CCML 2021 第十八届中国机器学习会议
The 18th China Conference on Machine Learning
2021年8月6-8日 中国·长沙
AUG 6-8, 2021 CHANGSHA, CHINA

Abstract

In recent years, the target detection task has made encouraging progress, but there are still challenges for densely distributed objects, which is particularly evident in high-resolution remote sensing images. Aiming at the problem that the existing target detection models have poor performance in dense target detection, this thesis proposes a deformable convolution and attention model for dense target detection in remote sensing images. Firstly, in the feature extraction stage, the feature pyramid module is designed, and the multi-layer features are fused by the way of down-sampling and jumping connection to get the feature map containing more information. Secondly, attention mechanism is adopted to focus on pixel information, channel information and spatial information in feature map to reduce interference information. Finally, the deformable convolution module is used to learn the deformation information of dense targets.In the open datasets UCAS-AOD and NWPU VHR-10, the mAP of the proposed model reaches 94.76% and 87.93%, respectively. Compared with other models, the detection accuracy is more advantageous.

Introduction

Remote sensing images bring the following challenges to the target detection task due to the specificity of imaging: (1) small targets, remote sensing images image a large area of the ground, most man-made targets occupy a small number of pixels, and these small targets may be surrounded by complex backgrounds; (2) the target distribution may be dense, e.g., vehicles, ships, oil storage tanks, etc. Traditional target detection algorithms rely on manually extracted features, such as histogram of oriented gradient features (HOG) , bag-of-words features (BOW) , etc., which have poor robustness and limited applications. With the development of deep learning, deep models provide feasible means of automatic feature extraction and greatly improve target detection performance. Faster R-CNN (Faster region-based convolutional neural networks) , a two-stage target detection model based on convolutional neural networks (CNN), has good performance and is a model that combines the steps of target candidate region extraction, Faster R-CNN (Faster region-based convolutional neural networks) has good performance and is an end-to-end model that integrates the steps of target candidate region extraction, target location regression and target classification, which improves the detection accuracy and reduces the consumption of computational resources. However, when the Faster R-CNN model is directly used for remote sensing image dense target detection, due to the dense and irregular distribution of targets to be measured, overlap occurs when predicting the bounding box, resulting in poor detection accuracy of dense targets.

To solve the problems faced in the above study, we propose the ADF-RCNN (Attention and deformable Faster R-CNN) model. First, we enhance the detail and semantic information in the top-down structure of the feature pyramid by down-sampling and jump-join operations, and increase the detail information in the top feature map by fusing the top feature map with the bottom feature down-sampling. The feature map of the top layer in top-down reduces the transfer of information in the intermediate layers by jumping connections, so that the semantic information in the top feature map is effectively transferred to the bottom feature map to enhance the semantic information in the bottom feature map; secondly, the attention to dense targets is enhanced and the influence of useless information is attenuated by pixel attention and channel and spatial attention; finally, deformable convolution is used to learn the output feature deformation information to improve the detection efficiency of dense targets. The contributions of this paper are (1) for small targets, a jumping feature pyramid module is designed for small target detection by feature fusion; (2) for dense targets, an attention mechanism is designed to reduce the adverse effects of interference information in the feature map, and the problem of restricted geometric deformation modeling of the target to be detected is solved by means of deformable convolution.

Method/Model

The structure of ADF-RCNN model is shown in Figure 1. ADF-RCNN model is built on the basis of Faster R-CNN and consists of feature pyramid module, attention module, deformable convolution module, RPN (Region Proposal Network), target region pooling (ROI Pooling), target classification and regression modules The backbone network is ResNet 50.

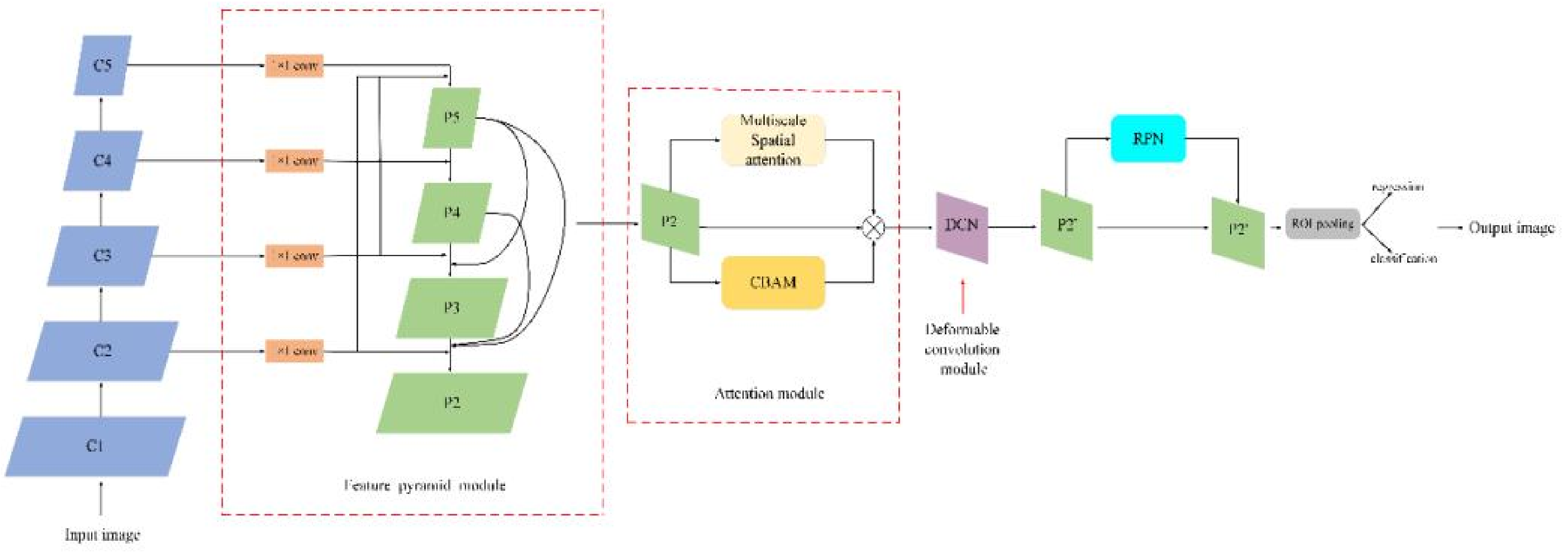

Fig. 1. ADF-RCNN Model structure.

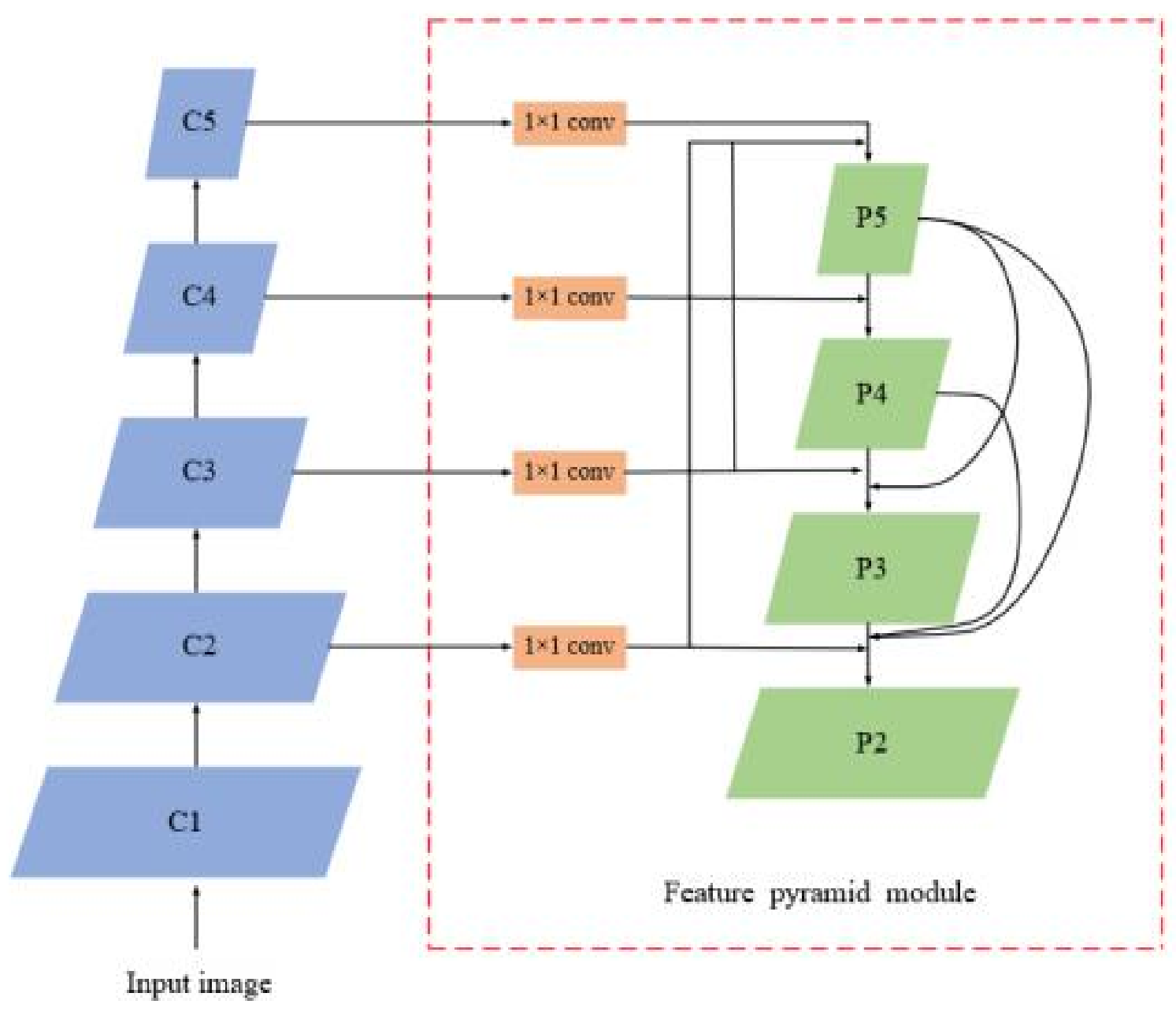

Fig. 2.The structure of pyramid model.

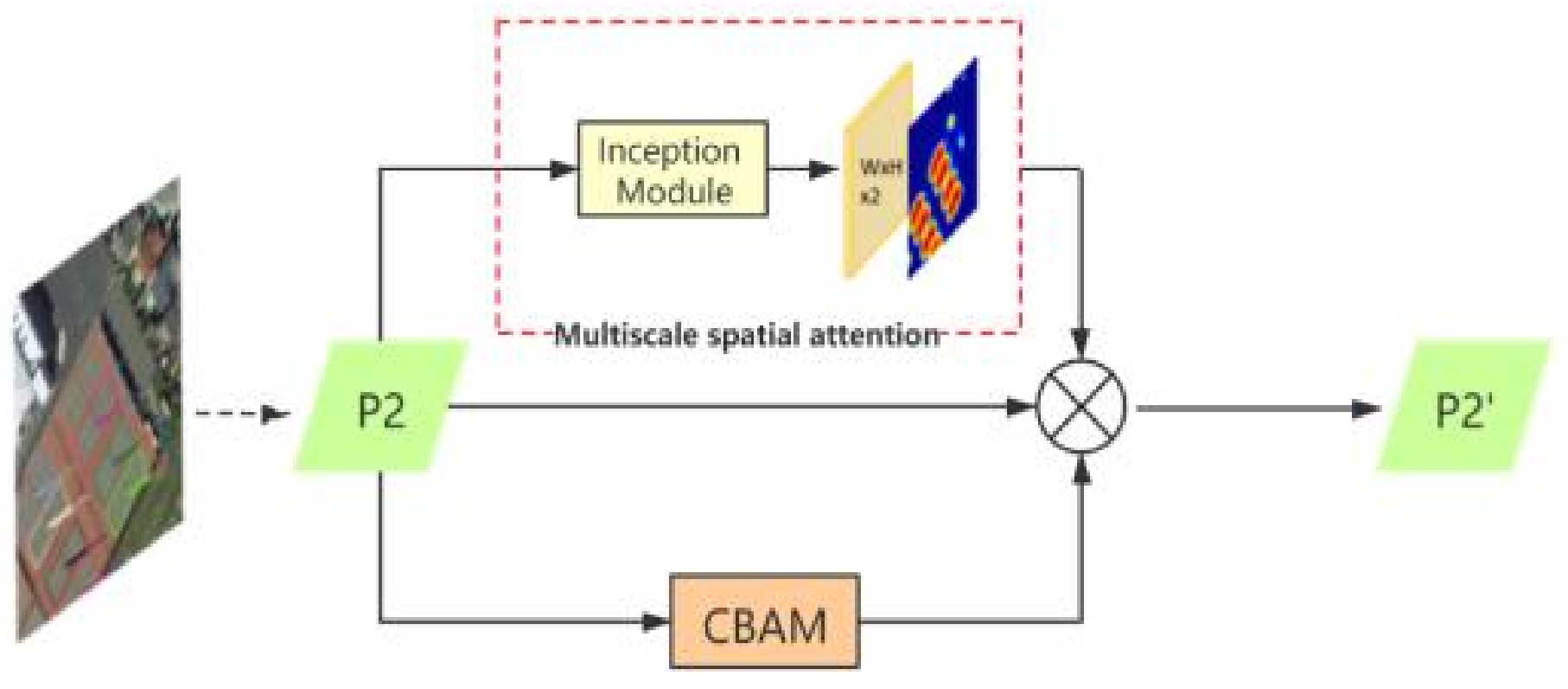

Fig. 3.The structure of CBAM.

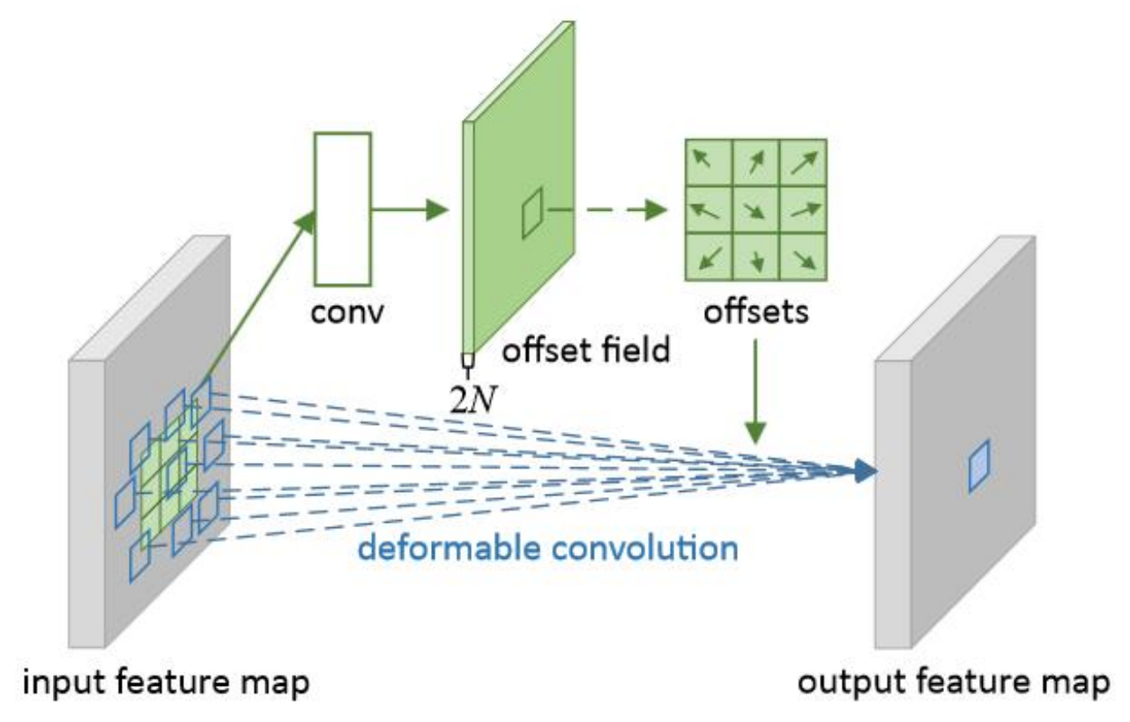

Fig. 4.The structure of the deformable convolution.

Experiments

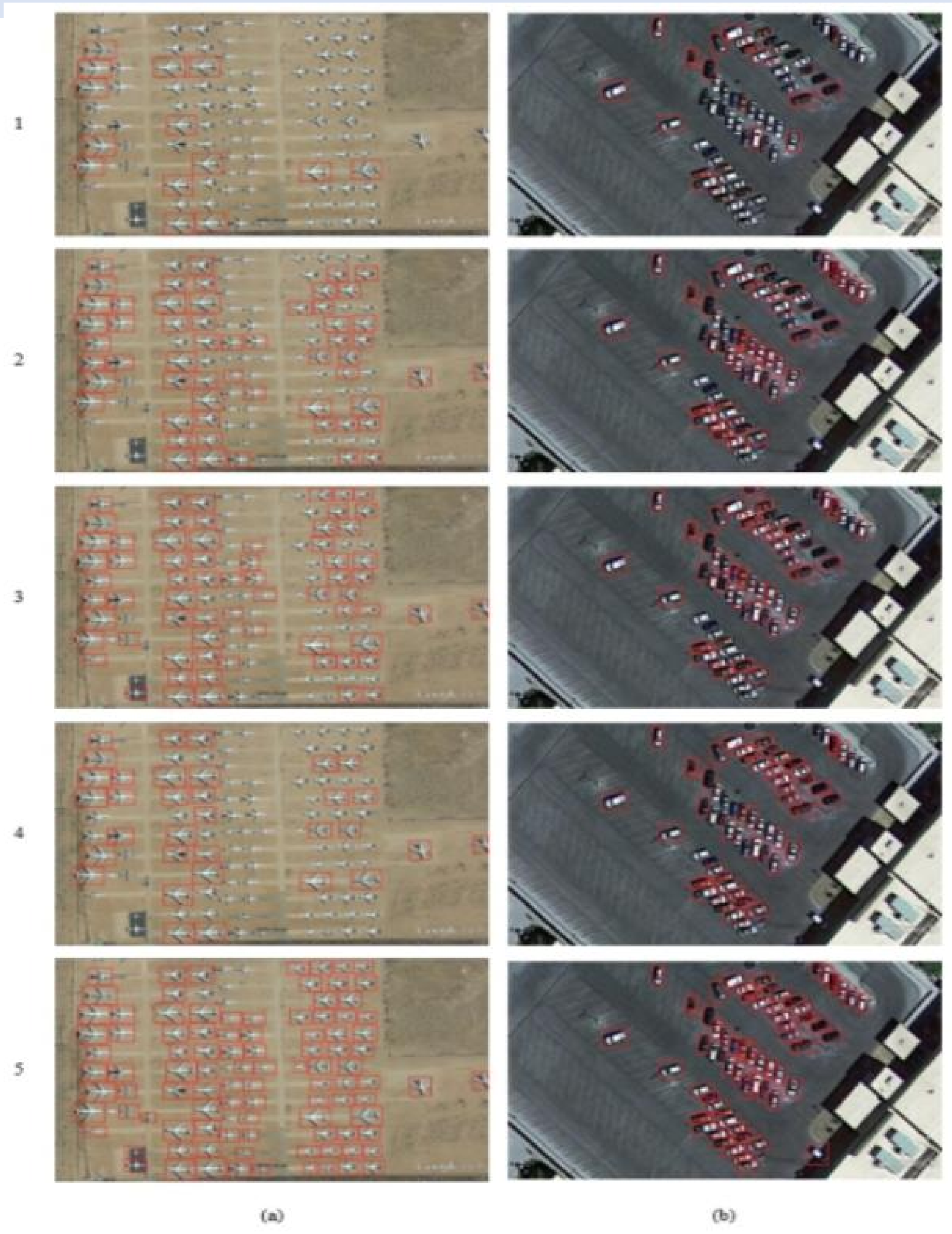

Fig. 5. Part of the results of the ablation experiment. (a) Test results of aircraft; (b) Testresults of car. 1-5 correspond to: Faster R-CNN (baseline), baseline + Feature pyramid module, baseline + Attention module, baseline + Deformable convolution module、 our module.

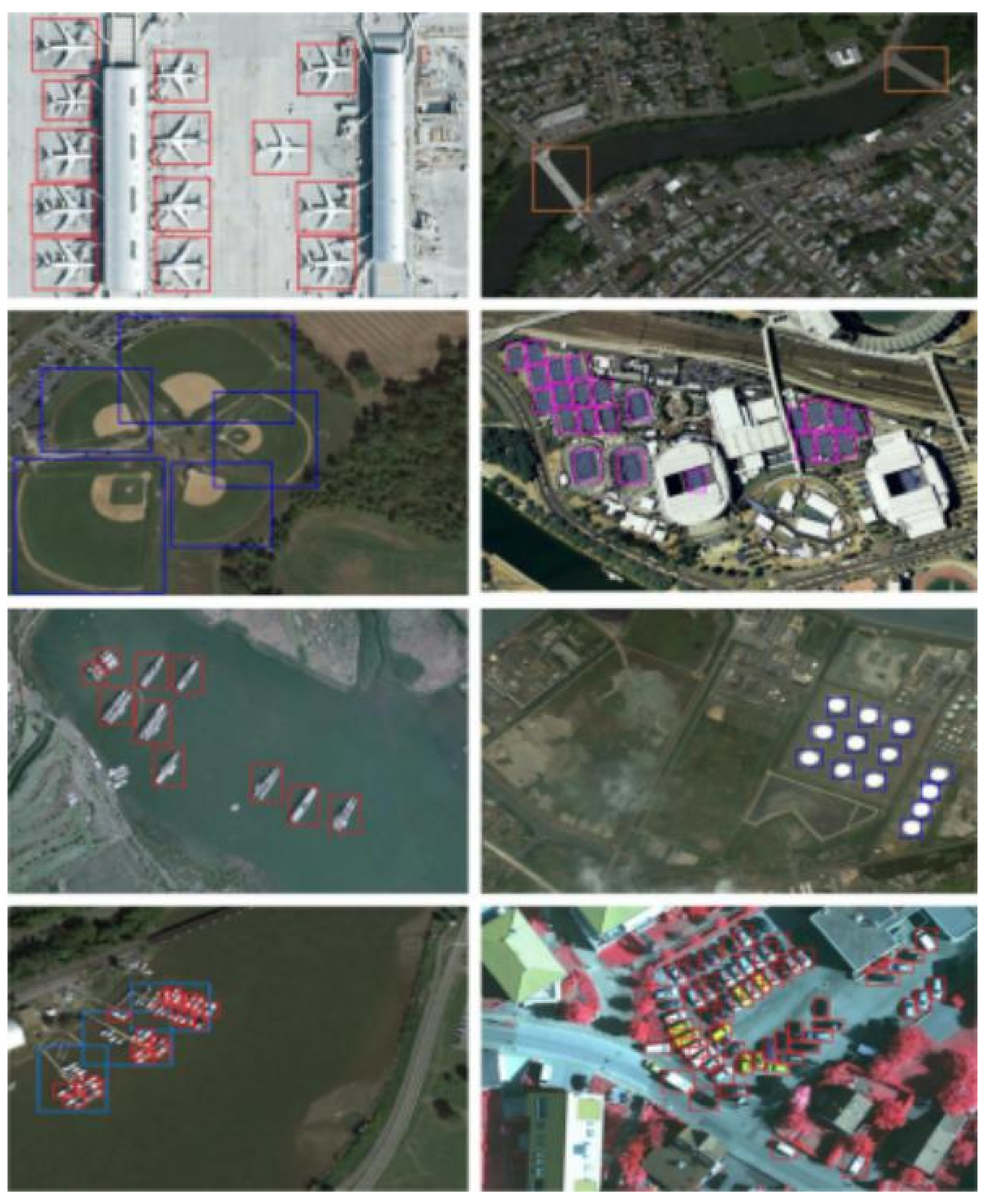

Fig. 6. Detection results of some objects in NWPU VHR-10.

Table 1. Comparison of the proposed method and other methods on the UCAS-AOD dataset

Method	Plane(%)	Car(%)	mAP(%)
SSD	88.13	85.09	86.61
Faster R-CNN	92.83	89.93	91.38
ADF-RCNN	**96.15**	**92.97**	**94.76**

Conclusion

To solve the problem of difficult detection of small and dense targets, the ADF-RCNN model is proposed in this paper. By constructing a feature pyramid structure, the bottom feature map is fused with the top feature map by using downsampling and jump connection to enrich the detail and semantic information of the feature map. By combining pixel attention, channel and spatial attention, the feature maps in the pyramid structure can focus on the features of dense targets more effectively, and the deformable convolution is used to learn the deformation results of dense targets. The experimental results on UCAS-AOD and NWPU VHR-10 datasets demonstrate the effectiveness of the proposed model.

投稿人 姓名：汪亚妮 梁敏 汪西莉
单位： 陕西师范大学