BIDIRECTIONAL PATHWAY FEATURE PYRAMID NETWORKS AND REVERSE SCALE-TRANSFER LAYER FOR DETECTING MULT-SCALE SHIPS

Guanhua Jiang[1], Yanan You[1], Gang Meng[2], Bohao Ran[1], Fang Liu[1]

[1]School of Artificial Intelligence, Beijing University of Posts and Telecommunications
[2]Beijing Institute of Remote Sensing Information ,100192,Beijing,China

IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium | 978-1-6654-0369-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/IGARSS47720.2021.9553241

ABSTRACT

Multi-scale ship detection for remote sensing images is always a popular research field in civil and military application. In this paper, in order to solve the feature information single pathway flow in feature layer causes the lack of detailed information in the deep feature layer, we propose a bidirectional pathway feature pyramid networks (BP-FPN) method, which enables the deep feature layer to have strong semantic information as well as rich detailed information. At the same time, the Reverse Scale-transfer Layer down-sampling method is proposed to reduce the information loss of feature layer in the process of down-sampling. It ensures that the feature layer maintain information during the down-sampling process. Experimental results on a dataset collected from Google Earth have quantitatively and qualitatively demonstrated the effectiveness of our approach

Index Terms— remote sensing, high-level semantic, multi-scale ship detection, convolution neural network

1. INTRODUCTION

Ship detection has always been a very important issue in both civilian and military fields, and multi-scale object detectionis a challenge in high-resolution optical remote sensing images. In fact, variety ships usually berthed around the port, but they have different shapes and scales, so it is difficult to accurately detect them all, which is an urgent concern in the current ship detection task.

With the development of deep learning, various object detection network models proposed at present have achieved excellent performance in object detection tasks. At the same time, targeted network structures such as R2CNN [1] and OPD-Net [2] are proposed for ship detection based on remote sensing images, which have a good performance for ship detection. At present, the main method to solve the multi-scale problem of ships is to fuse the feature of the each network layer [3,4,5,6]. The most critical module is feature pyramid networks (FPN) [3]. FPN uses different feature layers to focus on different object scales, making the network enable to detect multi-scale objects.

Due to less sampling times and more detailed information, the low-level feature layers tend to be more localized [7], which is suitable for detecting small objects. On the contrary, the deep feature layer has stronger semantic information [8,9] because of high sampling times and is suitable for detecting large objects. FPN uses a single pathway from top to down, which makes up for the shortcomings of low-level feature layer, more detailed information but insufficient semantic information, and enhances the detection of small and medium objects in the network. However, such a network design would make the feature extracted from the high-level feature layer relatively simple, with strong semantic information but missing the original detail information [10]. As a result, it has a low reliability of the network in detecting large objects, with the high false alarm rate, and the prediction box is inaccurate, as shown in Fig 1. Therefore, in order to solve the multi-scale object detection problem, we proposed bidirectional pathway feature pyramid networks (BP-FPN) . It retains the FPN from top to down pathway while adding the pathway from down to top. So that the deep feature layer has not only strong semantic information but also abundant details information from low-level features. In order to maintain more details information of the low-level feature layer in the process of subsampling, we also proposed a Reverse Scale-transfer Layer modified based on the Scale-transfer Layer [11,12] In this work, R2CNN is seletected as the basic

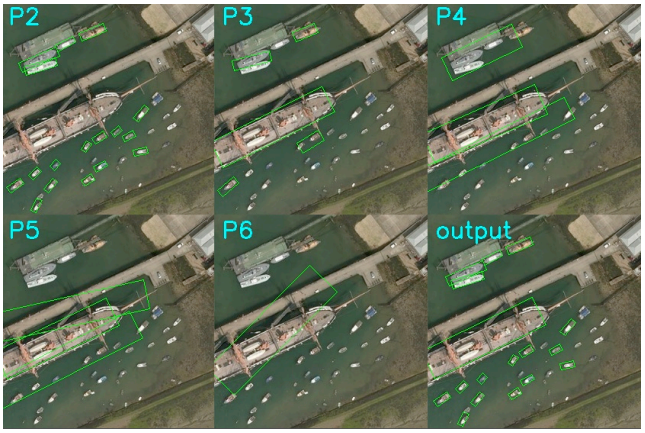

Fig. 1.The first five figures are respectively the predicted output of P2, P3, P4, P5 and P6 of each RPN layer, and the last figure is the predicted output after the fusion of all layers (P5 and P6 have high false alarm and inaccurate prediction box).

framework to evaluate our method on a dataset collected from Google Earth.

2. METHODS

In this section, we will first introduce the Reverse Scale-transfer layer, which serves as the basic module for the down-sampling in the BP-FPN used later. Then the structure of BP-FPN will be introduced.

2.1 Reverse Scale-transfer Layer

The commonly used downsampling methods in deep learning, such as interpolation and pooling, tend to reduce the amount of information. So that the feature layer loses certain details and semantic information, which is not conducive to dealing with position sensitive tasks. Inspired by the Scale-transfer Layer used in [4,5], we propose a Reverse Scale-transfer Layer, so that the feature layer does not lose its information during the downsampling process, and the detail information and semantic information of the feature layer are better retained.

The Reverse Scale-transfer Layer takes the tensor with the shape of H×W×C as the input, takes [m, n] as the kernel size and S as the stride to slide on the feature map. For each sub-tensor with the shape of m×n×C, cut it into m×n tensors with the shape of 1×1×C and concatenate. Then transform the sub-tensor into the shape of $1\times1\times C\cdot m\cdot n$. Finally, concatenate each sub-tensor. When m=n=2 and S=2, the

realization process of Reverse Scale-transfer Layer is shown in Fig. 2. Its mathematical formula is shown as follows:

$$H' = floor(\frac{H-m}{s})+1$$
$$W' = floor(\frac{W-n}{s})+1 \qquad (1)$$
$$RST(T_{H\times W\times C}) = T'_{H'\times W'\times C\cdot m\cdot n}$$

where RST is the Reverse Scale-transfer Layer, T is the input tensor and T' is the output tensor.

By hiding feature layer information into the channel, the Reverse Scale-transfer Layer not only ensures the downsampling of feature map, but also ensures that the information of feature Layer will not be lost in the downsampling process. More detail information and semantic information will be saved, so that the feature map can obtain more comprehensive information and make prediction more accurate.

2.2 BP-FPN

The top-down pathway of FPN well solves the shortcoming of the lack of semantic information in the lower feature layer. Making the lower feature layer have strong semantic information while giving consideration to rich detail information, which is suitable for detecting small object accurately. However, the disadvantage of this single

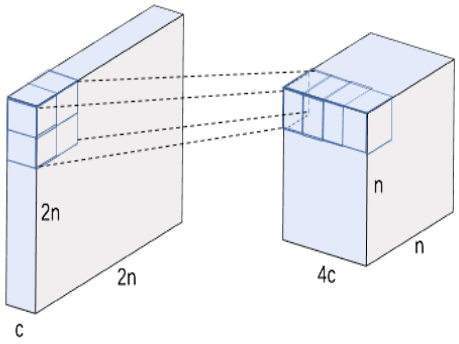

Fig. 2.Reverse Scale-transfer Layer, m=n=2, S=2

pathway is that it only endows the lower feature layer with strong semantic information, but it does not solve the problem that the upper feature layer lacks detailed information. Therefore, our BP-FPN is designed to overcome this shortcoming and make the feature layer information flow in two directions. From top to down, the upper feature layer endows the lower feature layer with strong semantic information. From down to top, the lower feature layer endows the upper feature layer more detail information. Through bidirectional pathway, relatively comprehensive feature information can be extracted from both the lower feature layer and the upper feature layer, which is suitable for object position sensitive tasks and makes mult-scale object detection more accurate.

The network structure of BP-FPN is designed as shown in Fig. 3, with ResNet101 as the backbone to extract the feature layer as {C2, C3, C4, C5}. From the top-down pathway, the upper feature layer is endowed with richer semantic information by up-sampling to the lower feature layer, generating feature layer $\{P2_d, P3_d, P4_d, P5_d, P6_d\}$, where $P6_d$ is obtained by down-sampling $P5_d$, and the up-sampling method is replaced by Scale-transfer Layer. On the contrary, the channel from down to top endows the upper feature layer with richer details information through down-sampling, generating the feature Layer $\{P2_t, P3_t, P4_t, P5_t, P6_t\}$, $P6_t$ is obtained by $P5_t$ through down-sampling, which is replaced by Reverse Scale-transfer Layer. So that the feature layer does not lose too much detail information in the process of subsampling. The top-down pathway formula is as follows:

$$P_i = \begin{cases} Conv_{3\times3}(Conv_{1\times1}(C_i) + \sum_{j=i+1}^{5} ST_{2(j-i)}(P_j)) & i = 2,3,4 \\ Conv_{1\times1}(C_i) & i = 5 \\ Conv_{3\times3}(RST_{r=2}(P_{i-1})) & i = 6 \end{cases} \qquad (2)$$

from down-top pathway formula is as follows：

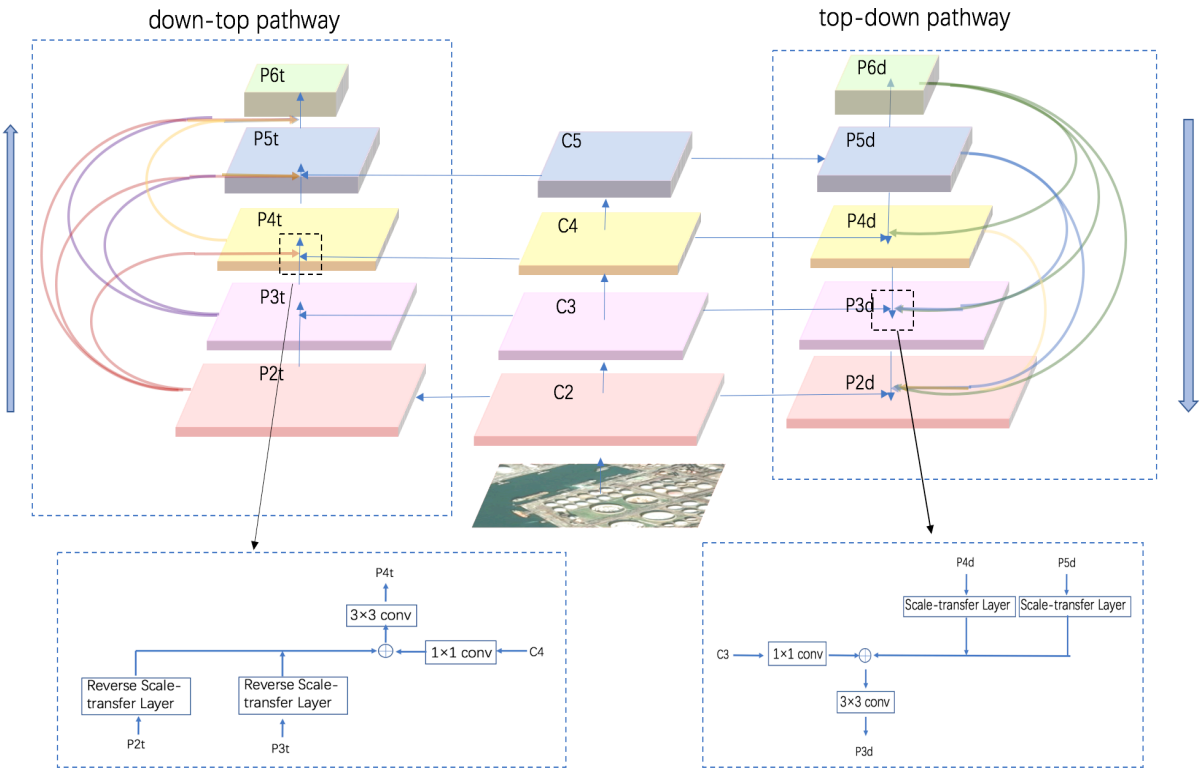

Fig. 3 The structure of BP-FPN with Reverse Scale-transfer Layer and Scale-transfer Layer.

$$P_i = \begin{cases} Conv_{1\times1}(C_i) & i=2 \\ Conv_{3\times3}(Conv_{1\times1}(C_i) + \sum_{j=2}^{i-1} RST_{r=2(j-i)}(P_j)) & i=3,4,5 \\ Conv_{3\times3}(RST_{r=2}(P_{i-1})) & i=6 \end{cases} \quad (3)$$

where ST is the Scale-transfer Layer with up sampling factor r and RST is the Reverse Scale-transfer Layer with down sampling factor r.

Concatenate the two pathway outputs, and the feature layer {P2, P3, P4, P5, P6} is finally output.

3. EXPERIMENTS

This section will introduce some implementation details during the experiment, and then verify the contribution of each module used to the detection accuracy rate.

3.1 Implementation Details

We build an optical remote sensing dataset collected from Google Earth, which contains 2728 images of 12227 ships. The dataset covers ships of different length ranging from 20 to 450 pixels. Such a dataset can better verify the effect of our proposed method. In addition, in order to make the scale distribution more balanced, we added scale balance on the basis of the rotation and resize data enhancement, this makes the network easier to be trained. The size of each image is 768 by 768.

The basic network structure used in the experiment is R2CNN, and it is suitable for determining the direction of the target in the image. The pre-training parameter of the network is generated based on ImageNet. The BP-FPN structure is imported into the region proposal network (RPN) of R2CNN. The base anchor size set for each layer is different, {P2, P3, P4s, P5, P6} corresponding to the base anchor size is {20, 50, 100, 150, 200}, and the scale of anchor is [0.5, 1.0, 2.0, 1/4, 4.0, 1/6, 6.0].

The optimizer used in the training process is the stochastic gradient descent optimizer (SGD). The learning rate is 0.001 with a learning rate decay ratio 0.1. The GPU used in the experiment was NVIDIA Tesla T4.

3.2 Experiment Results

The method we proposed mainly includes two function modules: BP-FPN and Reverse Scale-transfer Layer in RPN stage. Via the above modules, the extracted features possess both richer detailed information and stronger semantic information. To test for these improvements, a series of tests are ultilized to verify its correctness and validity.

Table 1. Comparison of the performance of each method.

Methods	MAP	MAP (L)
R2CNN	79.43%	81.91%
R2CNN + FPN	86.09%	90.423%
R2CNN + BP-FPN	88.981%	92.697%
R2CNN + BP-FPN + Reverse Scale-transfer Layer	89.729%	93.744%

Meanwhile, in order to more reasonably evaluate the improvement of our method, mean average precision (MAP) is used as the evaluation index. At the same time, using our method is more about enriching the information of the deep feature layer, thus making the detection of large objects more accurate. Therefore, MAP is also calculated separately for objects with a length greater than 150 pixels. As shown in Table 1, the experimental results show that the two methods proposed by us (BP-FPN and Reverse Scale-transfer Layer) can improve the detection of multi-scale ships to a certain extent. Especially for large scale ship objects, the enhancement is more obvious. This is because our proposed method enables the deep feature layer to have richer information and is more conducive to the location-sensitive task. Some detection results as show in Fig s4.

4. CONCLUSION

In this paper, BP-FPN and Reverse Scale-transfer Layer are proposed to solve the multi-scale problem in ship object detection. Through the bidirectional pathway flow mechanism and the retention of more feature information, each feature layer has not only stronger semantic information but also richer detailed information. In the experiments, we used mult-scale dataset collected from Google earth, evaluated the improvement of each method by ablation experiments. The results prove the feasibility and accuracy of the method.

5. ACKNOWLEDGEMENT

This work is Supported by Beijing Natural Science Foundation, China (Grant No. 4214058).

6. REFERENCS

[1] Jiang Y, Zhu X, Wang X, et al. R2cnn: rotational region cnn for orientation robust scene text detection[J]. arXiv preprint arXiv:1706.09579, 2017.

[2] You Y, Ran B, Meng G, et al. OPD-Net: Prow Detection Based on Feature Enhancement and Improved Regression Model in Optical Remote Sensing Imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020.

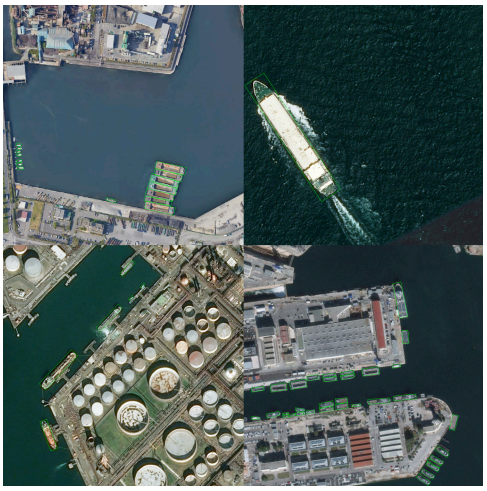

Fig. 4 Some detection results.

[3] Lin T Y, Dollár 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.

[4] Singh B, Davis L S. An analysis of scale invariance in object detection snip[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 3578-3587.

[5] Liu S, Huang D. Receptive field block net for accurate and fast object detection[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 385-400.

[6] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[J]. arXiv preprint arXiv:1511.07122, 2015.

[7] Luo W, Li Y, Urtasun R, et al. Understanding the effective receptive field in deep convolutional neural networks[J]. arXiv preprint arXiv:1701.04128, 2017.

[8] Zhang S, Zhu X, Lei Z, et al. S3fd: Single shot scale-invariant face detector[C]//Proceedings of the IEEE international conference on computer vision. 2017: 192-201.

[9] Wang J, Sun K, Cheng T, et al. Deep high-resolution representation learning for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2020.

[10] Yang X, Sun H, Fu K, et al. Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks[J]. Remote Sensing, 2018, 10(1): 132.

[11] Zhou P, Ni B, Geng C, et al. Scale-transferrable object detection[C]//proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 528-537.

[12] Liu N, Cui Z, Cao Z, et al. Scale-Transferrable Pyramid Network for Multi-Scale Ship Detection in Sar Images[C]//IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019: 1-4.