# Inception Parallel Attention Network for Small Object Detection in Remote Sensing Images

Shuojin Yang[1,3,4], Liang Tian[1,3,4], Bingyin Zhou[2,3], Dong Chen[1,3,4], Dan Zhang[1,3], Zhuangnan Xu[1,3], Wei Guo[2,3], and Jing Liu[1,3,4,5(✉)]

[1] College of Computer and Cyber Security, Hebei Normal University, Shijiazhuang 050024, Hebei, China
liujing01@ict.ac.cn

[2] School of Mathematical Sciences, Hebei Normal University, Shijiazhuang 050024, Hebei, China
guowei@chmiot.net

[3] The Key Laboratory of Augmented Reality, School of Mathematical Sciences, Hebei Normal University, Shijiazhuang 050024, Hebei, China

[4] Hebei Provincial Engineering Research Center for Supply Chain Big Data Analytics and Data Security, Hebei Normal University, Shijiazhuang 050024, Hebei, China

[5] The Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

**Abstract.** Small object detection in remote sensing images is a major challenge in the field of computer vision. Most previous methods detect small objects using a multiscale feature fusion approach with the same weights. However, experiments shows that the inter feature maps and the feature map in different scales have different contribution to the network. To further strengthen the effective weights, we proposed an inception parallel attention network (IPAN) that contains three main parallel modules i.e. a multiscale attention module, a contextual attention module, and a channel attention module to perform small object detection in remote sensing images. In addition, The network can extract not only rich multiscale, contextual features and the interdependencies of global features in different channels but also the long-range dependencies of the object to another based on the attention mechanism, which contributes to precise results of small object detections. Experimental results shows that the proposed algorithm significantly improves the detection accuracy especially in complex scenes and/or in the presence of occlusion.

**Keywords:** Parallel · Small objects · Attention mechanism · Complex scene · Occlusion

The research was jointly supported by the National Natural Science Foundation of China (Grant No.: 61802109, 61902109), the Science and Technology Foundation of Hebei Province Higher Education (Grant No.: QN2019166), the Natural Science Foundation of Hebei province (Grant No.: F2017205066, F2019205070), the Science Foundation of Hebei Normal University (Grant No.: L2017B06, L2018K02, L2019K01).


Y. Peng et al. (Eds.): PRCV 2020, LNCS 12305, pp. 469–480, 2020.
https://doi.org/10.1007/978-3-030-60633-6_39

## 1 Introduction

With the development of earth observation technology and the improvement in the performance of computation processing of large-scale data, object detection in remote sensing images has attracted increasing attention [2,7,18,19,21]. However, for remote sensing images with complex backgrounds, small size and in the presence of occlusion, it is still a significant challenge. Deep learning algorithms have demonstrated good performance in object detection in recent years, including region based methods and single shot methods. e.g., faster region-based CNN (faster-RCNN), spatial pyramid pooling networks (SPP-Net) [5], you only look once (YOLO) [15] and the single-shot multibox detector (SSD) [10]. However, these innovations usually fail to detect very small objects, as small object features are always lost in the downsampling process of deep CNNs (DCNNs). The above DCNN methods cannot effectively extract the features of small objects. especially from remote sensing images in which objects usually small and blurry, which has created considerable challenges.

To increase the accuracy of the networks in detecting small objects in remote sensing images, one method is to employ features of the middle layers of the CNN and then exploit multiscale features with multilevel information. Ding et al. [2] directly concatenated multiscale features of a CNN to obtain fine-grained details for the detection of small objects. Lin et al. adopted pixelwise summation to incorporate the score maps generated by multilevel contextual features of different residual blocks for the segmentation of small objects such as a vehicle. Mou et al. [13] proposed a top-down pathway and lateral connection to build a feature pyramid network with strong semantic feature maps at all scales. The method assigned the feature maps of different layers to be responsible for objects at different scales. Yang et al. [21] introduced the dense feature pyramid network (DFPN) for automatic ship detection; each feature map was densely connected and merged by concatenation. Furthermore, Li et al. [7] proposed a hierarchical selective filtering layer that mapped features of multiple scales to the same scale space for ship detection at various scales. Gao et al. [4] designed a tailored pooling pyramid module (TPPM) to take advantage of the contextual information of different subregions at different scales. Qiu et al. proposed A2RMNet [14] to improve the problem of wrong positioning caused by excessive aspect ratio difference of objects in remote sensing images. Xie et al. proposed a fully connect network model NEOONet [20], which solved the problem of category imbalance in remote sensing images.

Although the above methods have achieved promising detection results by aggregating multiscale features, there are still some problems. When the objects are in the complex scene or partly exposed the previous methods can not extract discriminative features for detection.

To address these issues, First, we proposed a multiscale attention module to guide the network to learn more useful information in low-level feature maps with long-range dependencies, which is important for the detection of small objects. Then we presented a contextual attention module to extract interdependencies of foreground and background in feature maps, So the module can learn more of the

correlated background and foreground information with long-range dependencies which can efficiently detect objects in complex scenes and in the presence of occlusion. In addition, we designed a channel attention module to model the importance of each feature map, which suppresses the background or unrelated category features with long-range dependencies, and improves the accuracy of the model.

## 2 Proposed Work

Previous methods cannot effectively extract the features of small objects because of the contradiction between resolution and high-level semantic information; i.e., low-level feature maps have high resolution but little semantic information, and high-level feature maps have low resolution but rich semantic information, which are all useful for accurately detecting objects. In addition, when an object is in a complex scene and in the presence of occlusion, the accuracy of the previous algorithms will decrease. Therefore, motivated by DANet [3], the proposed method uses a parallel self-attention module to enhance the accuracy of small object detection in remote sensing images.

As illustrated in Fig. 1, three types of attention modules are designed, and they contain a series of matrix operations to synthetically consider multiscale features, local contextual features and global features in the residual network after every residual block (resblock). In addition, in order to extract the feature of direction sensitive objects we use Deformable convolution and upsampling bottleneck operations to add deep feature maps to the shallow feature maps and we use four-scale feature maps to predict which are advantageous for detecting small objects. Before prediction, 3 3 convolutional layers are used to prevent aliasing effects.

### 2.1 Multiscale Attention Module

Feature maps of different scales have different semantic and spatial information; in high-level feature maps, semantic information are rich, and in the low-level feature maps, spatial information are rich, however, both types of information are useful for detecting small objects in remote sensing images. To strengthen the small object feature representation, a multiscale attention module is proposed to combine deep and shallow features. The structure of the multiscale attention module is illustrated in Fig. 2.

The feature maps of $A$ and $B$ are used obtained by resblock1 and resblock2 to calculate the attention map. $H$, $W$, and $C$ represent the height, width and the number of channels of the feature maps respectively. $\mathbf{B_1} \in \mathbb{R}^{C\times H\times W}$, $\mathbf{B_2} \in \mathbb{R}^{2C\times H/2\times W/2}$, $1 \times 1$ convolution is used to set $B_2$ to $\mathbf{B} \in \mathbb{R}^{C\times H/2\times W/2}$. We reshape $A$ to $\mathbb{R}^{C\times N}$ and $B$ to $\mathbb{R}^{C\times N/4}$, and $N = H \times W$ is the number of pixels. Then, a matrix multiplication of the transpose of $A$ and $B$ is performed and a softmax layer is applied to normalize the weights and calculate the multiscale

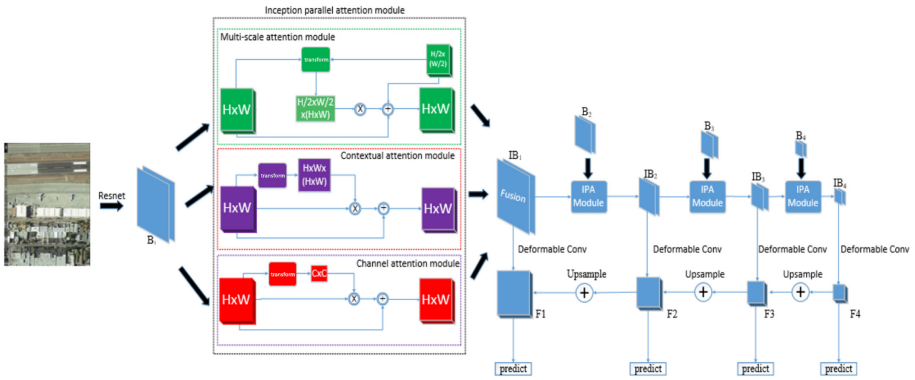


**Fig. 1.** The architecture of the IPAN. The proposed method use three parallel modules to obtain rich multiscale, contextual features and global features in different channels. Next, an element-wise summation operation is performed on the feature maps of the above three modules to obtain the final representations reflecting the long-range contexts. Finally, four scales with a multiscale feature fusion operation are used for prediction.

attention map $\mathbf{M} \in \mathbb{R}^{N \times H/4}$:

$$M_{ji} = \frac{\exp\left(A_i \cdot B_j\right)}{\sum_{i=1}^{N} \exp\left(A_i \cdot B_j\right)} \tag{1}$$

where $M_{ji}$ measures the $i$ th position's impact on the $j$ th position at different scales. Then, we perform a matrix multiplication of $B$ and the transpose of $M$ and reshape the result to $R^{C \times H \times W}$. Finally, the result is multiplied by a scale parameter $\alpha$ and perform an elementwise summation is performed with $A$ to obtain the final output $\mathbf{E} \in \mathbb{R}^{C \times H \times W}$ as follows:

$$E_j = \alpha \sum_{i=1}^{N} \left(M_{ji} B_i\right) + B_{1j} \tag{2}$$

The feature map $B$ is deeper than $A$ in resnet, which has more semantic information. Therefore, the attention map can guide $B$ to learn more low-level information. As a result, $E$ combines the low-level position information and high-level semantic information, which is effective for small object detection. From the heat map in $E$, we can see that more small object features are activated.

### 2.2 Contextual Attention Module

A discrimination of foreground and background features is essential for object detection and can be achieved by capturing contextual information. Some studies [6,17] have used contextual information to improve the detection result. To

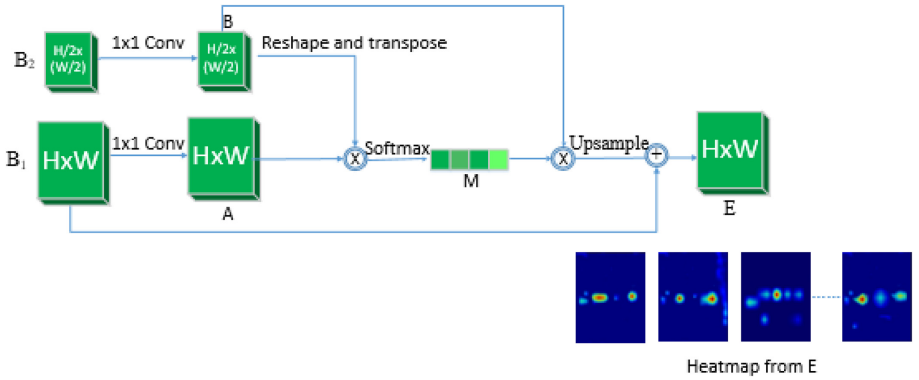


**Fig. 2.** The architecture of multiscale attention module. "×" represents matrix multiplication.

model rich contextual relationships over local features, we introduce a contextual attention module. The contextual attention module encodes a wide range of contextual information into the local features, thus enhancing the representation capability. Next, we elaborate the process to adaptively aggregate spatial contexts.

As illustrated in Fig. 3, given a local feature $\mathbf{B_1} \in \mathbb{R}^{C\times H\times W}$, the local feature is first fed into convolution layers to generate two new feature maps $C$ and $D$ the size of the filter is $7 \times 7$ (for example, $B_1$). With the deepening of the network, the filter sizes are $7 \times 7$, $5 \times 5$, and $3 \times 3$ at different scales, and large-scale filters can extract more contextual features, where $\mathbf{C}, \mathbf{D} \in \mathbb{R}^{C\times H\times W}$. Then, reshape them $\mathbb{R}^{C\times N}$, where $N = H \times W$ is the number of pixels. Then, a matrix multiplication of the transpose of $C$ and $B$ is performed, and a softmax layer is applied to calculate the contextual attention map $\mathbf{P} \in \mathbb{R}^{N\times N}$:

$$P_{ji} = \frac{\exp\left(C_i \cdot D_j\right)}{\sum_{i=1}^{N} \exp\left(C_i \cdot D_j\right)} \tag{3}$$

$$F_j = \beta \sum_{i=1}^{N} \left(P_{ji} B_{1i}\right) + B_{1j} \tag{4}$$

where $M_{ji}$ measures the $i$ th position's impact on the $j$th position with contextual information. The more similar feature representations of the two positions contribute to a greater correlation with contextual information between them. Then, we perform a matrix multiplication of $B_1$ and the transpose of $M$ and reshape the result to $\mathbb{R}^{C\times H\times W}$. Finally, we multiply the result by a scale parameter $\beta$ and perform an elementwise summation operation with the features $B_1$ to obtain the final output $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$.

The attention map from $B$ and $C$ contains contextual information with long-range dependencies. From the heat map shown in $F$, it can be observed that there

are many features around the objects, which contributes to the detection of the small objects is activated.

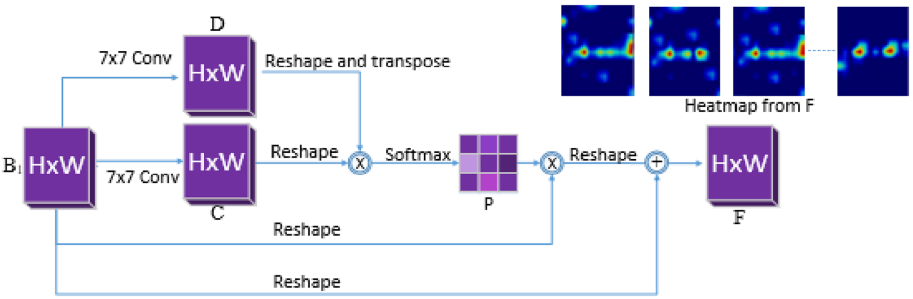


**Fig. 3.** The architecture of contextual attention module. "×" represents matrix multiplication.

### 2.3 Channel Attention Module

Each channel map has a different class and spatial position response, some contribute to object detection but others do not. To strengthen the positive response, weaken the negative response and exploit the interdependencies between channel maps, it can be emphasized that interdependent feature maps can improve the feature representation of the class and specific position. Therefore, we build a channel attention module to explicitly model the interdependencies between the channels and learn the long-range dependencies in the feature maps.

The structure of the channel attention module is illustrated in Fig. 4. Different from the contextual attention module, we directly calculate the channel attention map $\mathbf{M} \in \mathbb{R}^{C\times C}$ from the original features $\mathbf{B_1} \in \mathbb{R}^{C\times H\times W}$. Specifically, $B_1$ is reshaped to $\mathbb{R}^{C\times N}$, and then a matrix multiplication of original $B_1$ and the transpose of $B_1$ is performed. Finally, a softmax layer is applied to obtain the channel attention map $\mathbf{Q} \in \mathbb{R}^{C\times C}$:

$$Q_{ji} = \frac{\exp\left(B_{1i} \cdot B_{1j}\right)}{\sum_{i=1}^{C} \exp\left(B_{1i} \cdot B_{1j}\right)} \tag{5}$$

$$G_j = \gamma \sum_{i=1}^{N} \left(Q_{ji} B_{1i}\right) + B_{1j} \tag{6}$$

where $M_{ji}$ measures the $i$ th channel's impact on the $j$ th channel. In addition, we perform a matrix multiplication of the transpose of $M$ and $B_1$ and reshape the result to $\mathbb{R}^{C\times H\times W}$. Then, we multiply the result by a scale parameter $\gamma$ and perform an elementwise summation operation with $B_1$ to obtain the final output $\mathbf{G} \in \mathbb{R}^{C\times H\times W}$. From the heat map shown in $G$ from Fig. 4, we can see that more features strongly associated with the objects are activated.

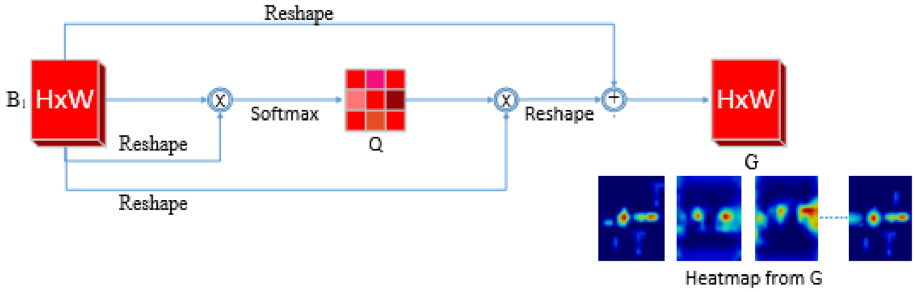


**Fig. 4.** The architecture of channel attention module. "×" represents matrix multiplication.

### 2.4 Embedding with Networks

As shown in Fig. 1, we set our parallel attention module after every resblock. After every IPA module, the feature maps have multiscale contextual and global information with long-range dependencies. Positive feature response achieves mutual gains, thus increasing differences in the objects and backgrounds. Then, every scale feature map is fed to RPN [16] for a further prediction. In addition, the model can exploit spatial information at all corresponding positions to model the channel correlations. Notably, the proposed attention modules are simple and can be directly inserted into the existing object detection pipeline. The modules do not require too many parameters yet effectively strengthen the feature representations.

## 3 Experiment

We performed a series of experiments with RSOD [11] including objects that are small in a complex scene and in the presence of occlusion. The proposed method achieved state-of-the-art performance: 95.9% for aircrafts and 95.5% for cars. We first conducted evaluations of the remote sensing object detection dataset to compare the performance to that of other state-of-the-art object detection methods, and evaluated the robustness of the proposed method using the COCO [9] dataset.

### 3.1 Comparison with State-of-the-Arts Methods in Remote Sensing Object Detection Dataset

We chose the ResNet-101 model as the backbone network. The model was initialized by the ImageNet classification model and fine-tuned with the RSOD [11]. We randomly split the samples into 80% for training and 20% for testing. In all experiments, we trained and tested the proposed method based on the TensorFlow deep learning framework. First, we resized the images to 800 × 800 pixels and applied stochastic gradient descent for 50k iterations to train our model.

The learning rate was 0.001 and it was decreased to 0.0001 after 30 k iterations. We adopted only one scale anchor for one scale to predict with three ratios, 1:1, 1:2 and 2:1, with areas of 32 × 32 pixels, 64 × 64 pixels, 128 × 128 pixels, and 256 × 256 pixels, which performed well with the dataset.

Figure 5 shows the detection results of comparing the proposed algorithm with popular deep learning methods. The red boxes in the first three rows are the results of aircraft (small object) detection in complex scene, in the presence of occlusion respectively. The last row shows the results for the cars, the blue box is the car in the complex scene and the red box is the car in the presence of occlusion.

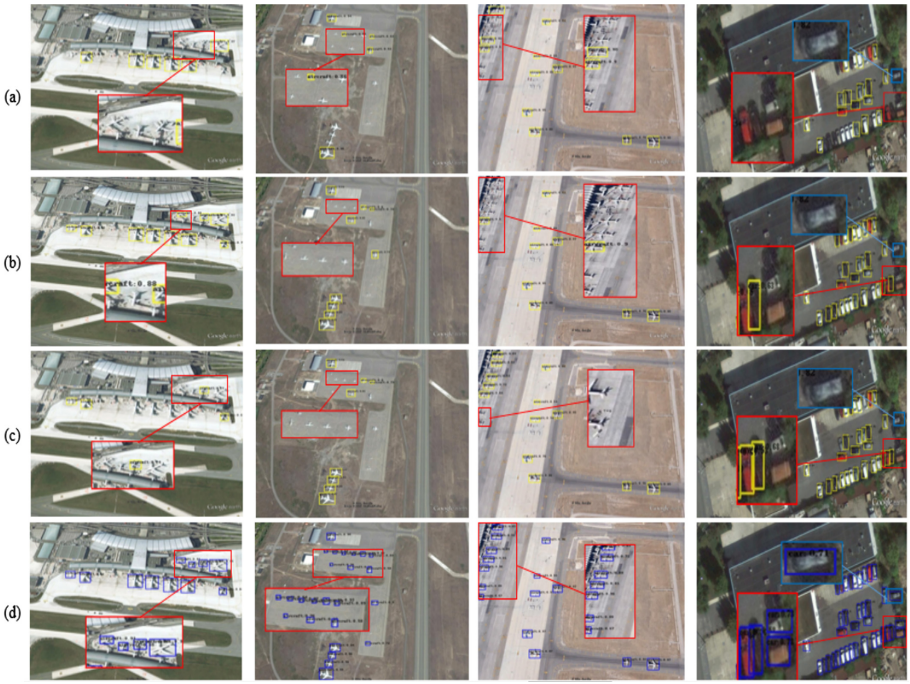


**Fig. 5.** Visualization of the small object detection results for a complex scene in the presence of occlusion for the proposed algorithm with popular deep learning methods on the aircraft and car datasets: (a) is the results of FPN [8], (b) is the results of Modified faster-RCNN [17], (c) is the results of R-FCN [1] and (d) the proposed method. (Color figure online)

Firstly, we compared the proposed approach with seven common state-of-the-art methods for object detection and five state-of-the-art methods for remote sensing object detection. As shown in Table 1, the deep learning method is obviously better in terms of accuracy than traditional methods such as the HOG [19] and SIFT [18]. Because of the superiority of the proposed method in small

object detection and the strong ability to handle complex scene with occlusion, the proposed method has higher accuracy and robustness when detecting aircrafts and cars. From the AP and recall (R) shown in Table 1, we can observe that in terms of AP over all the aircraft and the car categories, the proposed approach improve about 7% on average to other deep learning methods, and outperforms the previously most accurate USB-BBR [11] method and AARMN [14] method by 1% and 1%, respectively. The proposed method can obtain the highest AP on the premise of relatively high recall.

**Table 1.** Compares the proposed method to other methods in the remote sensing object detection datasets. The symbol "–" indicates that the method does not provide relevant results.

| Methods | Aircraft AP (%) | Aircraft R (%) | Car AP (%) | Car R (%) |
|---|---|---|---|---|
| YOLO [15] | 84.92 | 82.81 | 71.73 | 70.93 |
| faster-RCNN [16] | 80.15 | 78.91 | 82.52 | 81.21 |
| FPN [8] | 91.22 | 90.18 | 85.12 | 86.33 |
| R-FCN [1] | 84.3 | 95.26 | 89.3 | 88.2 |
| USB-BBR [11] | 94.69 | 93.09 | – | – |
| NMMDPN [12] | – | – | 91.36 | 90.56 |
| NEOON [20] | 94.49 | – | 93.22 | – |
| MFast-RCNN [17] | 84.5 | 85.1 | 80.1 | 80.5 |
| AARMN [14] | 94.27 | 94.18 | 94.65 | 92.16 |

### 3.2 Ablation Study

An ablation study is performed to validate the contribution of each module of the proposed method. In each experiment, one part of our method is omitted, and the remaining parts are retained. The APs are listed in Table 2. It can be seen that the AP increased by almost 1% by including the contextual attention module and the channel attention module, because the contextual attention module contains contextual information about the object. Second, we used the proposed channel attention module and the multiscale attention module after the resblock, and the result improved by 2.1%. Which indicates that global information and multiscale information are important for detection. Finally, we interposed the contextual attention module and the multiscale attention module into the network, and the AP improved by 2.3%. Thus, the contextual information and multiscale information are effective in helping the network detect small objects. It is obvious that when all modules are applied, the proposed method can improve the AP by 4.7%.

Figure 6 shows the visualized results of the ablation study for situations 1–5 in Table 3, corresponding to (b)–(f). It can be easily seen that due to the

use of the multiscale attention module, the proposed method obtains more spatial structural information and semantic information about small objects, so it can enhance the feature representation of small objects. The contextual attention module and the channel attention module enable the network to learn the correspondence between the background and foreground and the global interdependence in different channels and further strengthen the feature representation, so the network can efficiently detect the objects in a complex scenes and in the presence of occlusion.

**Table 2.** The result of the ablation study for the proposed framework in the aircraft detection task. N: No, Y: Yes.

| IPAN | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Multiscale attention | N | N | Y | Y | Y |
| Contextual attention | N | Y | N | Y | Y |
| Channel attention | N | Y | Y | N | Y |
| Aircraft AP(%) | 91.2 | 92.1 | 93.3 | 93.5 | 95.9 |

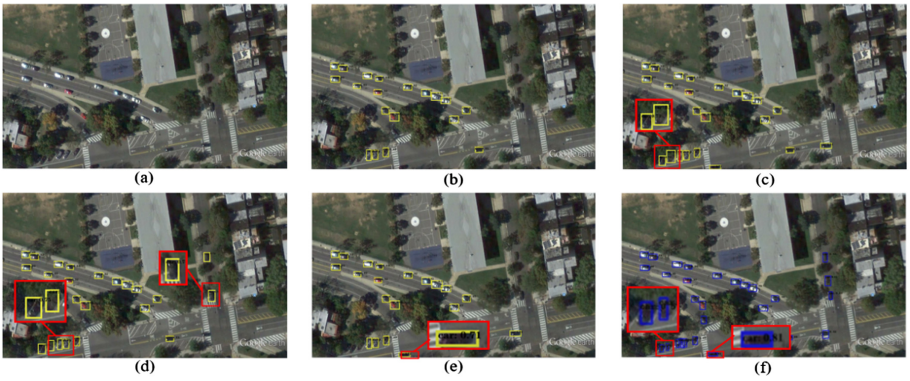


**Fig. 6.** The visualized results of the ablation study. (a) is the original image and (b)–(f) are situations from "1" to "5" in Table 2.

### 3.3 Robustness Experiments

In order to verify the robustness of the proposed method, we evaluated the performance with the COCO [9] dataset. As shown in Table 3, the dataset has more classes and considerably smaller objects so the proposed algorithm has better performance than other methods.

**Table 3.** Compares the proposed method in the COCO dataset to the state-of-the-art methods.

| Methods | COCO mAP(%) |
|---|---|
| R-FCN [1] | 29.9Y |
| FPN [8] | 36.2 |
| SSD [10] | 31.2 |
| YOLO [15] | 33 |
| Proposed method | 38.3 |

## 4 Conclusion

In this paper, an IPAN is presented for small object detection in remote sensing images, which enhances the feature representation and improves the detection accuracy for small objects, especially in complex scenes and in the presence of occlusion. Specifically, we introduced three parallel modules: a multiscale attention module guiding the model extract more information of small objects, a contextual attention module to capture contextual correlation, and a channel attention module to learn global interdependencies in different channels, In addition, the IPAN can capture long-range dependencies which helps to detect objects. The experiments and ablation study show that the network is effective and achieve precise detection results. The proposed IPAN consistently achieves outstanding performance with remote sensing object detection datasets. The future work will focus on further enhancing the robustness of our model.

## References

1. Dai, J., Li, Y., He, K., Sun, J.: R-FCN: object detection via region-based fully convolutional networks. In: Advances in Neural Information Processing Systems, pp. 379–387 (2016)
2. Ding, J., Chen, B., Liu, H., Huang, M.: Convolutional neural network with data augmentation for SAR target recognition. IEEE Geosci. Remote Sens. Lett. **13**(3), 364–368 (2016)
3. Fu, J., et al.: Dual attention network for scene segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3146–3154 (2019)
4. Gao, X., et al.: An end-to-end neural network for road extraction from remote sensing imagery by multiple feature pyramid network. IEEE Access **6**, 39401–39414 (2018)
5. Kaiming, H., Xiangyu, Z., Shaoqing, R., Jian, S.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**(9), 1904–1916 (2015)
6. Li, K., Cheng, G., Bu, S., You, X.: Rotation-insensitive and context-augmented object detection in remote sensing images. IEEE Trans. Geosci. Remote Sens. **56**(4), 2337–2348 (2017)

7. Li, Q., Mou, L., Liu, Q., Wang, Y., Zhu, X.X.: HSF-NET: multiscale deep feature embedding for ship detection in optical remote sensing imagery. IEEE Trans. Geosci. Remote Sens. **56**(12), 7147–7161 (2018)
8. Lin, T.Y., Dollar, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125 (2017)
9. Lin, T.-Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
10. Liu, W., et al.: SSD: single shot MultiBox detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
11. Long, Y., Gong, Y., Xiao, Z., Liu, Q.: Accurate object localization in remote sensing images based on convolutional neural networks. IEEE Trans. Geosci. Remote Sens. **55**(5), 2486–2498 (2017)
12. Ma, W., Guo, Q., Wu, Y., Zhao, W., Zhang, X., Jiao, L.: A novel multi-model decision fusion network for object detection in remote sensing images. Remote Sensing **11**(7), 737 (2019)
13. Mou, L., Zhu, X.X.: Vehicle instance segmentation from aerial image and video using a multitask learning residual fully convolutional network. IEEE Trans. Geosci. Remote Sens. **56**(11), 6699–6711 (2018)
14. Qiu, H., Li, H., Wu, Q., Meng, F., Ngan, K.N., Shi, H.: A2RMNet: adaptively aspect ratio multi-scale network for object detection in remote sensing images. Remote Sens. **11**(13), 1594 (2019)
15. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
16. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
17. Ren, Y., Zhu, C., Xiao, S.: Small object detection in optical remote sensing images via modified faster R-CNN. Appl. Sci. **8**(5), 813 (2018)
18. Singh, B., Davis, L.S.: An analysis of scale invariance in object detection snip. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3578–3587 (2018)
19. Xiao, Z., Liu, Q., Tang, G., Zhai, X.: Elliptic Fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote sensing images. Int. J. Remote Sens. **36**(2), 618–644 (2015)
20. Xie, W., Qin, H., Li, Y., Wang, Z., Lei, J.: A novel effectively optimized one-stage network for object detection in remote sensing imagery. Remote Sens. **11**(11), 1376 (2019)
21. Yang, X., et al.: Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks. Remote Sens. **10**(1), 132 (2018)