Computer Science and Information Systems 20(4):1289–1310 https://doi.org/10.2298/CSIS230315054Y

M^2F^2-RCNN: Multi-functional Faster RCNN Based on Multi-scale Feature Fusion for Region Search in Remote Sensing Images

Shoulin Yin[1], Liguo Wang[2], Qunming Wang[3], Mirjana Ivanović[4], and Jinghui Yang[5]

[1] College of Information and Communication Engineering, Harbin Engineering University
Harbin, 150001 China
yslin@hit.edu.cn
[2] College of Information and Communications Engineering, Dalian Minzu University
Dalian, 116000 China
wangliguo@hrbeu.edu.cn
[3] College of Surveying and Geo-Informatics, Tongji University
Shanghai, 200092 China
111wqm@163.com
[4] Faculty of Sciences, University of Novi Sad
21000 Novi Sad, Serbia
mira@dmi.uns.ac.rs
[5] School of Information Engineering, China University of Geosciences
Beijing, 100083, China
yang06081102@163.com

Abstract. In order to realize fast and accurate search of sensitive regions in remote sensing images, we propose a multi-functional faster RCNN based on multi-scale feature fusion model for region search. The feature extraction network is based on ResNet50 and the dilated residual blocks are utilized for multi-layer and multi-scale feature fusion. We add a path aggregation network with a convolution block attention module (CBAM) attention mechanism in the backbone network to improve the efficiency of feature extraction. Then, the extracted feature map is processed, and RoIAlign is used to improve the pooling operation of regions of interest and it can improve the calculation speed. In the classification stage, an improved non-maximum suppression is used to improve the classification accuracy of the sensitive region. Finally, we conduct cross validation experiments on Google Earth dataset and the DOTA dataset. Meanwhile, the comparison experiments with the state-of-the-art methods also prove the high efficiency of the proposed method in region search ability.

Keywords: remote sensing images, region search, multi-functional faster RCNN, multi-scale feature fusion, convolution block attention module.

1. Introduction

Region search is widely used in automatic driving, video, image index and remote sensing etc. With the rapid development of deep learning, regional convolutional neural network (RCNN)-based algorithms [1-3] have transformed the traditional manual feature extraction into feature learning, which plays an important role in the field of remote sensing

images processing. In order to quickly obtain useful information from massive remote sensing images, region search has a high application value. Rapid detection of sensitive areas through remote sensing images is of great significance for improving military operational efficiency, civil route planning, safety search and rescue, etc [4,5].

RCNN algorithm adopts selective search(SS) [6] to extract 1000 2000 suggestion bounding boxes from remote sensing images. It adds 16 pixels around each candidate box as the border of the average pixel value. Then it subtracts the average pixel value of the suggestion box from all candidate box pixels, and inputs the results into AlexNet network for feature extraction. Then support vector machine (SVM) classification is used to determine the category of candidate boxes. And the non-maximum suppression (NMS) algorithm is adopted to reduce the number of redundant candidate boxes [7]. Finally, the remaining candidate boxes are modified through the detection box regression model to correct the final position.

The training stage of RCNN is limited to the selection of candidate boxes. The existing problems are summarized as follows:

1. When RCNN reads the image information, it needs to fix the image size, but cutting the image will lead to the loss of the image information.
2. CNN needs to calculate the candidate boxes in each region, and repeated feature extraction will bring huge computational waste.

Compared with RCNN algorithm, Fast RCNN algorithm only extracts the features of each image once, and shares the extracted features in the calculation process to improve the speed of training and testing. In view of the problem that fast RCNN cannot achieve real-time detection and end-to-end training test, Ren et al. [8] proposed faster RCNN algorithm, using regional proposal network (RPN) to replace selective search (SS) algorithm, which could effectively extract the candidate regions in the original image. Different from the aforementioned detection algorithms based on proposal regions, Redmon et al. [9] proposed a YOLO (You Only Look Once) detection algorithm based on regression. The framework abandoned the preset candidate box strategy, regarded the detection task as a regression problem, and directly carried out border regression and category determination in the output layer, which could meet the real-time detection requirements, but the detection accuracy was low. In order to give equal consideration to detection speed and accuracy, Liu et al. [10] proposed the Single Shot MultiBox Detector (SSD) framework, integrating the idea of preset candidate boxes and YOLO. It conducted multi-layer detection through feature pyramid simultaneously to improve detection accuracy. However, the detection effect of small objects was not satisfactory.

The imaging characteristics of remote sensing images are different from those of natural scenes commonly captured by digital cameras. Remote sensing images are mainly shot at high altitude, covering a wide range of ground objects, complex image background, dense target distribution and small size. And the image quality of remote sensing is not as good as that of digital camera [11,12]. If the existing deep learning detection frameworks are directly applied to remote sensing image for region detection, the detection accuracy cannot be the same as that of natural scene image. At present, a large number of scholars have improved the deep learning detection frameworks and applied them to the detection of sensitive areas in remote sensing images. Based on the RCNN model and the cascade AdaBoost algorithm, Tang et al. [13] proposed a coarse-to-fine object region recognition

algorithm, which reduced the amount of computation and improved the detection accuracy of high-resolution remote sensing images. Based on faster-RCNN, Wu et al. [14] proposed a target detection algorithm based on feature fusion combined with soft judgment, which improved the detection accuracy of small target areas. Li et al. [15] overcame the defect of losing small target information in deep feature by integrating shallow feature with deep feature after up-sampling, and improved the fast detection accuracy of target region in aerial remote sensing images. Cui et al. [16] optimized RPN according to the aspect ratio characteristics of the target region. Meanwhile, the online difficult sample mining method was adopted to balance positive and negative samples, which improved the performance of the model algorithm. Aiming at the uncertainty of the direction of the target region in remote sensing images, Pazhani et al. [17] integrated the rotating region network into the Fast RCNN, introduced the convolution layer in front of the fully connected layer of the network, reduced the dimension of the feature map. It improved the performance of the classifier, and achieved better detection results. Based on the YOLO-V2 detection framework, Anuar et al. [18] introduced the transfer learning strategy to improve the detection accuracy of high-resolution remote sensing image regions in small samples, but the detection accuracy of small targets was not high. Based on the SSD detection framework, ResNet50 was used to replace the front-facing network VGG16, and the target detection was carried out on the aerial data set [19]. The detection effect was significantly improved, but the error was also relatively large. Aiming at the problem of low detection accuracy of small targets in complex scenes, Chen et al. [20] proposed MultDet, a lightweight multi-scale feature fusion detection framework, which improved the characterization ability of small scale aircraft targets and realized high-precision detection of aircraft targets.

Previous researches on remote sensing image target region extraction mainly rely on the basic features of the image, such as spectrum, shape, contour, texture, color, shadow, etc. For example, Zheng et al.[21] proposed an object-based Markov Random field (OMRF) model for building extraction. The model constructed a weighted region adjacency map based on region size and edge feature information. Then it used OMRF with a region penalty term to achieve accurate building area extraction. Mag et al. [22] proposed a region extraction method based on saliency analysis, which extracted multi-scale texture and edge features of remote sensing images by Fourier transform and adaptive wavelet. Wang et al. [23] used Extended multi-resolution segmentation (EMRS) and back-propagation (BP) network to extract building areas. EMRS was used to represent multi-scale spatial resolution features, and BP network was used to classify pixels with different building areas. Ni et al. [24] proposed a local competitive super-pixel segmentation method, which could effectively integrate spatial resolution and multi-scale features of remote sensing images, and completed accurate extraction of building areas. Zhu et al. [25] proposed a building extraction method based on mixed sparse representation, which divided remote sensing images into subgraph combinations with different components, then used sparse representation to express different subgraph features. Finally, support vector machine was adopted to complete the extraction of building areas. The target region extraction methods based on the above basic features have achieved a certain effect. However, due to the lack of extraction of deep semantic features and global spatial features in remote sensing images, the segmentation and extraction results still have some problems, such as boundary information loss and incomplete shape structure.

In summary, this paper collects remote sensing image data sets from Google Earth, DOTA data sources, and then establishes a typical region target data sets of remote sensing images by manual annotation. Therefore, we propose a multi-functional faster RCNN based on multi-scale feature fusion model for region search. The stages of region search method are as follows. Firstly, the feature extraction network is based on ResNet50 and the dilated residual blocks are utilized for multi-layer and multi-scale feature fusion. Then, a path aggregation network with a convolution block attention module (CBAM) attention mechanism is added in the backbone network to improve the efficiency of feature extraction. Thirdly, the extracted feature map is processed, and RoIAlign is used to improve the pooling operation of regions of interest and it can improve the calculation speed. Finally, in the classification stage, an improved non-maximum suppression is used to improve the classification accuracy of the sensitive region. Meanwhile, abundant experiments also show the effectiveness of the proposed method.

The structure of this paper is as follows. In section 2, the related works are introduced including Faster-RCNN, RPN. Section 3 detailed states the proposed method for region search in remote sensing images. Rich experiments are shown in section 4. A conclusion is shown in section 5.

2. Related Works

In recent years, deep learning [26,27] has made new progress, and convolutional neural network (CNN) has been applied to target region detection. Since the appearance of the image classification competition based on the large image database ImageNet, various deep learning detection algorithms have been proposed successively, which are mainly divided into One-stage and Two-stage. Herein, YOLO-v3 [28] is a representative algorithm in One-stage. The representative algorithms in the two-stage method are Faster RCNN and other algorithms. RCNN-based algorithms have been developed from Faster RCNN to Mask RCNN. By using RoIAlign, a mask branch is added to achieve the detection and segmentation at the instance level. One-stage algorithm is superior to Two-stage algorithm in detection speed, but it is lower that two-stage algorithm in terms of accuracy. However, these mainstream target detection algorithms aim at the image of the natural scene, without considering the directivity of the target and other features.

Deep learning has achieved great success in detecting the target region in natural scenes. Therefore, a large number of researches on remote sensing image target region detection [29,30] based on deep learning continue to emerge. An important characteristic of the target region to be detected in remote sensing images is the uncertainty of its direction and shape. In addition, most target regions in remote sensing images are densely distributed, so the horizontal box method may also affect the results of non-maximum suppression (NMS) after detection, resulting in poor detection effect. In reference [31], a text detection algorithm based on the rotating region was proposed. The algorithm firstly used the region proposal network to generate horizontal candidate boxes, and then used the features after multi-scale pooling to predict slanted text boxes. In reference [32], Region Proposal Network (RPN) in Faster R-CNN network was improved and it added rotation information, so that the RPN network could directly generate an angled prediction box for multi-directional text detection.

2.1. Faster-RCNN

Faster RCNN is a region-based convolutional neural network framework, which mainly consists of four parts: feature extraction layer, RPN layer, ROI Pooling layer and detection subnetwork, as shown in Figure 1. Feature extraction layer extracts features of input images through convolutional neural networks. The candidate region generation network carries out convolution operation on the obtained image feature map to generate the candidate frame that may contain the object. The region of interest (ROI) pooling layer [33] converts the region feature maps corresponding to different candidate boxes into feature maps with the same size. The detection network identifies the target in the ROI and corrects the position of the candidate box to get the final detection result.

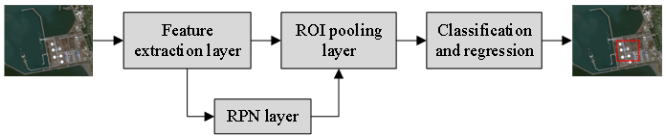

Fig. 1. Structure of Faster RCNN

2.2. Region Proposal Network (RPN)

RPN is a full convolutional network, which reduces the number of candidate boxes in the detection process and improves the detection efficiency of Faster RCNN. It can be seen from Figure 2 that RPN network mainly has two branches: one is the classification layer, which is used to classify positive and negative samples, and the other is the regression layer, which is used to regress the candidate box position of positive samples. The RPN network implementation process is as follows. A 3×3 sliding window is used to slide the feature graph (suppose the size is $N \times 60 \times 60$), and each pixel of the feature graph is traversed to generate a low-dimensional feature graph ($256 \times 60 \times 60$). In addition, k predefined candidate boxes are generated at each pixel position on the feature graph. Then two 1×1 convolution operations are performed on the low-dimensional feature graphs. $2k$ probability values and $4k$ candidate box offsets are obtained at each pixel. Finally, combined with the pre-defined candidate boxes, post-processing operations such as cross boundary clipping, small candidate box removal and Non-Maximum Suppression (NMS) are carried out to obtain the ROI candidate boxes.

The training of RPN network adopts multi-task loss function, as shown in Equation (1):

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

Where i is the index value of anchor. p_i and t_i are the $i - th$ anchor containing the probability value of the target and four coordinate values corresponding to anchor respectively. N_{cls} and N_{reg} are standardized constants. λ is the equilibrium coefficient.

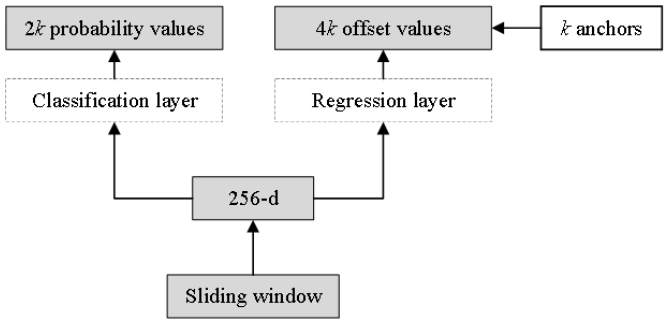

Fig. 2. RPN structure

When anchor is a positive sample, $p_i^* = 1$, otherwise $p_i^* = 0$. t_i^* is the coordinate value of the real region corresponding to the positive sample. L_{cls} and L_{reg} are the classification loss function and the regression loss function respectively. The expressions are as follows:

$$L_{cls}(p_i, p_i^*) = -log[p_i p_i^* + (1 - p_i)(1 - p_i^*)]. \tag{2}$$

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*). \tag{3}$$

Where R is Smooth L1 function, reflecting the robustness of the function, it is defined as follows.

$$R(x) = \begin{cases} 0.5x^2 & |x| \leq 1 \\ |x| - 0.5 & |x| > 1 \end{cases} \tag{4}$$

3. Proposed M^{2}F^2-RCNN for Region Search

Considering the detection application requirements of remote sensing satellite image, this paper adopts the Faster RCNN framework with higher detection accuracy as the basic prototype. The proposed M^{2}F^2-RCNN region search model in this paper is shown in Figure 3.

Our contributions are mainly in the following three aspects:

1. In the basic network stage, we construct a feature pyramid network to perform multi-scale feature fusion. We added path aggregation network with Convolution Block Attention Module (CBAM) to improve feature extraction efficiency.
2. RoIalign is used to replace RoI pooling, and a convolution layer is added into the classification network to improve the classification efficiency.
3. A new non-maximum suppression algorithm is improved to the overlap problem of large amount of similar candidate boxes.

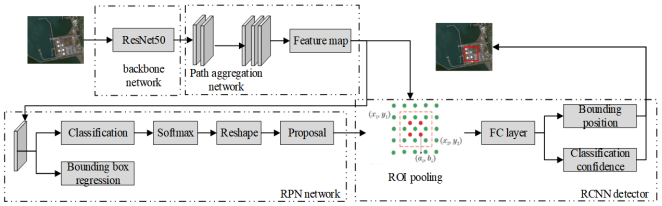

Fig. 3. M^{2}F^2-RCNN structure

3.1. Multi-scale Feature Fusion

Compared with the original feature extraction layer VGG, ResNet50 has a deeper network layer and can extract deeper and more abstract features [34]. Moreover, the residual structure of ResNet50 is conducive to solving the problems of gradient dispersion and gradient explosion when the network performance is not significantly improved. Therefore, ResNet50 is selected as the basic feature extraction network in this paper. Table 1 shows the Resnet50 network structure. In Table 1, two multiplied numbers indicate the size of the convolution kernel. The next number is the number of convolution kernels. In the case of conv1, 7×7 means the size of the convolution kernel and 64 means the number of convolution kernel. The matrix represents the residual block, and the number in the matrix represents the composition of the residual block. Taking conv3 as an example, conv3 is composed of 4 residual blocks, each of which is composed of 128 convolution kernels with 1×1 size, 128 convolution kernels with 3×3 size, and 512 convolution kernels with 1×1 size. The step size in convolution kernel with 3×3 size in the first residual block is 2, and the rest step size is 1.

Table 1. Structure of ResNet50

Layer	Structure
Convolutional layer (conv1)	7×7, 64, step size=2
Pooling layer	3×3, step size=1
Convolutional layer (conv2)	$\begin{matrix} 1 \times 1 & 64 \\ 3 \times 3 & 64 \\ 1 \times 1 & 256 \end{matrix} \times 3$, step size=2
Convolutional layer (conv3)	$\begin{matrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 512 \end{matrix} \times 4$, step size=2
Convolutional layer (conv4)	$\begin{matrix} 1 \times 1 & 256 \\ 3 \times 3 & 256 \\ 1 \times 1 & 1024 \end{matrix} \times 6$, step size=2
Convolutional layer (conv5)	$\begin{matrix} 1 \times 1 & 512 \\ 3 \times 3 & 512 \\ 1 \times 1 & 2048 \end{matrix} \times 3$, step size=2

However, if it directly uses Convl-5 as the feature extraction layer in the Faster RCNN structure, it cannot significantly improve the model detection accuracy. In order to solve

the problem that the feature extraction layer is deepened and the detection accuracy is not significantly improved, It considered conv5 and the full connection layer as the final detection subnetwork based on the structure of Networks on Convolutional Feature Maps (NoCs). The specific structure is shown in Figure 4. Moving conv5 to the detection network can improve the classifier performance and thus improve the overall detection accuracy.

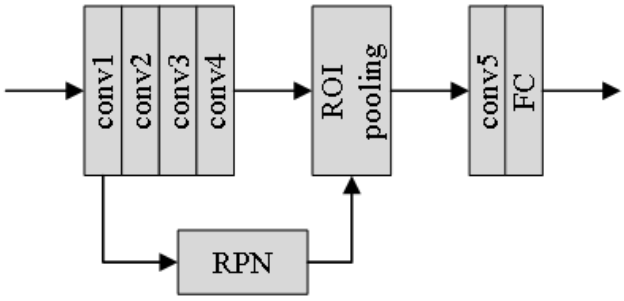

Fig. 4. ResNet50 with Faster RCNN

Figure 5(a) is the original residual block in Resnet50. The dilated residual blocks in FIG. 5(b) and 5(c) are the structures formed by introducing the dilated convolution with expansion rate 4. With the same spatial resolution, the dilated residuals can enlarge the receptive field of the deep network, which is conducive to extracting the deep semantic information of the target. In the experiment, a feature fusion model is designed by using the dilated residual block. The specific structure is shown in Figure 5(d).

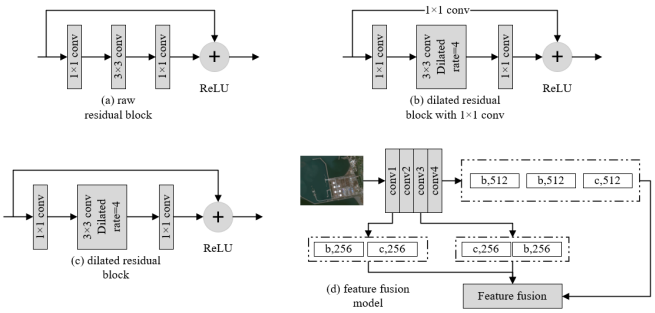

Fig. 5. Feature fusion process

Firstly, because conv5 in the original ResNet50 is moved to the detection network, the feature extraction layer only has the first four layers and the network depth is shallow. So after conv4 layer, the dilated residual block with 1×1 convolution layer mapping (Figure 5(b)) is used once time, and then the dilated residual block with 1×1 convolution

layer mapping (Figure 5(c)) is used twice as the new fifth layer in the feature extraction network. The combination of dilated residuals ensures that the spatial resolution of fifth layer is consistent with conv4, which can reduce the loss of feature information. In order to reduce the amount of computation and memory consumption, the channel number of the residual block is set to 512. Secondly, due to the large area size in the remote sensing image, the spatial resolution decreases and the semantic information of small targets is seriously lost after down-sampling. Hence, the features generated by convl, conv3 and the new fifth layer are fused to make full use of the deep semantic information under low spatial resolution and the shallow geometric, textural information under high spatial resolution. However, although the shallow feature retains high spatial resolution and abundant small target information, the shallow information is too shallow, and the directly fusion of shallow feature and deep feature cannot significantly improve the regional detection accuracy. Therefore, before feature fusion, a structure consistent with the fifth layer is introduced after the features of convl and conv3. The number of residual block channels is set to 256. To fully learn the target information under high spatial resolution, improve the positioning accuracy of large targets, and enhance the detection ability of the network model, the down-sampled shallow features and the up-sampled deep features are fused, so the improved feature extraction network is obtained.

The resolution of remote sensing image is too large, resulting in many small target areas. However, Faster RCNN uses part of the VGG Net network layer as shallow feature extractor to extract basic features such as points and edges of targets. Therefore, the detection effect of small target region is not good. In this paper, ResNet50 is selected as the basic feature extraction network, and multi-scale features are added to enhance the detection ability of small target regions. The ResNet solves the degradation problem when the traditional network is deepened, and the deepest network can reach 152 layers. Compared with traditional networks, deep residual networks have better generalization ability and lower complexity. ResNet is deeper than the VGG network. It is able to learn detailed features in the image. ResNet adds a batch normalization layer between the convolution and pooling layers to speed up training, while residuals are used to make training the depth model easier. Residual network introduces a learning framework based on residual blocks. The input can be propagated forward faster through cross-layer connections.

Faster RCNN is only classified according to the output features of the last layer in the basic network, which requires less computing and less memory. However, the features of the last layer belong to the high-level features, so the network is not expressive enough for small-scale targets. Generally, the high level features of convolutional neural networks are characterized by low resolution and high level semantic information. In contrast, low-level features are characterized by high resolution, low-level semantic information. By combining high level features with low level features, semantic information at multiple scales can be utilized at the same time. According to the idea of reference [19], feature pyramid is added to feature extraction in this paper to improve the final detection effect, as shown in Figure 6. The feature extraction process forms a bottom-up path, a top-down path and a horizontal connection. Formally, for the ROI with width w and height h, it is assumed that k is the reference value, representing the number of feature layers. k is the number of feature levels corresponding to the ROI. The formula assigned to the feature pyramid is:

$$k = [k_0 + b(\sqrt{wh}/224)]. \tag{5}$$

The feed-forward calculation of convolutional neural network is a bottom-up path. The basic network in this paper uses the characteristics of each residual block to activate the output corresponding to the output of different convolutional layers. It has different step sizes, which can deal with small targets region well. Compared with the feature extracted only by the last convolutional layer, this structure can utilize more high-level semantic information. For the small target region, the operation on the larger feature map increases the resolution of the feature map, so that more useful information about the small target region can be obtained.

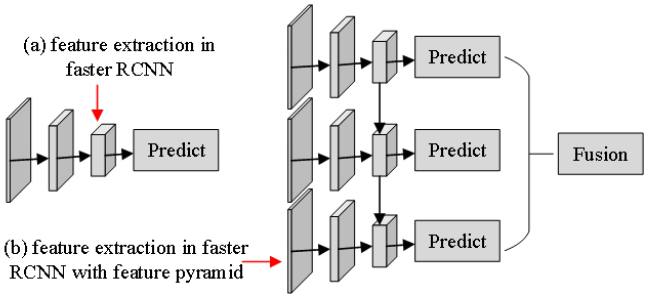

Fig. 6. Multi-scale feature extraction process

In remote sensing images, there is a large scale difference between different regions, and region targets with different scales have different features, so feature maps of different scales should be used for detection. In the underlying feature map, there are more details and the location of region targets is clear, which is suitable for detecting region targets with small scale. In the top-level feature map, the semantic information is rich, which is suitable for the detection of large scale region targets. However, the Faster RCNN model only uses the top-level feature map for target detection. The top-level feature map has rich semantic information but less detailed information after multiple sub-sampling operations, so the detection effect for small targets is not ideal.

In 2017, Lin et al. [36] proposed Feature Pyramid Network (FPN). FPN network fuses multi-scale feature maps and uses the fused feature maps for target detection. In this way, details in the underlying feature map can be fully utilized. However, the structure of FPN network is relatively concise, which leads to two shortcomings of FPN network for complex target detection tasks. Firstly, the FPN network uses the nearest neighbor interpolation algorithm with low precision to realize the up-sampling operation. The loss of the top layer information in the downward transmission process is too large, resulting in less semantic information fusion in the feature graph near the bottom. Secondly, the feature fusion path of FPN network is top-down, and a large number of details in the bottom feature map cannot be fused in the top feature map. As a result, the contour information

between overlapping regions cannot be effectively extracted when the target detection is carried out on the top feature map.

To solve the above problems in FPN networks, a path aggregation network with Convolution Block Attention Module (PACBAM) is proposed in this paper aiming to enhance the fusion effect between feature maps with different scales. Figure 7 shows the PACBAM network structure. The dimensions of feature map F1, F2, F3 and F4 are $336 \times 192 \times 256$, $168 \times 96 \times 512$, $84 \times 48 \times 1024$ and $21 \times 12 \times 2048$, respectively. The PACBAM network first convolves the four feature graphs with 1×1 to unify them into 256-channel. The bilinear interpolation algorithm is used to realize the double up-sampling operation of the feature graph, and the up-sampled feature graph is added from top to bottom to realize the first fusion of the feature graph. The convolution kernel with the step size of 2 and the size of 3×3 is used for down-sampling of the fused feature graph. The sub-sampled feature maps are added from bottom to top to realize the second fusion of feature maps. PACBAM network adopts bilinear interpolation to realize the up-sampling operation of feature graph. The bilinear interpolation carries out linear interpolation in two directions respectively, and the pixel value information of four pixel points around the target point can be utilized. The nearest neighbor interpolation only selects the nearest pixel as the pixel value of the target point. Bilinear interpolation reduces the information loss in the process of feature graph transmission and improves the up-sampling accuracy. In addition, PACBAM network adds a bottom-up feature fusion path on the basis of FPN network to realize multiple fusion of multi-scale feature maps, so that the details of the bottom layer can be better transferred to the top layer feature map.

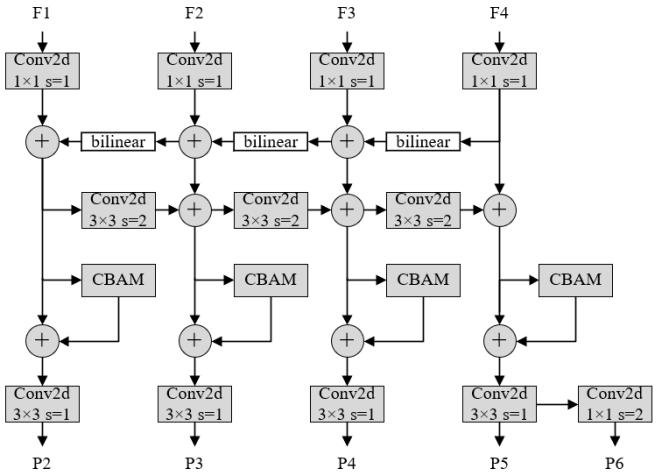

Fig. 7. MPACBAM network

The feature graphs that have been fused twice are weighted using CBAM attention mechanism, which helps to obtain the feature information that contributes more to the detection results. CBAM includes a channel attention module and a spatial attention module. Its structure is shown in Figure 8.

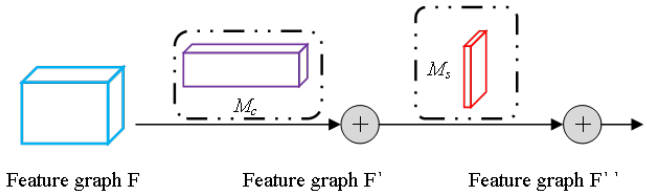

Fig. 8. CBAM structure

Here M_c is the weight of channel attention, M_s is the weight of space attention.

The channel attention module focuses on the real content of the target to be detected. The specific calculation method is as follows.

$$M_c(F') = \sigma(MLP(AvgPool(F) + MLP(MaxPool(F)))). \tag{6}$$

Where $M_c(F')$ is the channel attention weight. σ is the Sigmoid function. MLP denotes multilayer perceptron. $AvgPool$ and $MaxPool$ are global average pooling and maximum pooling, respectively. F represents the input feature graph.

Spatial attention module focuses on the location information of the target to be detected, which can reduce the interference of background information. The specific calculation method is:

$$M_s(F'') = \sigma(f^{7\times7}([AvgPool(F'); MaxPool(F')])). \tag{7}$$

Where $M(F'')$ is the weight of spatial attention. $f^{7\times7}$ means that the convolution kernel size is 7×7. F' represents the feature map weighted by channel attention.

The weighted feature graph is convolved with a convolution kernel with step size of 1 and size of 3×3, and a total of 4 feature graphs from P2 to P5 are obtained. The P6 feature map is obtained by the maximum pooled up-sampling operation with the step size of 2 on the P5 feature map. Finally, there are 5 feature maps from P2 to P6.

3.2. Pooling Layer of Region of Interest

The RoI Pooling layer proposed in Fast R-CNN can significantly accelerate training and testing and improve the detection accuracy. For boxes with different sizes, RoI Pooling can also obtain feature maps with fixed sizes. Floating point rounding occurs when the image reaches the feature mapping through the convolutional network to obtain the position of the candidate frame and when the RoI Pooling is applied to the position of each small grid. These two quantifications are likely to lead to the deviation of the position of the candidate box. First, the continuous coordinates (x_1, y_1) and (x_2, y_2) of ROI should be integer quantized. Let downward rounding and upward rounding of coordinate component x be $floor(x)$ and $cell(x)$ respectively. The discrete eigenvalues w_{ij} on the feature graph are calculated by summative operation, and the nearest neighbor sampling operation is carried out. Quantization will result in a mismatch between the image and the

feature map. The calculation formula of ROI pooling result $r_{pooling}(x_1, y_1, x_2, y_2)$ on the feature map is:

$$r_{pooling}(x_1, y_1, x_2, y_2) = \frac{\sum_{i=floor(x_1)}^{cell(x_2)} \sum_{i=floor(y_1)}^{cell(y_2)} w_{ij}}{(cell(x_2) - floor(x_1) + 1) \times (cell(y_2) - floor(y_1) + 1)}. \tag{8}$$

The RoIAlign Pooling method is proposed in Mask R-CNN algorithm, which can effectively reduce the errors generated by RoI Pooling quantization operation. RoIAlign takes no quantization and uses bilinear interpolation to convert pixel values on the image to floating point numbers, thus converting the entire feature aggregation processing into a continuous operation. In order to eliminate the quantization error of RoI Pooling discretization, the center point obtained by bilinear interpolation operation within the range from the point to the top, bottom, left and right discretization points of $N = 4$. We use formula (3) to operate each successive point (a_i, b_i) once. Bounding box is the corresponding region of ROI region on the feature graph, so the pooling result of ROI region align can be expressed as:

$$r_{align}(x_1, y_1, x_2, y_2) = \sum_{i=1}^{N} f(a_i, b_i)/N. \tag{9}$$

Where $f(\cdot)$ represents the eigenvalue of the feature graph. N is the number of feature points.

The connection model of fully connected layer of convolutional neural network is different from that of convolutional layer and pooling layer, which contains a large number of parameters. According to the study of reference [21], fully connection layer is easy to lead to over-fitting, thus reducing the generalization ability of the network. The main idea of improving the classification network is to reduce the number of parameters in the full connection layer to reduce the computation and prevent over-fitting. Therefore, this paper modifies the classification network, as shown in the dotted line box at the classification network in Figure 2. A convolution layer is added before the fully connected layer to reduce the number of parameters in the feature graph and make the classifier more powerful. In this paper, 3×3 convolution kernel is used. Small convolution kernel cannot only realize the function of large convolution kernel, but also reduce the number of parameters and speed up the calculation. In addition, this operation can prevent the over-fitting phenomenon caused by the large dimension of the fused feature, and reduce the feature size by 1/2, which is convenient for the subsequent calculation.

3.3. Bounding Box Optimization with Improved NMS

The NMS algorithm is widely used in edge and target detection. It can solve the problem that a large number of candidate boxes overlap when the target is surrounded by a large number of candidate boxes during classification. The steps of the NMS algorithm are as follows:

1. Sorting all the candidate boxes according to their score and selecting the candidate box with the highest score.

2. Comparing the remaining candidate boxes with the candidate box (highest score) in turn. If the overlap area between the two boxes is greater than a certain threshold, the box will be deleted.
3. Repeat step 2 by selecting the highest score box from the unprocessed candidate box (one that does not overlap the highest scoring box), leaving only the target in the optimal box.

In object detection process, the NMS algorithm can be understood as a process of scoring boundary boxes, and its linear weighting function can be expressed as:

$$s_i^{NMS} = \begin{cases} s_i^{NMS}, X_{IoU}(M, b_i) < N_t \\ 0, X_{IoU}(M, b_i) > N_t \end{cases} \tag{10}$$

Where s_i^{NMS} is the IOU corresponding to the $i - th$ prediction box. N_t indicates the suppression threshold. b_i indicates the $i - th$ prediction box to be filtered. M is the boundary box with the maximum score. X_{IOU} is the union of the intersection ratio of the predicted boundary frame area A and the actual boundary frame area B, which can be expressed as:

$$X_{IOU} = \frac{A \cap B}{A \cup B}. \tag{11}$$

As can be seen from the preceding steps, the original prediction box will be deleted when detecting the same type of objects with a large amount of overlap. As shown in Figure 9, the detection result should output two boxes, i.e. 0.75 and 0.90. However, the traditional NMS algorithm may delete the solid box with low confidence. As a result, the confidence level of the box with a confidence level of 0.75 becomes 0.00 or 0.35. The detected actual output has two boxes. If the IOU of the solid and dotted boxes on the NMS is greater than the threshold and the score of the solid boxes is low, so the IOU will be deleted. As a result, only the targets in the dotted box can be detected, reducing the recall rate of the targets. In order to solve this problem, the Soft-NMS algorithm is used to replace the NMS, and the score can be recursively scored according to the current score, instead of directly deleting the adjacent boxes with lower scores. When the same type of objects are highly overlapped, the situation of deleting the prediction boxes by mistake is reduced. Soft-NMS algorithm does not need to introduce any hyperparameters in the training stage. The hyperparameters used to adjust the Soft-NMS algorithm only occur during the test or demonstration phase and do not increase the computational complexity. The procedure of the Soft-NMS algorithm is as follows:

1. Grouping tags according to different types to predict all candidate regions in different tags.
2. All boxes in each category are denoted as E, and the set of filtered boxes are denoted as D. a) Select box M with the highest score and add it to D; b) Calculate the overlap area between the remaining frame and M. If it is larger than the set threshold N, it will be discarded; otherwise, it will be retained; c) If all boxes obtained in step b) are empty, return to Step 2), otherwise continue to perform step a);
3. After the above process is completed, it keeps all categories in the collection of valid boxes.

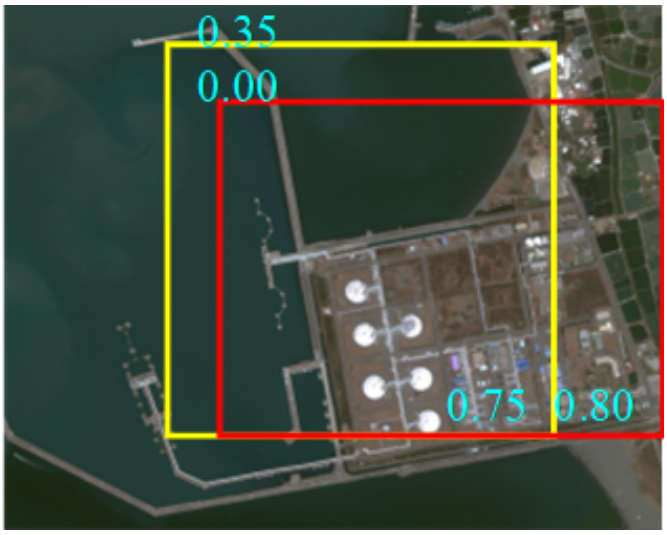

Fig. 9. NMS process

The linear weighting function in Soft-NMS can be expressed as:

$$s_i^{Soft} = \begin{cases} s_i^{Soft}, X_{IoU}(M, b_i) < N_t \\ s_i^{Soft}[1 - X_{IoU}(M, b_i)], X_{IoU}(M, b_i) > N_t \end{cases} \tag{12}$$

Where s_i^{Soft} is the classification score corresponding to the $i - th$ prediction box. Reference [37] compared the mAP of Soft-NMS and NMS under different thresholds with the same data. The results showed that the best effect was achieved when the IOU was set between 0.45 and 0.70, so the IOU value was set as 0.70 in the next experiment.

4. Experiments and Analysis

4.1. Procedure of Experiment

This paper builds a deep learning framework based on MATLAB2017a, and the relevant configuration of the test platform is Intel(R)Xeon(R) 2.30GHz CPU, NVIDIA Tesla 1060 GPU. The optimizer is stochastic gradient descent (SGD) algorithm. Initial learning rate is 5×10^{-3}, momentum factor is 0.9. Decay learning rate is 8×10^{-4}. Iteration number is 7000, batch size is 4, epoch number is 20. In RPN networks, the threshold of Soft-NMS is 0.3. The number of prediction boxes retained by each image after being suppressed by Soft-NMS is 2000.

4.2. Datasets

Different from natural images, remote sensing images have different target sizes and directions. The background environment is also complex. This paper collects and compares

a variety of remote sensing data sets, and analyzes their advantages and disadvantages. Considering that the data quality of DOTA data set is high, with the highest resolution of 4000×4000 pixels, and it contains up to 15 categories. The samples are more balanced than other data sets, and the scale changes greatly, so it is taken as the data set of the experiment. At the same time, part of the data is collected from Google Earth and marked by roLabellmg tool to provide data for the experimental test. There are 3000 images in the sample set, including 1500 training samples, 500 verification samples and 500 test samples.

The positive sample threshold of IoU is set at 0.7 and the negative sample threshold is set at 0.3 to maintain a certain ratio of positive and negative samples. When the IoU between anchor and GT(ground truth) is greater than 0.7, the anchor is considered as a positive sample. When IoU is less than 0.3, the anchor is considered as a negative sample. The IoU of anchor is within the (0.3,0.7). The IoU threshold of the NMS is set to 0.7.

4.3. Evaluation Index

Precision, Recall and Average Precision (AP) values under different IOU thresholds are used as evaluation indicators. Accuracy is the ratio of the number of target regions correctly detected by the model to the total number identified as target regions in the test set. Recall rate is the ratio of the target area correctly detected by the model to the total number of samples in the target area in the test set. AP value is a comprehensive evaluation index, which is determined by the area under P-R curve drawn by Precision and Recall. The calculation formula of each index is as follows:

$$Precision = \frac{TP}{TP + FP}. \tag{13}$$

$$Recall = \frac{TP}{TP + FN}. \tag{14}$$

$$AP = \int_0^1 P(r)dr. \tag{15}$$

Where FN is the target sample wrongly considered as the background. TP represents the correctly identified target sample. FP represents the background sample of the misjudged target. r represents the Recall value, and $P(r)$ represents the Precision value corresponding to the r value.

4.4. Training Data Analysis

The model training results are shown in Table 2 and Figure 10, where the mAP is used to test the training effect of the model, and mAP represents the comprehensive detection accuracy of the model. With the increase of the iteration number, the mAP curve grows smoothly. The reason is that the sample quantity of the last iteration in each Epoch is less than that of the previous iterations, and the sample representativeness of the last

iteration is poor, which slightly changes the iteration direction of the model and leads to the larger loss value. With the increase of the iteration number, the distance between the model training result and the global optimal solution gradually decreases. At this time, the decrease of the iteration number of samples will not significantly change the direction of iteration. This phenomenon will not affect the final training result of the model. After 150 network iterations, the model gradually converges, with mAP tending to 0.68.

Table 2. The comparison of mAP with iterations

Iteration number	mAP
50	0.55
100	0.65
150	0.68
200	0.68
250	0.68
300	0.68

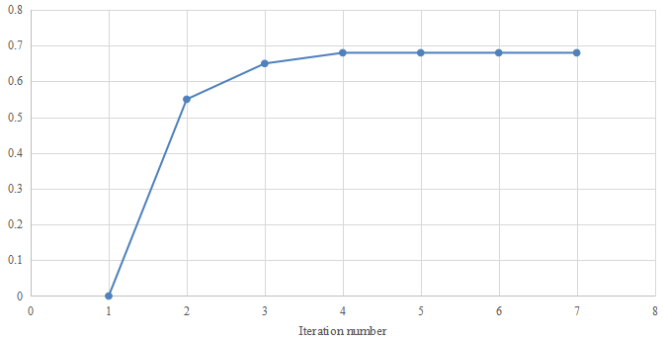

Fig. 10. The curve of loss value and average accuracy with the number of iterations

4.5. Comparison of Pooling Methods

Table 3 shows the detection results of five types of targets by two pooling methods. It can be seen that, compared with RoI Pooling method, the detection accuracy of RoIAlign pooling method is increased by 3%. Because there are a large number of small and medium targets in remote sensing images and they are very dense, RoIAlign reduces pixel deviation during pooling, so the detection rate is improved to a certain extent compared with RoI pooling method. In addition, this paper also calculates the test time of the two methods, testing multiple images from the DOTA data set for testing, and calculates the average detection time of the two methods.

Table 3. AP comparison with different pooling methods/%

Method	small-vehicle	large-vehicle	Plane
RoIPooling	60.1	83.2	89.6
RoIAlign	65.5	86.5	92.7

4.6. Comparison with different State-of-the-art methods

In this paper, the training model of the M^2F^2-RCNN is used to carry out relevant tests on the DOTA data set, including small vehicle, large vehicle, Plane. Table 5 shows the values of Precision, Recall and AP.

Table 4. AP results with M^2F^2-RCNN

Object	Precision/%	Recall/%	AP/%
small-vehicle	78.3	83.6	65.5
large-vehicle	89.7	84.5	86.5
Plane	91.7	93.4	92.7

Table 4 shows the results of target recognition on DOTA data set by the M^2F^2-RCNN in this paper. It can be seen that large vehicle and plane have good recognition effect, with an average accuracy of more than 85%. On the other hand, the identification effect of small vehicle is poor, with an average accuracy of 65.5%. The reason is that large vehicle and plane have significant shape, color and texture features, and the environment is relatively simple, it is relatively easy to identify. However, Bridges are generally located in areas with dense ground objects, with a large length-width ratio and a small number of plane in the data set, so it is difficult to detect.

In order to demonstrate the effectiveness of the M^2F^2-RCNN, other advanced methods are used for comparison. The models used are AOR [38], HASS [39], EAN [40] and OSHP [41], and the results are shown in Table 5.

Table 5. AP results with different methods/%

Object	AOR	HASS	EAN	OSHP	M^2F^2-RCNN
small vehicle	58.2	58.7	59.5	63.7	65.5
large vehicle	75.6	78.2	81.7	83.6	86.5
Plane	80.9	82.1	84.7	85.2	92.7

Table 5 shows that compared with AOR, HASS, EAN and OSHP, the M^2F^2-RCNN is greatly improved in AP. M^2F^2-RCNN in Table 5 has higher AP values than other methods. For large vehicle, AP of M^2F^2-RCNN is 92.27%, which is higher than that of AOR (75.6%), HASS (78.2%), EAN (81.7%) and OSHP (83.6%). Looking through the data of the whole table, it is not difficult to find that the proposed scheme can effectively detect the region, and the results are satisfactory.

5. Conclusions

For the problems existing in the Faster RCNN algorithm in region detection, we propose M^2F^2-RCNN model. Firstly, the feature extraction network is based on ResNet50 and the dilated residual blocks are utilized for multi-layer and multi-scale feature fusion. Then, a path aggregation network with a convolution block attention module (CBAM) attention mechanism is added in the backbone network to improve the efficiency of feature extraction. Thirdly, the extracted feature map is processed, and RoIAlign is used to improve the pooling operation of regions of interest and it can improve the calculation speed. Finally, in the classification stage, an improved non-maximum suppression is used to improve the classification accuracy of the sensitive region. The experimental results show that the M^2F^2-RCNN can get better detection results. Future work will apply this theme in practical engineering. How to use deep learning technology to realize the rapid detection of remote sensing image is still the focus of future research.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grant 62071084 and National Natural Science Foundation of China under Grant 62001434. Also supported by Talents Project of the State Ethnic Affairs Commission.

References

1. Wang Z, Li P, Cui Y, et al. "Automatic Detection of Individual Trees in Forests Based on Airborne LiDAR Data with a Tree Region-Based Convolutional Neural Network (RCNN)," *Remote Sensing,* 2023, 15(4): 1024.
2. Seetharaman K, Mahendran T. Leaf Disease Detection in Banana Plant using Gabor Extraction and Region-Based Convolution Neural Network (RCNN)[J]. Journal of The Institution of Engineers (India): Series A, 2022, 103(2): 501-507.
3. Y. Yuan, Z. Xu and G. Lu, "SPEDCCNN: Spatial Pyramid-Oriented Encoder-Decoder Cascade Convolution Neural Network for Crop Disease Leaf Segmentation," IEEE Access, vol. 9, pp. 14849-14866, 2021, doi: 10.1109/ACCESS.2021.3052769.
4. Yanmaz E. Joint or decoupled optimization: Multi-UAV path planning for search and rescue[J]. Ad Hoc Networks, 2023, 138: 103018.
5. Teng, L., Qiao, Y. BiSeNet-oriented context attention model for image semantic segmentation. Computer Science and Information Systems, vol. 19, no. 3, pp. 1409-1426. (2022), https://doi.org/10.2298/CSIS220321040T.
6. Uijlings J R R, Van De Sande K E A, Gevers T, et al. Selective search for object recognition[J]. International journal of computer vision, 2013, 104: 154-171.
7. Tang X, Xie X, Hao K, et al. A line-segment-based non-maximum suppression method for accurate object detection[J]. Knowledge-Based Systems, 2022, 251: 108885.
8. Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.
9. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
10. Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C] Computer Vision CECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

11. Karim, Shahid, Geng Tong, Jinyang Li, Akeel Qadir, Umar Farooq, and Yiting Yu. "Current Advances and Future Perspectives of Image Fusion: A Comprehensive Review." Information Fusion, Vol. 90, pp.185-217, February 2023.
12. S. Yin, L. Wang, M. Shafiq, L. Teng, A. A. Laghari and M. F. Khan, "G2Grad-CAMRL: An Object Detection and Interpretation Model Based on Gradient-weighted Class Activation Mapping and Reinforcement Learning in Remote Sensing Images," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023. doi: 10.1109/JSTARS.2023.3241405.
13. Tang C, Wu J, Hou Y, et al. A spectral and spatial approach of coarse-to-fine blurred image region detection[J]. IEEE Signal Processing Letters, 2016, 23(11): 1652-1656.
14. Wu D, Cao L, Zhou P, et al. Infrared Small-Target Detection Based on Radiation Characteristics with a Multimodal Feature Fusion Network[J]. Remote Sensing, 2022, 14(15): 3570.
15. Li Z, Wang H, Zhong H, et al. Self-attention module and FPN-based remote sensing image target detection[J]. Arabian Journal of Geosciences, 2021, 14: 1-18.
16. Cui Z, Leng J, Liu Y, et al. SKNet: Detecting rotated ships as keypoints in optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(10): 8826-8840.
17. Pazhani A A J, Vasanthanayaki C. Object detection in satellite images by faster R-CNN incorporated with enhanced ROI pooling (FrRNet-ERoI) framework[J]. Earth Science Informatics, 2022, 15(1): 553-561.
18. Anuar M M, Halin A A, Perumal T, et al. Aerial imagery paddy seedlings inspection using deep learning[J]. Remote Sensing, 2022, 14(2): 274.
19. Gao, Y., Wu, H., Wu, X., Li, Z., Zhao, X.: Human Action Recognition Based on Skeleton Features. Computer Science and Information Systems, Vol. 14, No. 3, 537-550. (2017), https://doi.org/10.2298/CSIS220131067G
20. Chen L, Yang Y, Wang Z, et al. Underwater Target Detection Lightweight Algorithm Based on Multi-Scale Feature Fusion[J]. Journal of Marine Science and Engineering, 2023, 11(2): 320.
21. Zheng C, Wang L. Semantic segmentation of remote sensing imagery using object-based Markov random field model with regional penalties[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 8(5): 1924-1935.
22. Ma F, Gao F, Wang J, et al. A novel biologically-inspired target detection method based on saliency analysis for synthetic aperture radar (SAR) imagery[J]. Neurocomputing, 2020, 402: 66-79.
23. Wang Z, Xu N, Wang B, et al. Urban building extraction from high-resolution remote sensing imagery based on multi-scale recurrent conditional generative adversarial network[J]. GIScience & Remote Sensing, 2022, 59(1): 861-884.
24. Ni K, Zhao Y, Wu Y. SAR Image Segmentation Based on Super-Pixel and Kernel-Improved CV Model[C]//IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2022: 3476-3479.
25. Zhu Z, Yin H, Chai Y, et al. A novel multi-modality image fusion method based on image decomposition and sparse representation[J]. Information Sciences, 2018, 432: 516-529.
26. Liguo Wang, Yin Shoulin, Hashem Alyami, et al. "A novel deep learning-based single shot multibox detector model for object detection in optical remote sensing images," *Geoscience Data Journal,* (2022). https://doi.org/10.1002/gdj3.162
27. Karnyoto, A. S., Sun, C., Liu, B., Wang, X.: Transfer Learning and GRU-CRF Augmentation for Covid-19 Fake News Detection. Computer Science and Information Systems, Vol. 19, No. 2, 639-658. (2022), https://doi.org/10.2298/CSIS210501053K
28. Farhadi A, Redmon J. Yolov3: An incremental improvement[C]//Computer vision and pattern recognition. Berlin/Heidelberg, Germany: Springer, 2018, 1804: 1-6.
29. Luo S, Yu J, Xi Y, et al. Aircraft target detection in remote sensing images based on improved YOLOv5[J]. Ieee Access, 2022, 10: 5184-5192.
30. Zhang K, Shen H. Multi-stage feature enhancement pyramid network for detecting objects in optical remote sensing images[J]. Remote Sensing, 2022, 14(3): 579.

31. Ma J, Shao W, Ye H, et al. Arbitrary-oriented scene text detection via rotation proposals[J]. IEEE transactions on multimedia, 2018, 20(11): 3111-3122.
32. Wang K, Du S, Liu C, et al. Interior attention-aware network for infrared small target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-13.
33. Singhal S, Passricha V, Sharma P, et al. Multi-level region-of-interest CNNs for end to end speech recognition[J]. Journal of Ambient Intelligence and Humanized Computing, 2019, 10: 4615-4624.
34. Theckedath D, Sedamkar R R. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks[J]. SN Computer Science, 2020, 1: 1-7.
35. Sun M, Zhao H, Li J. Road crack detection network under noise based on feature pyramid structure with feature enhancement (road crack detection under noise)[J]. IET Image Processing, 2022, 16(3): 809-822.
36. Lin T Y, Doll P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.
37. Addis A A, Tian W, Acheampong K N, et al. Small-Scale and Occluded Pedestrian Detection Using Multi Mapping Feature Extraction Function and Modified Soft-NMS[J]. Computational Intelligence and Neuroscience: CIN, 2022, 2022.
38. Hou Y, Yang Q, Li L, et al. Detection and Recognition Algorithm of Arbitrary-Oriented Oil Replenishment Target in Remote Sensing Image[J]. Sensors, 2023, 23(2): 767.
39. Lv N, Zhang Z, Li C, et al. A hybrid-attention semantic segmentation network for remote sensing interpretation in land-use surveillance[J]. International Journal of Machine Learning and Cybernetics, 2023, 14(2): 395-406.
40. Xiong W, Xiong Z, Cui Y. An explainable attention network for fine-grained ship classification using remote-sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-14.
41. Chen J, Li Z, Peng C, et al. UAV Image Stitching Based on Optimal Seam and Half-Projective Warp[J]. Remote Sensing, 2022, 14(5): 1068.

Shoulin Yin received the B.S. and M.S. degrees in image processing from Software College, Shenyang Normal University, Shenyang, China, in 2013 and 2015, respectively. He is currently a PHD with School of Information and Communication Engineering, Harbin Engineering University. He has published more than 20 technical articles in scientific journals and conference proceedings. His research interests include image fusion, object detection, and image recognition. Email:yslin@hit.edu.com.

Liguo Wang received his M.A. degree in 2002 and Ph.D. degree in signal and information processing in 2005 from Harbin Institute of Technology, Harbin, China. He held postdoctoral research position from 2006 to 2008 in the College of Information and Communications Engineering, Harbin Engineering University. He is currently a Professor with Information and Communications Engineering, Dalian Minzu University, Dalian, China. His research interests are remote sensing image processing and machine learning. He has published three books, 27 patents, and more than 200 papers in journals and conference proceedings. Email: wangliguo@hrbeu.edu.cn.

Qunming Wang received the Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong, in 2015. He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. From 2017 to 2018, he was a

Lecturer (Assistant Professor) with Lancaster Environment Centre, Lancaster University, Lancaster, U.K., where he is currently a Visiting Professor. His 3-year Ph.D. study was supported by the hypercompetitive Hong Kong Ph.D. Fellowship and his Ph.D. thesis was awarded as the Outstanding Thesis in the Faculty. He has authored or coauthored more than 70 peer-reviewed articles in international journals such as Remote Sensing of Environment, IEEE Transactions on Geoscience and Remote Sensing, and ISPRS Journal of Photogrammetry and Remote Sensing. His research interests include remote sensing, image processing, and geostatistics. Prof. Wang is an Editorial Board Member for Remote Sensing of Environment, and serves as an Associate Editor for Science of Remote Sensing (sister journal of Remote Sensing of Environment) and Photogrammetric Engineering & Remote Sensing. He was an Associate Editor for Computers and Geosciences (2017-2020).

Mirjana Ivanovic (Member, IEEE) has been a Full Professor with the Faculty of Sciences, University of Novi Sad, Serbia, since 2002. She has also been a member of the University Council for informatics for more than 10 years. She has authored or coauthored 13 textbooks, 13 edited proceedings, 3 monographs, and of more than 440 research articles on multi-agent systems, e-learning and web-based learning, applications of intelligent techniques (CBR, data and web mining), software engineering education, and most of which are published in international journals and proceedings of high-quality international conferences. She is/was a member of program committees of more than 200 international conferences and general chair and program committee chair of numerous international conferences. Also, she has been an invited speaker at several international conferences and a visiting lecturer in Australia, Thailand, and China. As a leader and researcher, she has participated in numerous international projects. She is currently an Editor-in-Chief of Computer Science and Information Systems Journal.

Jinghui Yang received the B.S., M.S., and Ph.D. degrees from Harbin Engineering University, Harbin, China, in 2010, 2013, and 2016, respectively. She is currently an associate professor with the School of Information Engineering, China University of Geosciences, Beijing, China. Her research interests include hyperspectral imagery, remote sensing, pattern recognition, and signal processing.

Received: March 20, 2023; Accepted: June 22, 2023.