However, the regression-based target detection method needs to cover the receptive field of the whole object. The length-independent interaction ratio proposed in this chapter can increase the recall rate of the long target to a certain extent. However, the regression-based target detection needs to cover the receptive field of the whole target, and the length-independent interaction ratio still cannot solve the extreme situation. The next chapter will propose another solution to this problem.

4. Experiment and Analysis

In this study, the training set and verification set of MLT are combined for the training model, and a total of 180000 iterations are trained. First, it is verified whether the proposed text center edge probability and text center direction can well separate pixels belonging to different text instances and combine pixels belonging to the same text instance. In the test, the model uses a single size test, and the long edge of the image is reduced to 1800. This study then compares the proposed method with other methods and confirms the proposed method in a broad context [23]. For multiscale experiments, the model downscales the long side of the image to {1000, 1800, 2600} pixels and mixes the multiscale results without maximum compression. The probability function of the text center edge is given in Table 5.

In order to prove that the center of text center edge probability can well separate instances belonging to different text lines, its gradient direction can be used to combine

instances belonging to the same text line. First, this study trains a semantic segmentation model, which only includes the text score module and another model including the text center edge probability module [24]. Table 5 provides the comparison of the results. It can be seen that the semantic segmentation model does not have enough information to separate and combine text lines, so its performance is very poor. However, after adding the probability of text center edge, the model can get enough center edge information for postprocessing to get better results. The role of text center direction is given in Table 6.

Although the model has obtained enough information from the text center edge probability to separate and combine text lines, adding additional text center direction training makes the model more robust [20]. In order to prove this, this study adds text center direction training to the model. As can be seen from Table 6, only adding the text center direction in the training can improve the model by 1%. It can be seen that there is a strong correlation between the direction of the text center and the probability of the edge of the text center. They learn different expressions of the same feature. Moreover, the text center direction can close the pixel features belonging to the same text instance and pull away the pixel features belonging to different text instances. However, using the text center direction in the test phase cannot improve the performance of the model, because the text center direction information has been well included in the text center edge probability. The comparison between TextMountain and other methods is given in Table 7.

The proposed method is compared with other advanced methods as given in Table 7. Results are based on ResNet-50. It can be seen that the proposed method achieves very excellent results [21]. In the single-scale test, TextMountain obtained 74.72% F-measure index, and in the setting of multiscale test, TextMountain reached 76.85% F-measure, which has an advantage of about 4% compared with the most competitive method. In this study, the performance of the proposed method is verified on ICDAR2015 data set, and 60000 iterations of the model are fine-tuned on ICDAR2015 training set. The same model zooms the image to 1280×768 in the test and shows the results in a single size. Table 8 compares the results of the proposed scheme with those of other methods. Compared with other segmentation schemes, the proposed method achieves better performance (precision: 88.51%, recall: 84.16%, F-measure: 86.28%). And benefiting from the parallel combination algorithm, Text-Mountain obtained the efficiency index of 10.5 FPS, which indicates that it also has very excellent efficiency [25]. The results on the ICDAR2015 dataset are given in Table 8.

The performance of the proposed method is tested on the ICDAR2015 data set, and the model is fine-tuned on the ICDAR2015 training package for 60,000 iterations. Similar to the model upscales the image to 1280×768 in experiments and displays the results to some extent. Table 8 compares the results of the proposed scheme with those of other methods. Compared with other segmentation schemes the proposed method has better performance (precision: 88.51%, recall: 84.16%, f-measurement: 86.28%). And benefiting from the parallel combination algorithm,

TABLE 3: The results on the DOTA test set indicating that the verification set is also used for training; otherwise, only the training set is used for training.

Method	Ours	Ours*	FR-O [6]	RoITransformer*	RRPN [23]	R2CNN [21]
Plane	89.57	89.96	79.09	88.64	80.95	88.53
BD	76.76	83.62	69.12	78.52	65.75	71.2
Bridge	51.29	53.43	17.17	43.44	35.34	31.66
GTF	71.64	76.02	63.49	75.93	67.44	59.4
SV	73.11	74.02	34.2	68.82	50.92	51.85
LV	77.16	77.15	37.16	73.68	50.91	56.20

TABLE 4: Task 1-DOAI2019 rotating frame ranking.

TeamName	USTC-NELSLIP	Pca-lab	Czh	AICyber	CSUIQQ	Peijin
Plane	89.3	88.2	89.0	88.5	87.8	80.9
BD	85.3	86.4	83.2	85.4	83.6	83.6
Bridge	57.4	59.5	54.5	56.8	56.7	55.2
GTF	80.9	80.0	73.8	74.5	74.4	70.7
SV	73.9	68.2	72.6	63.9	63.3	59.9
LV	81.3	75.6	80.4	72.7	71.0	76.4

TABLE 5: Comparison of model performance with or without text center edge probability.

Method	P	R	F
w/o TCBP	57.43	52.53	54.86
w/TCBP	80.92	67.70	73.71

TABLE 6: Function of text center direction.

Method	P	R	F
w/oTCD	80.92	67.69	73.72
w/TCD	82.79	67.92	74.61
w/TCD (train)	82.82	68.06	74.71

TABLE 7: Comparison between TextMountain and other methods.

Method	P	R	F
Linkage-ER-Flow	44.50	25.60	32.49
TH-DL	67.76	34.78	45.98
SARI_ FDU_ RRPN vl	71.17	55.51	62.37
Sensetime OCR	56.93	69.42	62.56
SCUT_ DLVClabl	80.28	54.55	64.97
Lyu et al.	83.81	55.6	66.8
TextMountain	88.82	68.26	74.72

TextMountain has obtained the efficiency index of 10.5 FPS, which indicates that it also has very excellent efficiency. On the SCUT CTW 1500 data set, this study examines the performance of the proposed method in detecting curved text lines, fine-tuning the pattern on the SCUT CTW 1500 training set for 60,000 iterations, and reducing the image length to 800 [25]. To be fair, the model is tested on a single scale. The test results are given in Table 9. Compared with other schemes, the model achieves good results (83.2% *f*-measurement), which shows that the text mountain curve

TABLE 8: Results on the ICDAR2015 dataset.

Methods	P	R	F	FPS
Zhang et al.	71	43	54	—
SegLink	73.2	76.9	75.0	—
EAST	83.57	73.48	78.20	13.2
EAST	83.26	78.33	80.73	—
He et al.	82	80	81	—
PixelILink	85.6	82.0	83.7	3.1
Lyu et al.	89.6	79.8	84.3	1

TABLE 9: SCUT ctw1500 dataset.

Method	P	R	F
SegLink*	42.4	40.0	40.8
SWT*	20.8	9.0	12.5
CTPN*	60.4	53.7	56.8
EAST*	78.7	49.1	60.4
DMPNet*	69.8	56.2	62.2
CTD + TLOC	77.4	69.8	73.4
TextSnake	67.9	85.4	75.7

[25] handles text lines well. The results of the SCUT ctw1500 packet are given in Table 9.

In the coarse detection stage, the image is divided into blocks to eliminate the color rendering index, and then the image is segmented according to a deep-learned segmentation model. When dividing into blocks, the number of blocks that can be cut vertically and horizontally in the figure is m and N, respectively, and the total number of blocks is mxn. Equation (13) is shown as follows:

$$m = \left\lceil \frac{H - O_v}{BH - O_v} \right\rceil,$$

$$n = \left\lceil \frac{W - O_v}{BW - O_v} \right\rceil. \tag{13}$$

Here, h and W represent the height and width of the original image, and O represents the overlap of adjacent overlapping parts, which can be adjusted according to the resolution of the image. In this format, the overlap is set to 100. BH and BW represent the height and width of the cutting block, respectively. After combining these properties, the object detector uses its own structure. Considering a separate rotation layer of the starting concept, this larger separable convolutional layer replaces the convolution core with two-layer convolution of LXK and KXL. The convolution operation reduces the amount of calculation on the premise that the calculation result remains unchanged. As shown in Figure 6, the concept module designed in the detector is added with deformable convolution operation based on the large separable convolutional layer on each branch to expand the receptive field and add semantic information. Finally, a new characteristic graph is obtained by adding elements on the two branch channels. Among them, the size value k of convolution kernel in separable convolutional layer is set to 15, and the inception module is shown in Figure 6.

The deformable convolutional layer used in the inception module was first introduced in Dai, which can directly obtain additional offset information through target task learning without adding redundant supervision information [26]. Unlike standard convolution, which must use regular sampling grid, deformable convolution allows arbitrary deformation of sampling grid, so as to enhance the spatial sampling and positioning ability of the network. In addition, the deformable convolutional layer is also trained end-to-end through the conventional back-propagation algorithm and can easily replace the standard convolution unit in any CNN structure. The first step of convolution operation using standard convolution unit is to sample the characteristic image in a fixed convolution window, and the second step is to multiply and add the values of each position in the window and the corresponding values in the convolution kernel. For example, when a certain point P0 in the characteristic graph I to be convoluted is used for standard convolution operation with a convolution kernel r of 3×3, it is shown in

$$S(p_0) = \sum_{p_n \in R} K(p_n) \bullet I(p_0 + p_n), \tag{14}$$

where P_n is a position in the convolution kernel R, I is the input characteristic diagram, and the value of R is as shown in

$$R = \{(-1,-1),(-1,0),\dots,(0,1),(1,1)\}. \tag{15}$$

In the deformable convolution operation, an offset $\{\Delta p_n = (\Delta p_n^x, \Delta p_n^y) \mid n = 1, \dots, (N\}$ is defined, and N is the number of elements in the receptive field area. The corresponding formula (15) can be rewritten as shown in the following:

$$S(p_0) = \sum_{p_n \in R} K(p_n) \bullet I(p_0 + p_n + \Delta p_n). \tag{16}$$

As a variant of conventional ROI pooling, PSROI pooling mainly encodes useful spatial information for
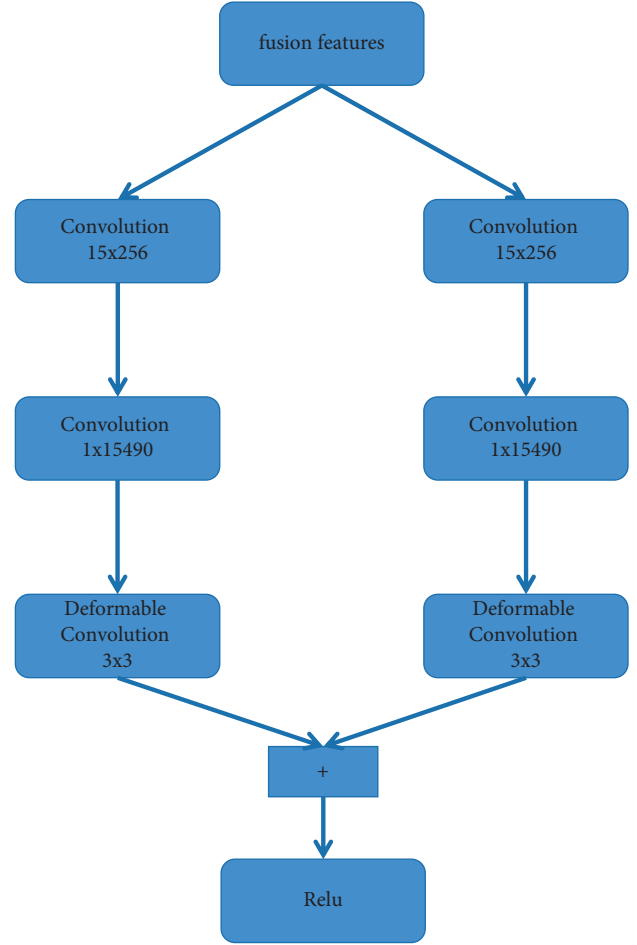

FIGURE 6: Inception module.

convolutional network classification and target location. Due to the arbitrariness of remote sensing target direction, some background irrelevant to the target proposed in the regional proposal network participates in the regional pooling, which seriously destroys the accuracy of the extracted target features. Therefore, the detector extends the PSROI pooling in the original light head R-CNN to a deformable PSROI pooling that can include offset information in the area pooling operation. The calculation formula of deformable PSROI pooling is shown in

$$Y(i,j) = \sum_{p^x, p^y \in \text{bin}(i,j)} z_{i,j} \frac{\left(p_0^x + p^x + \Delta \hat{p}_{ij}^x, p_0^y + p^y + \Delta \hat{p}_{ij}^y\right)}{n}. \tag{17}$$

The deep learning target detector usually uses the frame return to improve the positioning ability. In the two-order target detector, the horizontal rectangular boundary box is usually used to represent the target. Obviously, this representation method lacks the direction information of the target and is not suitable for multidirectional remote sensing target detection. Inspired by the inclined text detection algorithm in the field of text detection, we use the rotatable bounding box to detect the remote sensing target by

estimating the rotation direction of the target in the experiment. In this way, the direction is detected and estimated at the same time in a complete end-to-end way, and no additional detection and estimation pipeline are required. That is to say, like the position of the regression prediction target, the angle is also the information obtained through prediction and estimation. The key of direction estimation method is the trained network model. In this network, the existence of target category, the offset of coordinate position, and the approximate angle are based on the proposed area estimation. Compared with the traditional horizontal rectangular bounding box, the rotating bounding box can not only detect the target, but also estimate the direction of the target at the same time [27]. The rotatable target boundary frame is more compact, and the target can be easily distinguished from the background and has strong robustness to the interference of background pixels, which also helps to improve the detection results. In the designed detector, we regard the angle estimation problem as a regression task by using five parameters (x, y, W, h, and θ) to represent the bounding box of the rotating target with any direction, so as to integrate this problem into the bounding box regression, where θ is the estimated rotation angle value of the target direction. The regression method of the rotation boundary box is redefined as shown in

$$t_x = \frac{(x - x_a)}{w_a, t_y}$$

$$= \frac{(y - y_a)}{h_a},$$

$$t_w = \log \frac{w}{w_a},$$

$$t_h = \log\left(\frac{h}{h_a}\right),$$

$$t_\theta = \theta - \theta_a,$$

$$t_x^* = \frac{(x^* - x_a)}{w_a, t_y^*}$$ (18)

$$= \frac{(y^* - y_a)}{h_a},$$

$$t_w^* = \log \frac{w^*}{w_a},$$

$$t_h^* = \log \frac{h^*}{h_a},$$

$$t_\theta^* = \theta^* - \theta_a.$$

The loss function defined in each proposal box is the sum of ship/nonship classification loss and frame position regression loss. The classification loss function $L_{cls}(p, l)$. is defined as shown in

$$L_{cls}(p, l) = -\log pl.$$ (19)

In the experiment, a total of 51 large-scale remote sensing images were collected from several large ports and nearby waters such as Yokosuka Port and Santiago Port through Google Earth. These images do not necessarily contain dense ship targets, but contain some empty target images or typical detection interference, such as islands, clouds, and fog. The resolution of the selected image format is between 28000×16000 and 3000×3000. In the experiment, the image is cut into small image blocks with the size of 600×600 to 800×800, the cut images without ships are selected as negative training samples, and those with ships are selected as positive samples. A total of 3480 sample images are produced for the classification task, and the training set and test set are divided by $4:1$. After spending a lot of time tagging the targets in the positive sample classification image, the target detection data set with ship coordinate tagging information contains 1432 images, and the total number of ship samples is 3083. The statistics of the number of ships are given in Table 10, where ns represents the number of ships in the figure, and Ni represents the figure. As can be seen from the table, most of the pictures have only one or two boats, and the survey results have an average of 2.15 for boats in each picture. Compared with nearly 10,000 images of 20 categories in VOC archive images, the amount of data in the experiment is enough to train a single-category object detector that combines various antifouling skills. Table 10 provides the statistics on the number of people and ships.

The end-to-end rotating target detector DRBoX designed based on SSD is compared in the first-order detector, and the improved Fast R-CNN rotating target detector and R2CNN++ are compared in the second-order target detector. The improved Fast R-CNN detector is the baseline released for rotating target detection task on DOTA data set. R2CNN++ is a rotating target detection algorithm with excellent detection accuracy. In the experiment, the training model provided by relevant authors is used to test the image of the test set, and the test results are given in Table 11. It can be seen from the table that the two-order detection algorithm has obvious advantages over the first-order detection algorithm, the accuracy of the proposed method on the test set is 93.3070, and the recall rate is 92.7070, which are higher than other methods. This excellent detection performance is due to the proposed multilayer feature fusion method, which makes the detector perform well in small target detection. At the same time, due to the introduction of deformable convolution operation, fewer ships are lost and the recall rate is higher. The comparative experiment is given in Table 11.

Table 12 provides the ablation experimental results of the target detector proposed in this study. It can be seen from the table that when ResNet-50 is directly used instead of feature fusion in the basic network, the accuracy and recall are 92.1070 and 91.3070, respectively. After replacing deformable PSROI pooling with PSROI pooling and removing deformable convolution in perception module, the accuracy and recall of target detection are 90.7% and 90.1070,

TABLE 10: Statistics of images and ships.

Ns	1	2	3	4	5	6	7	8	9	10	11	12
Ni	714	328	189	75	43	28	21	13	7	8	6	0

TABLE 11: Performance comparison experiment of the detector proposed in this study.

Detector	Accuracy (%)	Recall rate (%)	$F1$ score
Faster R-CNN	84.6	82.6	0.835
R2CNN++	88.8	86.3	0.875
Proposed in this study	93.3	92.8	0.94
DR box	80.8	79.3	0.8

TABLE 12: Experimental results of the detector proposed in this study.

Feature fusion	Deformable convolution and pooling	Accuracy (%)	Recall rate (%)	$F1$-score
✕	✕	88.1	87.5	0.878
✓	✕	90.7	90.1	0.904
✕	✓	92.2	91.4	0.917
✓	✓	93.3	92.8	0.94

respectively. If we remove all the improvements, the final accuracy and recall are only 88.2% and 87.4070. The ablation experimental results show that the multilayer feature fusion method, inception module, and deformable pooling operation can significantly improve the detection performance of the detector for remote sensing image target detection tasks. The ablation experimental results of the detector are given in Table 12.

In the large-scale remote sensing image test experiment, we directly use large-scale remote sensing images for ship detection. The detection process includes image clipping and classification (coarse detection stage) and target detection (fine detection stage). In the rough detection stage, the remote sensing image is clipped and sent to the ship/nonship classification model. In the experiment, in order to reduce the damage of clipping to the target in the boundary area, the overlap of adjacent clipped images is set to be slightly larger than the average size of the target of 100. Since ship target detection depends on candidate region image (CRI), the wrong classification of CRI will directly lead to the loss of targets [28]. In the experiment, the confidence is used to represent the classification threshold when classifying and clipping image blocks. The higher the confidence is, the higher the reliability of CRI image is, and the changing confidence will affect the detection performance and efficiency. In order to extract CRI reliably, we perform recall rate (RS) analysis on the image blocks in the test set and determine the confidence threshold of the classification network. When the confidence is greater than 0.4, RS will gradually decrease, and the number of CRI images extracted in the cropped image block will decrease sharply, which may lead to missed detection of ships; if the confidence is less than 0.4, the number of CRI images will increase, resulting in a large number of detection false alarms and computational redundancy in the fine detection stage. In order to extract CRI stably, the curve inflection point value of 0.4 is used as
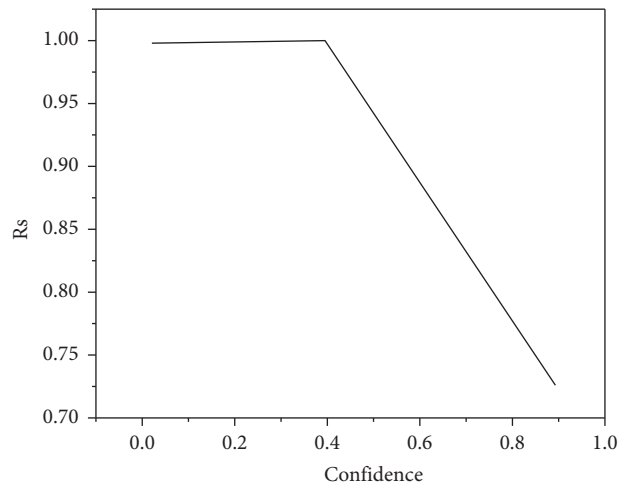

FIGURE 7: Recall rate curve of classified network.

the classification confidence in the experiment, so as to better balance the performance and efficiency. The recall curve of the classification network is shown in Figure 7.

In the fine detection stage, each CRI image is sent to the detector, respectively, and all the image detection results sent to the detector are spliced and mapped back to the original image to obtain the final detection result. In addition, in order to compare with the traditional sliding window detection method of large-scale remote sensing image, the classification process in the coarse detection stage is removed in the experiment, and the image block cut by the sliding window is directly sent to the detector for ship target detection. In the experimental results of traditional sliding window detection, the average running time of each large remote sensing image is 1934.63 Ms. Due to the processing of a large number of empty target images, the average detection time of this method is increased by 709.39 MS

TABLE 13: Statistics of test results of large remote sensing images.

Method	Accuracy (%)	Recall rate (%)	Average time (MS)
Sliding window detection method	92.3	93.7	1934.63
The method proposed in this study	94.2	93.8	1225.24

compared with the proposed detection method of coarse to fine detection strategy. At the same time, the sliding window-based method produces a large number of false alarms and redundancy in the noncandidate region image of the test results, which leads to the reduction of detection accuracy and recall. Comparative experiments show that it has superior detection performance in large-scale remote sensing image detection. It can be seen that the proposed detection method based on coarse to fine strategy has excellent accuracy and speed performance for detecting ship targets in large remote sensing images. The statistics of large remote sensing image test results are given in Table 13.

The classification of the proposed method in the experimental process and the production of target detection data set are introduced in detail, and the relevant training process and specific experimental details are described. In the detector performance comparison experiment, the three remote sensing target detection algorithms that can detect rotating targets are DRBoX, improved Fast R-CNN method, and R2CNN++. On the test data set, the detection accuracy of these three detectors and the proposed detector are 80.9070, 84.5070, 88.7%, and 93.3070, respectively. The results show that the remote sensing ship target detection in this list has the best detection ability. Finally, through experiments on large-scale remote sensing images, it is proved that the strategy for obtaining large-scale remote sensing images from rough to high quality planned in this study can effectively improve the search efficiency and determine the balance between detection accuracy, quality, and speed.

5. Conclusion

At present, goal-oriented research based on in-depth research has become an important technique in goal-oriented research, but goal-oriented research is often used as horizontal box annotation, and the distance between the object in the text and the image seen in the image is quadrilateral or curved of polygonal shape. Using horizontal rectangular box annotation will lead to large background noise and cannot surround the target well. Text and remote sensing targets are generally expressed in three forms: target vertex or contour coordinate point, target rotation rectangular box, and target instance segmentation. These three forms have their own advantages, disadvantages, and difficulties. For the method of regression coordinate points, it is very difficult to determine the regression order of quadrilateral or curved polygon vertices. No matter how to set this order, it will lead to ambiguity in some cases. For the method of regressing the angle of rotating rectangular box, because the angle is periodic, directly using a variable to regress the angle cannot reflect the periodicity of the angle. In addition to the regression-based method, there is also an instance segmentation scheme based on full convolution. For the instance

segmentation scheme, how to separate the pixels belonging to different instances and then combine the pixels belonging to the same instance is its main difficulty.

Data Availability

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

References

[1] Z. Li, H. Wang, H. Zhong, and Y. Dai, "Self-attention module and fpn-based remote sensing image target detection," *Arabian Journal of Geosciences*, vol. 14, no. 23, pp. 1–18, 2021.

[2] Y. You, B. Ran, G. Meng, Z. Li, and Z. Li, "Opd-net: prow detection based on feature enhancement and improved regression model in optical remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 1–17, 2020.

[3] V. P. Patil and C. Ram SinglaRam Singla, "Performance enhancement of image stitching process under bound energy aided feature matching and varying noise environments," *Journal of Computational and Theoretical Nanoscience*, vol. 17, no. 9, pp. 4419–4424, 2020.

[4] R. Wang, B. Jiang, C. Yang, Q. Li, and B. Zhang, "Magan: unsupervised low-light image enhancement guided by mixed-attention," *Big Data Mining and Analytics*, vol. 5, no. 2, pp. 110–119, 2022.

[5] V. R. Pandit and R. J. Bhiwani, "Morphology-based spatial filtering for efficiency enhancement of remote sensing image fusion," *Computers & Electrical Engineering*, vol. 89, no. 1, pp. 1–15, Article ID 107015, 2021.

[6] K. Ahmadian and H. R. Reza-Alikhani, "Single image super-resolution with self-organization neural networks and image laplace gradient operator," *Multimedia Tools and Applications*, vol. 81, no. 8, Article ID 10630, 2022.

[7] H. You, L. Yu, S. Tian, and W. Cai, "Dr-net: dual-rotation network with feature map enhancement for medical image segmentation," *Complex & Intelligent Systems*, vol. 8, no. 1, pp. 611–623, 2022.

[8] J. Li, R. An, Y. Li, and Y. Zhao, "Desert noise suppression for seismic data based on feature enhancement denoising network," *Izvestiya - Physics of the Solid Earth*, vol. 57, no. 6, pp. 935–949, 2021.

[9] K. Zhou, S. K. Oh, and J. Qiu, "Design of ensemble fuzzy-rbf neural networks based on feature extraction and multi-feature fusion for gis partial discharge recognition and classification,"

Journal of Electrical Engineering & Technology, vol. 17, no. 1, pp. 513–532, 2021.

[10] N. Yu, J. Li, and Z. Hua, "Detail enhancement decolorization algorithm based on rolling guided filtering," *Multimedia Tools and Applications*, vol. 81, no. 2, pp. 2711–2731, 2021.

[11] L. Rajesh and P. Satyanarayana, "Evaluation of machine learning algorithms for detection of malicious traffic in scada network," *Journal of Electrical Engineering & Technology*, vol. 17, no. 2, pp. 913–928, 2021.

[12] M. Chen, R. Liu, J. Yang, J. Zhu, and X. Li, "Pavement damage identification method based on point cloud multi-source feature enhancement," *International Journal of Pavement Research and Technology*, vol. 15, no. 2, pp. 257–268, 2021.

[13] J. S. Ko, J. Byun, S. Park, and J. Y. Woo, "Prediction of insufficient hepatic enhancement during the hepatobiliary phase of gd-eob dtpa-enhanced mri using machine learning classifier and feature selection algorithms," *Abdominal Radiology*, vol. 47, no. 1, pp. 161–173, 2021.

[14] T. Wen, W. Li, L. Chen, and X. Jia, "Semi-supervised label enhancement via structured semantic extraction," *International Journal of Machine Learning and Cybernetics*, vol. 13, no. 4, pp. 1131–1144, 2021.

[15] X. Li and Z. Chen, "Single image super-resolution reconstruction based on fusion of internal and external features," *Multimedia Tools and Applications*, vol. 81, no. 2, pp. 1589–1605, 2021.

[16] A. Syed, S. Mishra, and S. Jayanty, "7,7-bis(N, N-diethylethylenediamino)-8,8-dicyanoquinodimethane: effect of ethyl moiety on the photophysical property besides thermal stability," *Journal of Fluorescence*, vol. 32, no. 1, pp. 115–124, 2021.

[17] B. Dong, R. Wang, J. Yang, and L. Xue, "Multi-scale feature self-enhancement network for few-shot learning," *Multimedia Tools and Applications*, vol. 80, no. 25, Article ID 33883, 2021.

[18] S. Chen, Y. Liu, and C. Zhang, "Water-body segmentation for multi-spectral remote sensing images by feature pyramid enhancement and pixel pair matching," *International Journal of Remote Sensing*, vol. 42, no. 13, pp. 5025–5043, 2021.

[19] M. Verma and A. Nandi, "Design and analysis of algan/gan based dg moshemt for high-frequency application," *Transactions on Electrical and Electronic Materials*, vol. 21, no. 4, pp. 427–435, 2020.

[20] Z. Sun, M. Dai, X. Leng et al., "An anchor-free detection method for ship targets in high-resolution sar images," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, no. 99, pp. 7799–7816, 2021.

[21] H. Sun, Q. Liu, J. Wang et al., "Fusion of infrared and visible images for remote detection of low-altitude slow-speed small targets," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, no. 99, pp. 2971–2983, 2021.

[22] P. V. G. D. Prasad ReddyPrasad Reddy, "Blood vessel extraction in fundus images using hessian eigenvalues and adaptive thresholding," *Evolutionary Intelligence*, vol. 14, no. 2, pp. 577–582, 2020.

[23] X. Ma, K. Ji, B. Xiong, L. Zhang, S. FengFeng, and G. Kuang, "Light-yolov4: an edge-device oriented target detection method for remote sensing images," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, no. 99, Article ID 10820, 2021.

[24] Z. Niu, J. Zheng, T. Su, W. Li, and L. Zhang, "Radar high-speed target detection based on improved minimalized windowed rft," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, no. 99, p. 1, 2020.

[25] F. Atighi, A. Safdarinezhad, and R. Karimi, "Target detection improvements in hyperspectral images by adjusting band weights and identifying end-members in feature space clusters," *Journal of Geospatial Information Technology*, vol. 8, no. 3, pp. 103–122, 2021.

[26] Y. Shi, J. Li, Y. Zheng, B. Xi, and Y. Li, "Hyperspectral target detection with roi feature transformation and multiscale spectral attention," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 6, pp. 5071–5084, 2021.

[27] R. Bai, Z. Lv, X. Chen et al., "Precision detection technology: equipping precision oncology with wings," *Journal of Oncology*, vol. 2020, no. 22, 8 pages, Article ID 9068121, 2020.

[28] F. Yang, B. Ma, J. Wang, H. Gao, and Z. Liu, "Target detection of uav aerial image based on rotational invariant depth denoising automatic encoder," *Xibei Gongye Daxue Xuebao/ Journal of Northwestern Polytechnical University*, vol. 38, no. 6, pp. 1345–1351, 2020.