Title of conference: MIPPR 2019: Pattern Recognition and Computer Vision

**Object detection for rotated and densely packed targets in aerial images**

Xiangyu Liu, Hong Pan[[1]](#footnote-1)

Southeast University

**Purpose**

With the development of commercial satellite and unmanned aerial vehicle (UAV), aerial images with many recognizable targets are becoming easy to access. Several large-scale datasets with well-annotated aerial images have been promoting object detection research in this field. Object detection in aerial images is important yet challenging problem due to the overhead perspective, complicated background and large variation in scale, orientation and appearance of objects in aerial images. While object detection in ground-based images has benefited from remarkable researches into new deep learning approaches, utilizing those technologies with object detection in aerial images is nontrivial. This paper aims to propose a pipeline that accurately detects objects of variant scales, orientations and appearances in high resolution aerial images.

**Methods**

**Feature extraction** As the backbone of convolutional neural networks, the quality of extracted feature has an impact on the performance of the following regression and classification tasks. To address the problem of multi-scale objects in aerial images, feature pyramid structure is used to extract multi-scale feature. Experiments[1], [2] have shown that shorten information path and making information propagation more easily are useful for object detection. For better information propagation of multi-scale objects, we add a bottom-up path with lateral connection to corresponding FPN levels like [3] did, which ensures our network accurately extracts multi-scale feature from complicated background.

**Regression of oriented bounding-box** One of the problems transitioning generic object detection technologies to object detection in aerial images is that the arbitrary orientations due to the overhead perspective. The most effective way to handle the problem is to use rotated PSRoI align[4] or deformable PSRoI pooling[5] to get rotation-invariant feature representation of orientated targets. Our experiments found that it’s essential to get better performance for detection of oriented densely packed targets in aerial images.

**Bells and whistles** The problem of class imbalance is often encountered in object detection in aerial images due to the unbalanced distribution of different targets in real world. We used the Stratified Online Hard Example Mining which feeds hard examples to the backpropagation process to overcome the dilemma of unbalanced data. Another problem which may affect the performance of detector is appearance ambiguity, that is to say, small inter-class difference on target appearance. For example, appearances of bridge and road, basketball court and soccer ball field are quite similar from overhead perspective. Including contextual information of RoI can help our detector distinguish ambiguous targets.

The architecture of our algorithm is as follows.

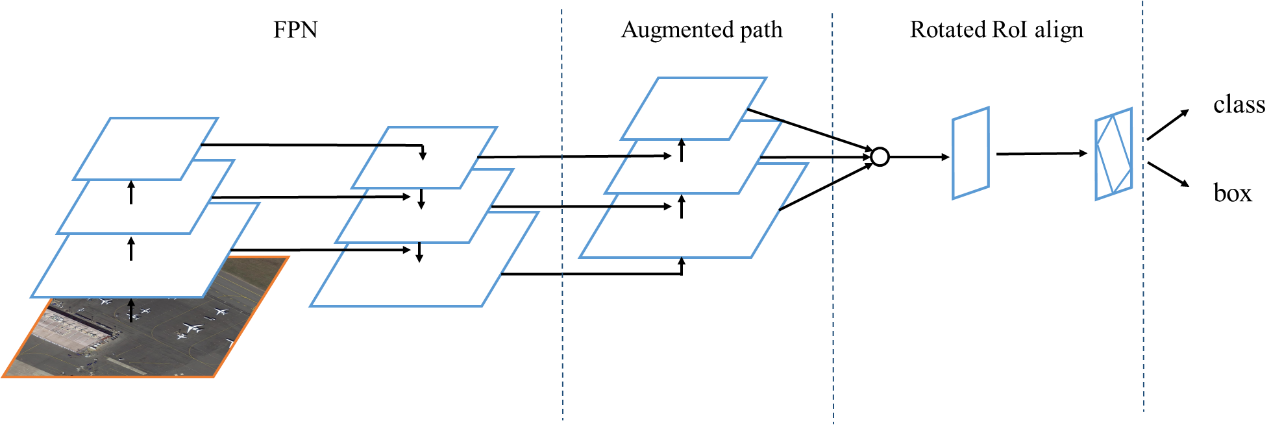


Figure 1. The architecture of our algorithm. We use FPN as our backbone, then the augmented bottom-up path is added through lateral connection with FPN. Multi-scale features are fused and rotated RoI align is conducted afterwards. Note that channel dimension is omitted for brevity.

**Results**

We test our network using a large-scale dataset named DOTA[6], which consists of high resolution aerial images with oriented bounding box annotated. It contains 2806 images with objects of 15 categories, including plane, baseball diamond(BD), bridge, ground track field(GTF), small vehicle(SV), large vehicle(LV), ship, tennis court(TC), basketball court(BC), storage tank(ST), soccer-ball field(SBF), roundabout(RA), harbor, swimming pooling(SP) and helicopter(HC). Our experiment results are shown in Table 1.

Table 1. Experiment results of our algorithm and some popular generic object detection networks on DOTA dataset.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | YOLOv2 | Faster R-CNN | YOLOv3 | ours |
| *Plane* | 78.72 | 87.96 | 88.04 | **89.27** |
| *BD* | 39.11 | 79.61 | 62.04 | **81.92** |
| *Bridge* | 22.64 | 51.12 | 44.56 | **56.35** |
| *GTF* | 34.71 | 62.2 | 56.06 | **69.85** |
| *SV* | 32.26 | 54.97 | 76.53 | **66.35** |
| *LV* | 23.12 | **69.23** | 57.31 | 65.76 |
| *Ship* | 49.53 | **85.20** | 78.59 | 84.99 |
| *TC* | 76.15 | **90.86** | 90.46 | 90.68 |
| *BC* | 43.53 | 78.69 | 77.15 | **82.67** |
| *ST* | 36.6 | 75.88 | 80.42 | **84.49** |
| *SBF* | 23.93 | 54.31 | 44.86 | **59.05** |
| *RA* | 43.88 | 61.72 | 58.56 | **65.57** |
| *Harbor* | 40.99 | 73.12 | 59.86 | **75.68** |
| *SP* | 40.53 | 68.53 | 62.29 | **72.17** |
| *HC* | 4.69 | 58.59 | 55.62 | **69.57** |
| ***mAP*** | 39.36 | 70.13 | 66.16 | **74.29** |

**Conclusion**

In this paper, we constructed a detection network for aerial images. Our network focuses primarily on oriented and densely packed objects, which is hard for generic object detectors due to the overlap of neighbored horizontal bounding boxes. We have tested our network on a large-scale aerial images dataset which includes many small, dense packed targets with arbitrary orientations. The result shows the effectiveness of our method.

**References**

[1] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[2] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.

[3] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “Path Aggregation Network for Instance Segmentation,” *ArXiv180301534 Cs*, Mar. 2018.

[4] J. Ding, N. Xue, Y. Long, G.-S. Xia, and Q. Lu, “Learning RoI Transformer for Detecting Oriented Objects in Aerial Images,” *ArXiv181200155 Cs*, Dec. 2018.

[5] Q. Yang *et al.*, “IncepText: A New Inception-Text Module with Deformable PSROI Pooling for Multi-Oriented Scene Text Detection,” *ArXiv180501167 Cs*, May 2018.

[6] G.-S. Xia *et al.*, “DOTA: A Large-scale Dataset for Object Detection in Aerial Images,” *ArXiv171110398 Cs*, Nov. 2017.

1. Corresponding author: mspanhong@hotmail.com [↑](#footnote-ref-1)