Bottle Detection from Low-Altitude UAV Imagery with the Rotation Bounding Box

1st Jinwang Wang
School of Electronic Information, Wuhan University
Wuhan, China
jwwangchn@whu.edu.cn

2nd Given Name Surname
School of Electronic Information, Wuhan University
Wuhan, China
email address

3nd Given Name Surname
School of Electronic Information, Wuhan University
Wuhan, China
email address

***Abstract*—This paper proposed a vision based system used in UAV for bottle detection. In UAV's perspective, bottles has huge variation in scale, ratios orientation and shape. Traditional annotation methods use the horizontal bounding box to annotate the objects, but we use the method of oriented bounding box which is easier to separate the objects from background. It is worth mentioning that we have collected $25,407$ images of bottles with size of 342×342 pixels using UAV in the different scenes. These bottles are in a wide variety of scales, ratios, orientations and shapes. And we also have annotated these images using oriented bounding box. These fully annotated images contain $34,791$ instances, each of which is annotated by an arbitrary $\{c_x, c_y, h, w, \theta\}$ quadrilateral. To build a baseline for bottle detection, we evaluate some state-of-the-art object detection algorithms like Faster R-CNN, SSD, YOLOv2 and RRPN on our UAV-Bottle Dataset (UAV-BD). Experiments demonstrate that our dataset contributes to environmental protection applications and are quite challenging due to the difficulties of locating the multi-angle bottles and separating them effectively from the background.**

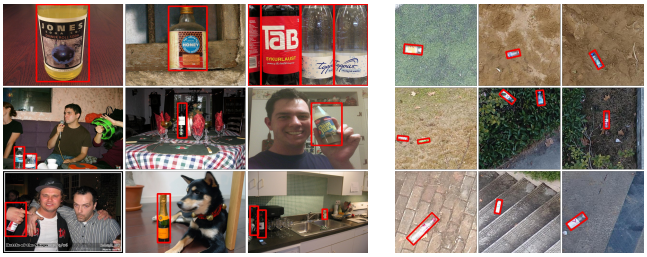
***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

At present, there is a lot of rubbish in the tourist attractions, including all kinds of bottles, these bottles are mainly recycled by manpower. This method is time-consuming, laborious and dangerous. In order to solve these problems, we use the unmanned aerial vehicle (UAV) to locate and even recycle bottles. We also have built a UAV bottle dataset (UAV-BD) to locate bottles more effictively.

Detecting objects in UAV images plays an important role in many ways and has received significant attention in recent years [1]. However, it is still a challenging problem due to the high resolution with the extremely high level of detail, various shooting platform, limited annotated data, and limited processing time for real-time applications [2]. In UAV images, the bottle looks completely different from the bottle in datasets such as PASCAL VOC [3], Microsoft COCO [4], etc. The difference between PASCAL VOC and our dataset is shown in Fig.1.

Identify applicable funding agency here. If none, delete this.



(a) Bottles in PASCAL VOC (b) Bottles in UAV images

Fig. 1. Comparison of bottles in PASCAL VOC dataset and UAV images

As to UAV images, detecting bottles exists several unique challenges. First, the size of bottles is very small, which is generally less than 50×50 pixels. At the same time, due to the different altitudes of the UAV, the size of bottles differs in scale. Second, in UAV images, the backgrounds of the bottles are very complex which results in poor performance of the general algorithm. Third, in contrast to conventional object detection datasets, where objects are generally oriented upward due to gravity [5], the bottles in UAV images often appear with arbitrary orientations depending on the shooting angle of the UAV camera, as illustrated in Fig.1(b). Fourth, the plastic bottles are transparent, so the background will can be seen through the bottle, increasing the difficulty of detection. In order to abtain the better performance of an algorithm for the bottle detection problem, we establish a UAV perspective bottle detection dataset and benchmark, which we call UAV-Bottle Dataset (UAV-BD).

Object detection is one of the most challenging tasks in computer vision and has attracted a lot of attention all the time [6]. As the development of deep learning, convolutional neural networks(CNN) has been applied for solving the object detection problem and the methods based on CNN have achieved state-of-the-art performance [6].

Most of the existing detection methods use the horizonal bounding boxes to locate objects in images. The horizonal

bounding box is a rotation variant data structure, but it works badly when the detector deals with orientation variations of target objects. To make the approach insensitive to objects in-plane rotation, some efforts are made either adjusting the orientation or trying to extract rotation insensitive features. Unlike these methods which try to eliminate the effect of rotation on the feature level, we prefer to make the rotation information useful for feature extraction so that the detection results involve the angle information. Therefore, the detection results are rotatable, whereas the performance of the detector is rotation invariant [6].

II. UAV-Bottle Dataset

A. Dataset collection

In the introduction section, we analyzed the challenges that bottle detection may encounter in the UAV perspective.

- The size of the bottles is very small and their scales change very much in UAV images. For solving this problem, we collect images at different flight altitudes.
- In UAV images, the backgrounds of the bottles are very complex. In order to increase the diversity of dataset, we classify the possible scenes and divide them into eight scenes. Eight scenes are illustrated in Fig.2 and Fig.3. In Fig.2, we show eight full images of eight scenes whose sizes are 5472×3078. In Fig.3, we show the segmented images of eight scenes, each scene contains three images whose sizes are 342×342.
- The bottles in UAV images often appear with arbitrary orientations. We find the orientation of bottles will affect the robustness of the trained model, so we annotate images by using the oriented bounding box.
- The plastic bottles are usually transparent, so the background can be seen through the bottle, increasing the difficulty of detection. Our dataset includes a lot of examples of transparent bottles, so we can use this large dataset to train a robust model.

The UAV platform used in this work is DJI Phantom 4 quadcopter integrated with a 3-axis stabilized gimbal.

Images are collected by a camera mounted on UAV. The resolution of the full images are 5472×3078. For dataset collection, we follow three key suggestions: (1) collecting images with bottles of a wide range of scale and aspect ratios; (2) collecting images with bottles of different background scenes; (3) collecting images with bottles of different orientations; (4) using as many types of bottles as possible.

To collect images covering bottles of a wide range of scales and aspect ratios, images at different flight altitudes ranging from $10m$ to $30m$ are collected. Eight background scenes are chosen and annotated in our UAV-BD, including *Bush forest land*, *Waste land*, *Step*, *Forest land*, *Flat ground*, *Plastic stadium*, *Sand land* and *Grassland*. In this work, UAV images are collected from two periods. For each period, images are collected by using different bottles and different flight altitudes. The background scenes are selected according to whether a kind of scene is common and its value for real-world applications [5].

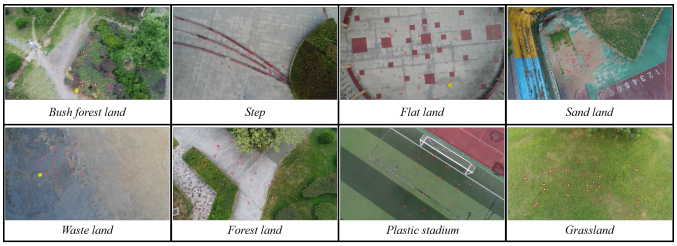


Fig. 2. Samples of annotated images in UAV-BD. We show one full image which size is 5472×3078 per each scene.

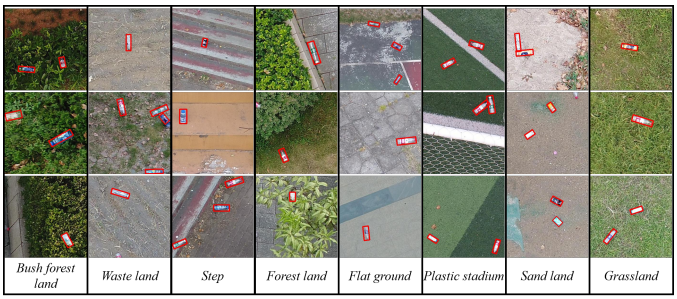


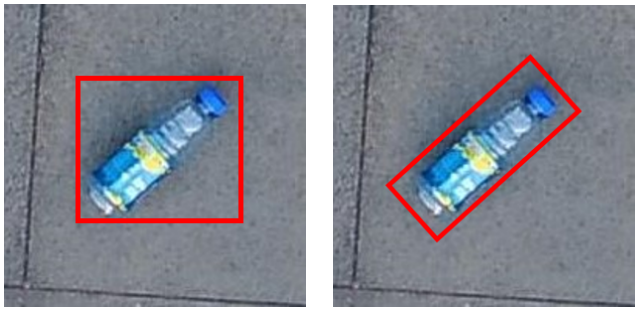
Fig. 3. Sample of annotated images in UAV-BD. We show three images which sizes are 342×342 per each scene.

B. Annotation method

We build the UAV-BD for the bottle detection problem by collecting bottle images using UAV. In the annotating stage, we consider different ways to annotate images. In computer vision field, many visual concepts such as region descriptions, objects, attributes, and relationships, are annotated with horizontal bounding boxes, as show in [5], [7]. A common description of horizontal bounding boxes is (c_x, c_y, w, h) or (x, y, w, h), where (c_x, c_y) is the center location of horizontal bounding box, w, h are the width and height and (x, y) is the top left location [5].

Objects without many orientations can be adequately annotated with this method. However, horizontal bounding box cannot accurately or compactly outline oriented instances such as the objects in UAV images. In UAV images, the overlap between two bounding boxes is sometimes very large that state-of-the-art object detection methods cannot diffetentiate them [5]. At the same time, horizontal bounding box may contain lots of backgrounds while annotating the target, it's especially the kind of objects with large aspect ratios. In order to remedy these, we need to find an annotation method suitable for oriented bottles in UAV images.

An option for annotating oriented objects is arbitrary quadrilatral bounding boxes, this annotation method can be denoted as $(x_i, y_i), i = 1, 2, 3, 4$, where (x_i, y_i) denotes the positions of the oriented bounding boxes' vertices in the image [5]. The vertices are arranged in a clockwise order. But due to bottles are rigid, almost no deformation, so we choose other way which is θ-based oriented bounding box which is adopted in some text detection benchmarks, namely (c_x, c_y, w, h, θ), where θ denotes the angle from the horizontal direction of



(a) Normal annotation method (b) Our annotation method

Fig. 4. Comparison of traditional bounding box and rotatable bounding box.

the horizontal bounding box [5]. The tool for annotating is roLabelImg[1].

C. Dataset Splits

In order to ensure that the training and testing data distributions approximately match, we randomly select 64% of the UAV-BD as the training data, 16% as validation data, and 20% as the testing data. We will publicly provide all the full images and segmented images with ground truth for UAV-BD.

D. Dataset Statistics

UAV images are usually very large in size compared to conventional images datasets. The size of full images in UAV-BD is 5472×3078 while most images in conventional datasets (e.g. PASCAL VOC and Microsoft COCO) are no more than 1000 [8]. We firstly make annotations on the full images without segmenting it into pieces to avoid the single instances is segmented into different pieces. But we find full image is too large to be trianed CNN based algorithms. So we segment full images into 144 small pieces, the size of piece is 342×342, note that we abandon the instances at the border. We will use small pieces to train CNN based detection model.

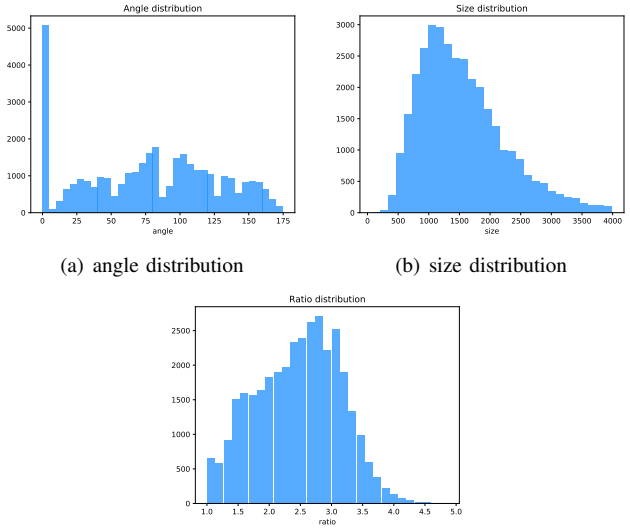
The statistics of the UAV-BD is shown in the TableI, where n_1 is the full images number for each scene, n_2 is the small images number for each scene, n_3 is the number of instances in full images for each scene, n_4 is the number of instances in small images for each scene. So UAV-BD contains about $34,791$ object instances in $25,407$ images. The "Grassland" scene has the largest number of object instance: $7,795$ instances in $5,785$ images. The "Step" scene has the smallest number of instances: 2106 instances in $1,325$ images.

Bottles are usually rigid body, so we can get some prior information for training out detection model, for example, we can use angle distribution, size distribution, ratio distribution, etc. to improve the performance of detection model. For UVA-BD, we plot angle, size and ratio distribution which are illustrated in Fig.5. We can see that bottles' angle in images are almost uniform in Fig.5(a). Bottles' size are usually range from 500pixel^2 to 3000pixel^2. Bottles' ratio are usually range from 1.0 to 4.0. Note that we use these statistics data to design detection models.

[1]https://github.com/cgvict/roLabelImg.git

TABLE I
IMAGES AND INSTANCES NUMBER IN UAV-BD

Scenes	n_1	n_2	n_3	n_4
Bush forest land	230	4134	1812	3047
Waste land	379	7598	4355	5800
Step	135	2691	1325	2106
Forest land	285	5724	3702	4891
Flat land	134	2803	1538	2142
Plastic stadium	336	6807	4180	4998
Sand land	249	5570	2704	4008
Grassland	456	9029	5778	7787
Total	2204	44356	25394	34779



(a) angle distribution (b) size distribution

(c) ratio distribution

Fig. 5. The angle, size and ratio distribution of UAV-BD

III. BASELINES AND METHODS

All experiments are performed using UAV-BD, the training, validation and testing date include 16258, 4065 and 5081 images, respectively. Note that the images for training, testing and validation are the size of 342. The whole UAV-BD contains 16258 images with 22211 instances for training, 5081 images with 6944 instances for testing and 4055 images with 5624 instances for validation.

Here, we present three different approaches to our task, which vary by their use of detection framework and data annotating method. For horizontal object detection, we select Faster R-CNN [9] and SSD [10] as our baseline testing algorithms for their excellent performance on general object detection. For oriented object detection, wo modify the original Rotation Region Proposal Networks(RRPN) algorithm [11] such that it can predict properly oriented bounding boxes denoted as $\{c_x, c_y, w, h, \theta\}$. Note that (c_x, c_y) is the central coordinate of the oriented bounding box, w and h is the width and height of the oriented bounding box, θ is the rotation angle of the oriented bounding box.

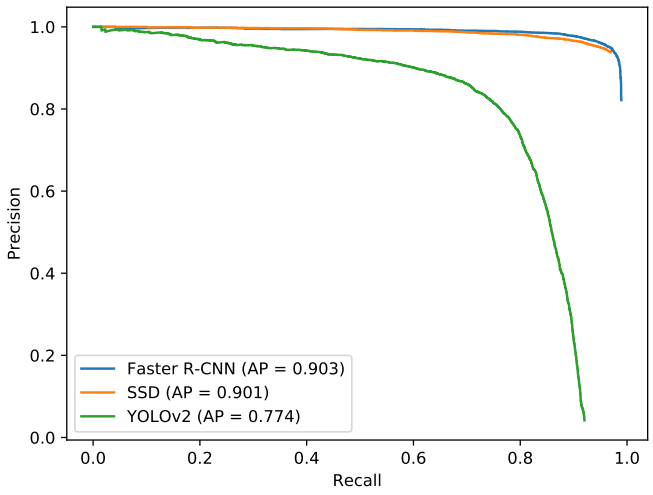


Fig. 6. Numerical results (AP) of baseline models evaluated with HBB ground truths.

A. Baselines with Horizontal Bounding Boxes

Ground truths for horizontal bounding boxes(HBB) experiments are generated by calculating the axis-aligned bounding boxes over original bounding boxes. To make it fair, we keep all the experiments' setting and hyper-parameters the same as depicted in corresponding papers [9], [10], [12].

The experimental results of HBB prediction are shown in Fig.9. In Fig.9, first row illustrates the results for Faster R-CNN, second row illustrates the results for SSD, third row illustrates the results for YOLOv2.

B. Baseline with Oriented Bounding Boxes

Prediction of oriented bounding boxes(OBB) is difficult because the state-of-the-art detection methods are not designed for oriented objects. Therefor, we choose Rotation Region Proposal Networks(RRPN) [11] as the framework for its accuracy and efficiency, then we modify it to adapt UAV-BD on the basis of dataset statistics in section II-D.

For RRPN, it's based on Faster R-CNN, in Faster R-CNN, Region of Interests(RoIs) generated by Region Proposal Network(RPN) are rectangles which can be write as $R = (x_{min}, y_{min}, x_{max}, y_{max}) = (c_x, c_y, w, h)$. These RoIs have regressed from k anchors which are generated by some predefined scales and aspect ratios. But in RRPN, it uses predefined scales, aspect ratios and *angles* to generate RoI, so RRPN can predict oriented bounding boxes which can be written as $R = (c_x, c_y, w, h, \theta)$. In the section II-D, we analyzed the size, aspect ratio and angle distributions of UAV-BD, so we can select reasonable scales, aspect ratios and angles value to generate new anchors which are shown in Fig.7.

The experimental results of RRPN prediction are shown in Fig.9. In Fig.9, forth row illustrates the results for RRPN.

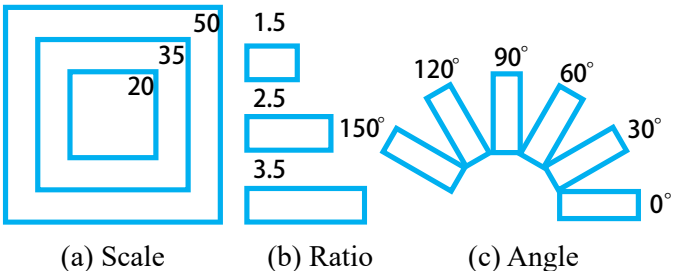


Fig. 7. Anchor strategy in our framework of RRPN

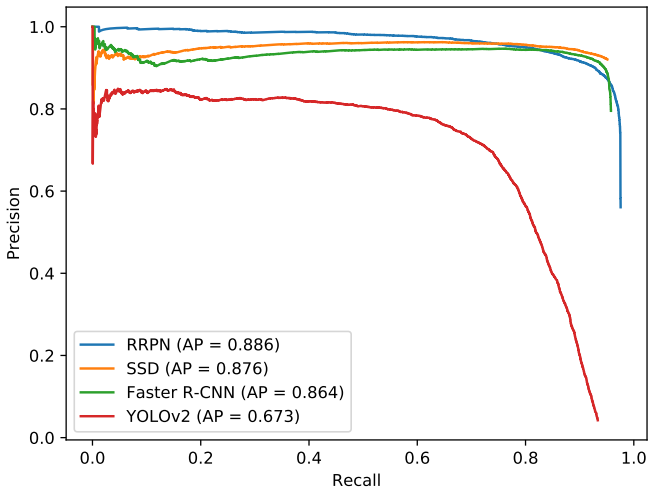


Fig. 8. Numerical results (AP) of baseline models evaluated with OBB ground truths.

C. Experimental Analysis

Fig.6 show the quantitative comparison result of three baseline models evaluated with HBB ground truth, measured by precision-recall curve and AP values. For evaluation metrics, we adopt the same AP calculation as for PASCAL VOC. As can be seen from it, Faster R-CNN, SSD, YOLOv2 obtain 90.3%, 90.1%, 77.4% performances, respectively.

Fig.8 show the quantitative comparison result of five baseline methods evaluated with OBB ground truth, measured by precision-recall curve and AP values. As can be seen from it, RRPN, SSD, Faster R-CNN, YOLOv2, DRBox obtain 88.6%, 87.6%, 86.4%, 67.3%, 32.7% performances, respectively.

In Fig.9, we compare the results between objects detection experiments of HBB and OBB. For oriented objects shown in Fig.9, location precision of objects in HBB experiments are much lower than OBB experiments and results are suppressed through progress operations. So OBB regression is the correct way for oriented object detection that can be really integrated to real applications.

IV. Conclusion

We establish a large-scale dataset for bottles detection in UAV images which we call UAV-BD. In contrast to general object detection benchmarks, we annotate a huge number of

Fig. 9. Visualization results of testing on UAV-BD using well-trained Faster R-CNN, SSD and RRPN. **TOP** to **Bottom** respectively illustrate the results for Faster R-CNN, SSD, YOLOv2 and RRPN.

well-distributed bottles with oriented bounding boxes. We also establish a benchmark for bottle detection in UAV-BD and show the feasibility to produce oriented bounding boxes by modifying a oriented bounding box based detection algorithm.

REFERENCES

[1] T. Moranduzzo and F. Melgani, "Automatic car counting method for unmanned aerial vehicle images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 3, pp. 1635–1647, 2014.

[2] Y. Xu, G. Yu, Y. Wang, X. Wu, and Y. Ma, "Car detection from low-altitude uav imagery with the faster r-cnn," *Journal of Advanced Transportation*, vol. 2017, 2017.

[3] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, Jan. 2015.

[4] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

[5] G.-S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "Dota: A large-scale dataset for object detection in aerial images," in *IEEE CVPR*, 2018.

[6] L. Liu, Z. Pan, and B. Lei, "Learning a rotation invariant detector with rotatable bounding box," *arXiv preprint arXiv:1711.09405*, 2017.

[7] R. Krishna, Y. Zhu, O. Groth, J. Johnson, K. Hata, J. Kravitz, S. Chen, Y. Kalantidis, L.-J. Li, D. A. Shamma *et al.*, "Visual genome: Connecting language and vision using crowdsourced dense image annotations," *International Journal of Computer Vision*, vol. 123, no. 1, pp. 32–73, 2017.

[8] C. Chen, M.-Y. Liu, O. Tuzel, and J. Xiao, "R-cnn for small object detection," in *Asian Conference on Computer Vision*. Springer, 2016, pp. 214–230.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[11] J. Ma, W. Shao, H. Ye, L. Wang, H. Wang, Y. Zheng, and X. Xue, "Arbitrary-oriented scene text detection via rotation proposals," *arXiv preprint arXiv:1703.01086*, 2017.

[12] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," *arXiv preprint arXiv:1612.08242*, 2016.