RSTDet: An Aerial Object Detector for Exploring Rotated and Small Targets

Siwei Wang 23020221154120, [1] **Ruofan Xiong 23020221154128,** [1]
Tanzhe Li 23020221154095, [1] **Yujie Liang 23020221154098** [1]

Abstract

Aerial object detection has broad application prospects in intelligent transportation, military and other fields, and has received more and more attention in recent years. Different from common object detection tasks, aerial objects are always non-axis aligned with arbitrary orientations having small size, which makes this task more challenging. Based on these observations, we proposed an aerial object **Det**ector for exploring **R**otated and **S**mall **T**argets, termed RSTDet, which can be well qualified for the task of oriented object and small object detection. To be specific, RSTDet uses the six-parameter rotating representation method. This simple design allows the network to generate high-quality oriented proposals at a lower cost. It not only avoids generating multiple redundant rotating bounding boxes, but also ensures high accuracy of detection. Secondly, there are a large number of small targets in aerial images, and we admit bottom-up path augmentation to solve this problem. With a simple branch in the feature pyramid, the network can more fully mine shallow features, which are particularly important for small target detection. Thirdly, in order to better focus on target features, especially extracting information about small objects, we also add channel and spatial attention to the backbone network. Our method has achieved competitive results on the general aerial object detection dataset, DOTA, which demonstrates the effectiveness of our proposed method.

Introduction

The rapid development of deep learning has made many breakthroughs on object detection in recent years. Nevertheless, small object detection (SOD) is still the bottleneck of object detection (Cheng et al. 2022). As one of the important applications of SOD, aerial object detection is even more difficult than general object detection task because of the small size of the detected objects with arbitrary rotation direction.

Aerial object detection is undoubtedly a significant but challenging task, which has attracted increasing attention. The mainstream method generally regards aerial object detection as an angle regression task. The most direct method is to add angle prediction to the general object detection framework (Lin et al. 2017a,b; Ren et al. 2017), or generate proposals with angle (Ma et al. 2018). These simple designs make the original horizontal bounding box become a rotating bounding box, effectively improving the detection effect on aerial objects. However, the simple design cannot meet the needs of real scenes. First, there is usually some deviation in the angle prediction of the model, as shown in Figure 1; second, generating angled proposals will bring a large number of redundant proposals, resulting in huge computing costs, as shown in Figure 2.

Figure 1: Some visualization results of Faster R-CNN towards oriented detection. It is obvious that the angle regression based method may lead to inaccurate results.

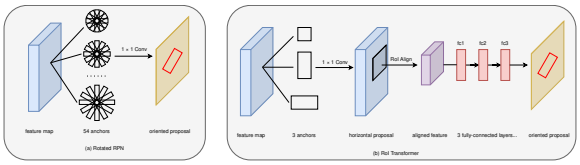

Figure 2: Mainstream methods generate oriented proposals through massive computation. (a) Rotated RPN generates a large amount of oriented anchors with different angles, ratios, and scales. (b) RoI Transformer adopts a lot of complicated operations including RPN, RoI Alignment, and regression, which are very time-consuming.

Recently, many researchers have improved on these issues. For example, Gliding vertex (Xu et al. 2020) adopts a new border representation method, which does not directly predict the target angle. Oriented RepPoints (Li et al. 2022) avoids the regression of angles by using anchor free manner to generate bounding boxes. RoI transformer (Ding et al. 2019a) generates accurate angled bounding boxes through a series of operations such as RPN, RoI Alignment and region. These methods improve the detection effect of rotating targets, but they all ignore the small size of aerial targets, and the recall rate of these methods is not ideal for small objects.

Based on the observation of the above phenomena, we propose RSTDet, a detection framework that can adapt to rotating objects and small objects, which is more suitable for the actual scene of aerial object detection. For rotating objects, we propose six-parameter rotating representation.

It refers to the midpoint offset representation proposed by Oriented RCNN (Xie et al. 2021) and adds two parameters α and β. The six-parameter rotating representation determines a rotation box through six parameters, and adds the operation of taking points from the circumscribed circle. In combination with the original border representation, we can adaptively generate boxes with angles, and solve the problems of inaccurate angle regression and high calculation cost. For small objects, RSTDet adopts a bottom-up path augmentation (Liu et al. 2018) in the feature pyramid, which enables the backbone network to better integrate multi-scale features. In addition, attention mechanism is also applied to the network of feature extraction and feature fusion, which helps to better extract the feature information of small objects. In summary, our contributions are listed as follows.

- We develop a novel aerial object detector for exploring rotated and small targets (RSTDet), which takes into account the two tasks of rotating and small object, making it more suitable and practical for aerial object detection.
- We design a six-parameter rotating representation, which can regress an accurate rotation box with a small amount of calculation. It helps the network solve the problem of rotation target.
- We use bottom-up path augmentation to make the shallow features and deep features deeply integrated, and thus the network has a stronger ability to explore targets.
- In addition, the attention mechanism introduced also enables the network to extract key features, which is conducive to discovering difficult objects. The above strategies alleviate the problem of small object detection.

Our method was validated on the DOTA benchmark, which is one of the most commonly-used dataset for aerial object detection. The results well demonstrate that our approach achieves substantial gains over competing methods and verify the effectiveness of it.

Related Work

Compared with traditional object detection, the challenge of aerial object detection is the angle uncertainty. It's tricksy to calculate by the Region Proposal Network in the classic object detection method (Ren et al. 2017). Secondly, the detection task has the bottleneck of small object detection due to the objects in aerial images are generally small.

Oriented Object detection. There are many methods for rotating targets. These methods are used to determine the rotation angle of the detected target and calibrate it with a rotation box. RoI Transformer (Ding et al. 2019a) uses a decoder to convert HRoI (horizontal anchor) to RRoI (rotation anchor). Though it greatly improves the detection accuracy of rotating targets, the decoder with three fully connection layers introduces many parameters. R3Det (Yang et al. 2021b) and CAD-Net (**?**) apply multi parameter regression to get the rotation angle of the object. Xie et al. (E, P, and X 2020) acquires boundary contour through center point regression and polar coordinate regression. Later, Gliding Vertex (Xu et al. 2021a) and RSDet (Qian et al. 2019) innovatively use quadrangles to locate objects. However, these methods may cause discontinuous boundary problems. Yang et al. (Yang and Yan 2020) solves the problem of angle periodicity by introducing periodicity, which changes angle prediction from regression task to classification task. Recently, Yang et al. (Yang et al. 2022) proposes a simple and more efficient SkewIoU approximate loss for oriented object detection. Different from the traditional way of using angle regression, Oriented R-CNN (Xie et al. 2021) proposes a box representation of midpoint offset, which not only avoids a lot of calculation processes, but also provides constraints for bounding box regression, greatly improving detection performance.

Small Object Detection. As a branch of object detection, small object detection is specially used to detect small objects and has the great significance for the practical application of various scenarios. Because of the limited from small objects and scarcity of small object datasets, small object's detection accuracy and precision has a big gap with medium-sized object detection.In this context, some approaches have been proposed.They are data-manipulation methods, scale-aware methods, feature-fusion methods,superresolution methods, context-modeling methods, and other approaches.Meanwhile, in order to alleviate the scarcity of data, some data sets for small target detection have been proposed, such as SOD,TinyPerson and SODA.

- data-manipulation methods Small objects usually only account for a small part of the data set. A simple and effective method is increasing the number of small objects.Related works are Oversampling-based augmentation strategy (Kisantal et al. 2019) and Automatic augmentation scheme (Zoph et al. 2020).

- scale-aware methods The early object detection methods based on deep learning are difficult to detect small objects because they only use high-level features. To solve this problem, some papers propose Multi-scale detection in a divide-and-conquer fashion (Kong et al. 2016) and Tailored training schemes (Li et al. 2019b).

- feature-fusion methods Due to the existence of subsampling layers of deep CNN, the features of small objects may gradually disappear with the depth change. To solve this problem, the common methods fuse the features of different layers or branches, so as to obtain better feature representation of small objects (Shrivastava et al. 2016).

- superresolution methods Most super-resolution methods use generative adversarial network (GAN) (Goodfellow et al. 2014) to obtain high-quality feature representations that are beneficial for small target detection, while other methods choose parameterized upsampling operation to enhance features (Deng et al. 2022).

- context-modeling methods Prior knowledge that captures semantic or spatial associations is called context; for example, humans can use the relationship between objects and the environment or between objects to facilitate the recognition of objects and scenes (Torralba 2003). Information context can provide more decision support in identifying objects with poor viewing quality than the objects themselves. Therefore, the detection of small objects can be enhanced by using context cues.

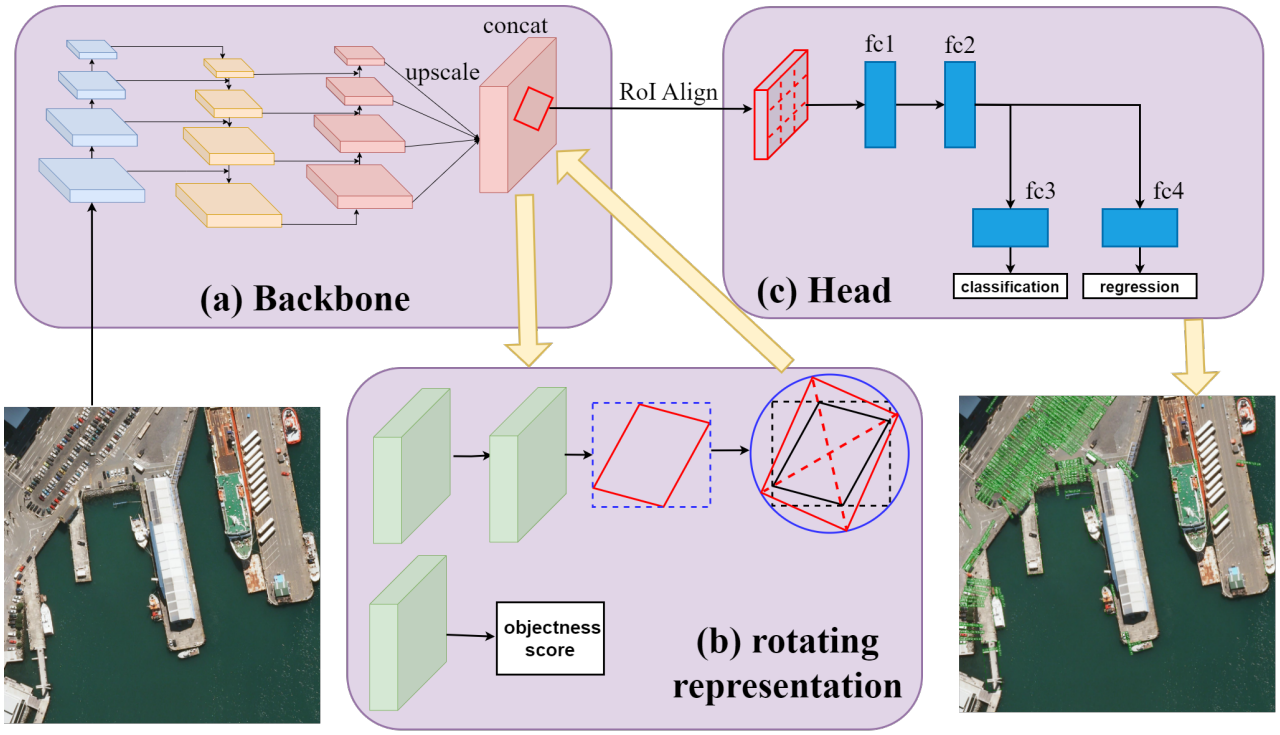

Figure 3: Overall framework of RSTDet. (a) The input image first passes through the backbone network with attention mechanism, which includes feature extraction, feature fusion, and bottom-up path augmentation. The obtained feature map is then sent to a two-stage detector. (b) In the first stage, six-parameter rotating representation gets the rotation bounding box. (c) In the second stage, the rotated RoI Align crops the feature and sends it to the fully connection layers in the second stage to complete the classification and regression tasks.

Proposed Solution

Framework Overview

As shown in Figure 3, after the image is input into the network, features at different levels are extracted through several convolution layers, and these features are sent to the feature pyramid network for feature fusion. These two parts both contain attention mechanisms, so that the network can better focus on effective features. Subsequently, bottom-up path augmentation further fuses the shallow features into the deep features. The obtained four level feature maps will be upsampled to the same size before concat. The merged feature map goes through two convolution branches, one of which gets the objectness score to evaluate the foreground and background score, and the other uses six-parameter rotating representation to get the rotation box. This rotation box is sent back to the feature map, and the features in it are cropped by rotated RoI Align and sent to the fully connection layers for classification and regression. The testing phase is similar to the training phase. The input image goes through the backbone network and the detector. After completing classification and regression tasks, the results are output.

Bottom-up Path Augmentation

In the neural network, the features extracted from the shallow layer are generally low-order, such as shape texture and so on, while those extracted from the high layer of the network are generally high-order features, such as whether there is hair or ears and so on. In the target detection, the shape texture features extracted from the bottom layer of the network are particularly critical for object positioning. The classical FPN aggregates the information of feature maps of different scales from the bottom to the top. Since the number of layers of the backbone feature extraction network is generally very deep, the information at the bottom will lose a lot of information after transmission. We argue that it is unfavorable for small object detection.

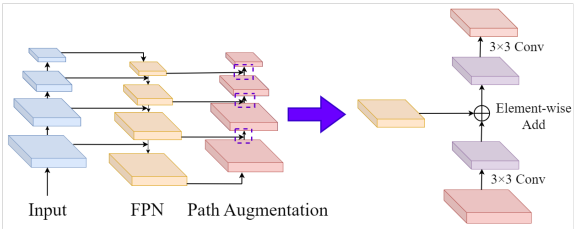

Figure 4: Illustration of bottom-up path augmentation.

In response to this issue, we add a bottom-up path augmentation module to our model. As shown in Figure 4, it consists of four feature layers similar to FPN. The high-resolution features of the bottom layer pass through a 3 ×

3 convolutional layer with stride 2 to reduce the spatial size. Then the features with the upper level features from FPN are fused by element-wise add. They are further fused by a 3 × 3 convolution layer. The above process is repeated to obtain features at all levels. The bottom-up path augmentation uses a few layers and preserves the beneficial features of various textures, shapes and other shallow layers, which is crucial for small object detection.

Six-parameter Rotating Representation

Due to the different rotation angles of objects, the traditional object detection methods can not be satisfied. We adopt the midpoint offset method proposed in Oriented R-CNN (Xie et al. 2021), and propose a novel six-parameter rotating representation. This method obtains the rotating object proposal by adding the horizontal box outside the rotating object proposal and offset parameters, which greatly improve the efficiency of anchor generation and improve the algorithm performance.

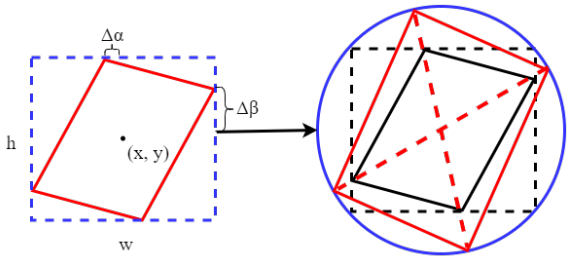

Figure 5: The six-parameter rotating representation diagram.

As shown in Figure 5, the angled object is represented by $O = (x, y, w, h, \Delta\alpha, \Delta\beta)$, where (x, y) represents the object's center coordinate, (w, h) represents the width and height of the external horizontal proposal of the oriented proposal where the object is located, and $(\Delta\alpha, \Delta\beta)$ represents the midpoint offset of the object proposal and the external horizontal proposal. In this way, on the basis of the original horizontal proposals regression, the regression of any two adjacent sides of the midpoint offset can realize the generation of oriented proposals. By decoding these 6 parameters, we get the coordinate set of the object representing $V = (v_1, v_2, v_3, v_4)$.The details are as follows:

$$\begin{cases} v_1 = (x, y - h/2) + (\Delta\alpha, 0) = (x + \Delta\alpha, y - h/2) \\ v_2 = (x + w/2, y) + (0, \Delta\beta) = (x + w/2, y + \Delta\beta) \\ v_3 = (x, y + h/2) + (-\Delta\alpha, 0) = (x - \Delta\alpha, y + h/2) \\ v_4 = (x - w/2, y) + (0, -\Delta\beta) = (x - w/2, y - \Delta\beta) \end{cases}$$

After obtaining the vertices set V of the oriented proposal through the midpoint offset method, we further adopt the operation of taking points from the circumscribed circle. This is because although the midpoint offset representation determines the rotation angle of the object, the constraints of the bounding rectangle usually make the bounding box too small. Based on this observation, we extended the diagonal of the proposal to the position of the circumscribed circle, effectively alleviating this problem.

With the generation of rotation proposals, they are sent to the feature map to crop object features. The rotated RoI alignment (Ding et al. 2019b) is applied to finish this work. The crop features are then used as input to the detection head, which is consist of four fully connection layers using for box classification and regression. The detection head is constructed on the basis of the original Faster R-CNN (Ren et al. 2017) realized by adding Angle regression parameters.

Attention Mechanism

The features of small objects are more likely to be lost in the process of down sampling, resulting in difficult detection. In order to alleviate this problem, we add attention mechanism to feature extraction and feature fusion, including channel attention and spatial attention. This enhances the feature information.

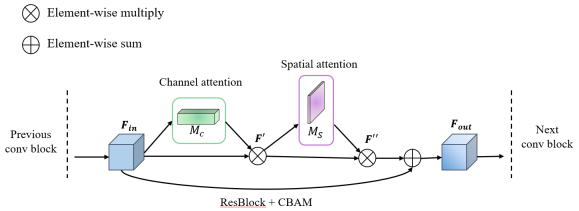

Figure 6: The CBAM module embedded in RSTDet.

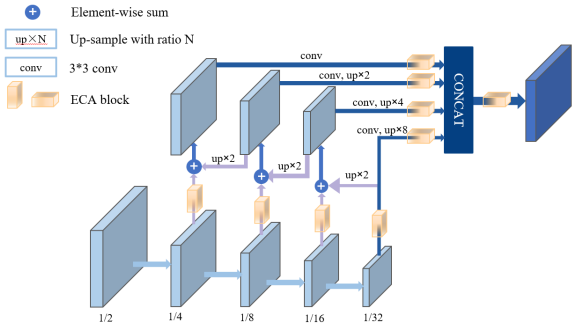

Figure 7: The ECA module embedded in RSTDet.

In the feature extraction stage, the network not only needs to pay attention to the spatial position of the tiny cues in the image but also needs to choose different feature information among different channels. Therefore, the mixed domain attention mechanism can be considered. In this regard, the CBAM module is integrated into each residual block in resnet. The embedding mode is shown in Figure 6. The output of each layer of resnet is adjusted by attention, for the purpose of extract detailed features.

In the feature fusion stage, the information of different levels is exchanged and fused. The channel attention mechanism can make the fusion process pay more attention to some special channels. To this end, the ECA module is added before and after feature fusion. For the fused feature map, we also add an ECA module in order to induce the model focus on features conducive to detection in the subsequent stage.

Method	Backbone	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
One-stage																	
RetinaNet-O†	R-50-FPN	88.67	77.62	41.81	58.17	74.58	71.64	79.11	90.29	82.18	74.32	54.75	60.60	62.57	69.67	60.64	68.43
DRN (Pan et al. 2020)	H-104	88.91	80.22	43.52	63.35	73.48	70.69	84.94	90.14	83.85	84.11	50.12	58.41	67.62	68.60	52.50	70.70
R3Det (Yang et al. 2021a)	R-101-FPN	88.76	83.09	50.91	67.27	76.23	80.39	86.72	90.78	84.68	83.24	61.98	61.35	66.91	70.63	53.94	73.79
PIoU (Chen et al. 2020)	DLA-34	80.90	69.70	24.10	60.20	38.30	64.40	64.80	90.90	77.20	70.40	46.50	37.10	57.10	61.90	64.00	60.50
RSDet (Qian et al. 2021)	R-101-FPN	89.80	82.90	48.60	65.20	69.50	70.10	70.20	90.50	85.60	83.40	62.50	63.90	65.60	67.20	68.00	72.20
DAL (Ming et al. 2020)	R-50-FPN	88.68	76.55	45.08	66.80	67.00	76.76	79.74	90.84	79.54	78.45	57.71	62.27	69.05	73.14	60.11	71.44
S^2ANet (Han et al. 2021)	R-50-FPN	89.11	82.84	48.37	71.11	78.11	78.39	87.25	90.83	84.90	85.64	60.36	62.60	65.26	69.13	57.94	74.12
Two-stage																	
ICN (Azimi et al. 2018)	R-101-FPN	81.36	74.30	47.70	70.32	64.89	67.82	69.98	90.76	79.06	78.20	53.64	62.90	67.02	64.17	50.23	68.16
Faster R-CNN-O†	R-50-FPN	88.44	73.06	44.86	59.09	73.25	71.49	77.11	90.84	78.94	83.90	48.59	62.95	62.18	64.91	56.18	69.05
CAD-Net (Zhang et al. 2019)	R-101-FPN	87.80	82.40	49.40	73.50	71.10	63.50	76.60	90.90	79.20	73.30	48.40	60.90	62.00	67.00	62.20	69.90
RoI Transformer (Ding et al. 2019b)	R-101-FPN	88.64	78.52	43.44	75.92	68.81	73.68	83.59	90.74	77.27	81.46	58.39	53.54	62.83	58.93	47.67	69.56
SCRDet (Yang et al. 2019)	R-101-FPN	89.98	80.65	52.09	68.36	68.36	60.32	72.41	90.85	**87.94**	86.86	65.02	66.68	66.25	68.24	65.21	72.61
RoI Transformer+	R-50-FPN	88.65	82.60	52.53	70.87	77.93	76.67	86.87	90.71	83.83	82.51	53.95	67.61	74.67	68.75	61.03	74.61
Gliding Vertex (Xu et al. 2021b)	R-101-FPN	89.64	85.00	52.26	77.34	73.01	73.14	86.82	90.74	79.02	86.81	59.55	70.91	72.94	70.86	57.32	75.02
FAOD (Li et al. 2019a)	R-101-FPN	90.21	79.58	45.49	76.41	73.18	68.27	79.56	90.83	83.40	84.68	53.40	65.42	74.17	69.69	64.86	73.28
CenterMap-Net (Wang et al. 2021)	R-50-FPN	88.88	81.24	53.15	60.65	78.62	66.55	78.10	88.83	77.80	83.61	49.36	66.19	72.10	72.36	58.70	71.74
FR-Est (Fu et al. 2021)	R-101-FPN	89.63	81.17	50.44	70.19	73.52	77.98	86.44	90.82	84.13	83.56	60.64	66.59	70.59	66.72	60.55	74.20
Mask OBB (Wang et al. 2019)	R-50-FPN	89.61	85.09	51.85	72.90	75.28	73.23	85.57	90.37	82.08	85.05	55.73	68.39	71.61	69.87	66.33	74.86
Oriented R-CNN	R-50-FPN	89.46	82.12	54.78	70.86	78.93	83.00	88.20	90.90	87.50	84.68	63.97	67.69	74.94	68.84	52.28	75.87
Oriented R-CNN	R-101-FPN	88.86	83.48	55.27	76.92	74.27	82.10	87.52	90.90	85.56	85.33	65.51	66.82	74.36	70.15	57.28	76.28
Ours																	
RSTDet	R-50-PAFPN	89.48	78.73	54.89	73.44	78.93	82.40	88.26	90.90	86.69	84.61	64.69	67.30	75.67	68.64	58.17	76.48
RSTDet	R-101-PAFPN	89.64	83.82	55.74	72.62	78.47	83.76	88.08	90.90	86.70	84.99	64.48	67.04	76.22	70.88	56.91	76.68
RSTDet‡	R-50-PAFPN	90.19	85.29	60.88	**80.37**	**80.02**	85.14	88.60	**90.90**	86.32	**87.83**	**71.39**	71.01	81.77	**79.23**	**74.33**	**80.88**
RSTDet‡	R-101-PAFPN	**90.33**	**85.99**	**62.13**	78.57	78.84	**85.26**	**88.66**	90.87	86.80	87.75	70.27	**71.19**	**82.95**	76.02	72.36	80.53

Table 1: Comparison with state-of-the-art methods on the DOTA dataset. † means the results from AerialDetection. ‡ denotes multi-scale training and testing. Bold indicates the best result, and underline indicates the best result without multi-scale training and testing.

Experiments

Dataset

We used the DATA set (Xia et al. 2018) released by Wuhan University in 2017. Classical data sets such as VOC and COCO contain relatively few small targets, and the marked objects are also horizontal due to gravity. In contrast, the DOTA dataset uses aerial or satellite images collected by aircraft, helicopters and drones, and the marking boxes used in the DOTA dataset are rotatable rather than horizontal. There are 2806 target images in the DOTA dataset, with resolution ranging from 800 * 800 to 4000*4000. It's about 35 GB in size and contains 15 categories. These categories include objects of all sizes and shapes.

Experimental Setting

Our experiments were carried out on a single RTX 1080Ti, using mmdetection framework. The optimizer uses SGD, where the weight attenuation and momentum of SGD are 0.0001 and 0.9, respectively. For DOTA dataset preprocessing, we cut the image into 1024*1024 squares, and scaled the original image to 0.5, 1.0 and 1.5 times of the original, and then obtained 1024*1024 squares by 524 step size. This is a general preprocessing operation. In the training phase, the learning rate starts at 0.005 and will gradually decrease with the increase of epoch. The number of epochs is 12 in total, and the NMS threshold is 0.1 when integrating image blocks. We used ResNet50 and ResNet101 as the feature extraction network of the model, and test the performance on the test set of the DOTA dataset.

Experimental Result

We compare the proposed method with 19 state-of-the-art methods on DOTA dataset. Table 1 reports the results. The proposed RSTDet reaches impressive results, which exceed all methods. Specifically, without multi-scale training and testing strategies, our method reaches 76.48% mAP with resnet50 and 76.68% mAP with resnet101, outperforming than other methods. Adding multi-scale trick makes the results turn to 80.88% mAP and mAP, respectively.

Conclusion

In this paper, we propose RSTDet, a unified object detection method that explores the rotated and small targets. We first design a six-parameter rotating representation to detect rotating targets in a low-cost and effective way. We also add the bottom-up path augmentation after the feature pyramid network, which makes better use of shallow features. The attention mechanism is further introduced to make the network focus on the important features of channel and spatial domain, which improves the performance of detection to some extent. Experimental results show that our method has competitive performance to the current state of the art methods.

References

Azimi, S. M.; Vig, E.; Bahmanyar, R.; Körner, M.; and Reinartz, P. 2018. Towards multi-class object detection in unconstrained remote sensing imagery. In *Proceedings of the Asian Conference on Computer Vision*, 150–165.

Chen, Z.; Chen, K.; Lin, W.; See, J.; Yu, H.; Ke, Y.; and Yang, C. 2020. PIoU Loss: Towards Accurate Oriented Object Detection in Complex Environments. In *Proceedings of the European Conference on Computer Vision*, 195–211.

Cheng, G.; Yuan, X.; Yao, X.; Yan, K.; Zeng, Q.; and Han, J. 2022. Towards large-scale small object detection: Survey and benchmarks. *arXiv preprint arXiv:2207.14096*.

Deng, C.; Wang, M.; Liu, L.; and Liu, Y. 2022. Extended Feature Pyramid Network for Small Object Detection. *IEEE Transactions on Multimedia*, 24: 1968–1979.

Ding, J.; Xue, N.; Long, Y.; Xia, G.-S.; and Lu, Q. 2019a. Learning RoI Transformer for Oriented Object Detection in Aerial Images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2844–2853.

Ding, J.; Xue, N.; Long, Y.; Xia, G.-S.; and Lu, Q. 2019b. Learning roi transformer for oriented object detection in aerial images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2849–2858.

E, X.; P, S.; and X, S. 2020. PolarMask: Single Shot Instance Segmentation with Polar Representation. *arxiv.org/abs/1909.13226*.

Fu, K.; Chang, Z.; Zhang, Y.; and Sun, X. 2021. Point-Based Estimator for Arbitrary-Oriented Object Detection in Aerial Images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(5): 4370–4387.

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, 2672–2680. Cambridge, MA, USA: MIT Press.

Han, J.; Ding, J.; Li, J.; and Xia, G. S. 2021. Align Deep Features for Oriented Object Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 1–11.

Kisantal, M.; Wojna, Z.; Murawski, J.; Naruniec, J.; and Cho, K. 2019. Augmentation for small object detection. *arXiv preprint arXiv:1902.07296*.

Kong, T.; Yao, A.; Chen, Y.; and Sun, F. 2016. Hypernet: Towards accurate region proposal generation and joint object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 845–853.

Li, C.; Xu, C.; Cui, Z.; Wang, D.; Zhang, T.; and Yang, J. 2019a. Feature-attentioned object detection in remote sensing imagery. In *Proceedings of the IEEE International Conference on Image Processing*, 3886–3890.

Li, W.; Chen, Y.; Hu, K.; and Zhu, J. 2022. Oriented reppoints for aerial object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1829–1838.

Li, Y.; Chen, Y.; Wang, N.; and Zhang, Z. 2019b. Scale-aware trident networks for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6054–6063.

Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; and Belongie, S. 2017a. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125.

Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; and Dollár, P. 2017b. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

Liu, S.; Qi, L.; Qin, H.; Shi, J.; and Jia, J. 2018. Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8759–8768.

Ma, J.; Shao, W.; Ye, H.; Wang, L.; Wang, H.; Zheng, Y.; and Xue, X. 2018. Arbitrary-Oriented Scene Text Detection via Rotation Proposals. *IEEE Transactions on Multimedia*, 20(11): 3111–3122.

Ming, Q.; Zhou, Z.; Miao, L.; Zhang, H.; and Li, L. 2020. Dynamic Anchor Learning for Arbitrary-Oriented Object Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Pan, X.; Ren, Y.; Sheng, K.; Dong, W.; Yuan, H.; Guo, X.; Ma, C.; and Xu, C. 2020. Dynamic refinement network for oriented and densely packed object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 11207–11216.

Qian, W.; Yang, X.; Peng, S.; Guo, Y.; and Yan, J. 2019. Learning Modulated Loss for Rotated Object Detection.

Qian, W.; Yang, X.; Peng, S.; Guo, Y.; and Yan, J. 2021. Learning modulated loss for rotated object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137–1149.

Shrivastava, A.; Sukthankar, R.; Malik, J.; and Gupta, A. 2016. Beyond skip connections: Top-down modulation for object detection. *arXiv preprint arXiv:1612.06851*.

Torralba, A. 2003. Contextual priming for object detection. *International journal of computer vision*, 53(2): 169–191.

Wang, J.; Ding, J.; Guo, H.; Cheng, W.; Pan, T.; and Yang, W. 2019. Mask OBB: A semantic attention-based mask oriented bounding box representation for multi-category object detection in aerial images. *Remote Sensing*, 11(24): 2930.

Wang, J.; Yang, W.; Li, H.-C.; Zhang, H.; and Xia, G.-S. 2021. Learning Center Probability Map for Detecting Objects in Aerial Images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(5): 4307–4323.

Xia, G.-S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; and Zhang, L. 2018. DOTA: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3974–3983.

Xie, X.; Cheng, G.; Wang, J.; Yao, X.; and Han, J. 2021. Oriented R-CNN for Object Detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 3500–3509.

Xu, Y.; Fu, M.; Wang, Q.; Wang, Y.; Chen, K.; Xia, G.-S.; and Bai, X. 2020. Gliding vertex on the horizontal bounding box for multi-oriented object detection. *IEEE transactions on pattern analysis and machine intelligence*, 43(4): 1452–1459.

Xu, Y.; Fu, M.; Wang, Q.; Wang, Y.; Chen, K.; Xia, G.-S.; and Bai, X. 2021a. Gliding Vertex on the Horizontal Bounding Box for Multi-Oriented Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(4): 1452–1459.

Xu, Y.; Fu, M.; Wang, Q.; Wang, Y.; Chen, K.; Xia, G.-S.; and Bai, X. 2021b. Gliding Vertex on the Horizontal Bounding Box for Multi-Oriented Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(4): 1452–1459.

Yang, X.; Liu, Q.; Yan, J.; Li, A.; Zhang, Z.; and Yu, G. 2021a. R3Det: Refined single-stage detector with feature refinement for rotating object. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Yang, X.; and Yan, J. 2020. Arbitrary-Oriented Object Detection with Circular Smooth Label. In Vedaldi, A.; Bischof, H.; Brox, T.; and Frahm, J.-M., eds., *Computer Vision – ECCV 2020*, 677–694. Cham: Springer International Publishing. ISBN 978-3-030-58598-3.

Yang, X.; Yan, J.; Feng, Z.; and He, T. 2021b. R3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(4): 3163–3171.

Yang, X.; Yang, J.; Yan, J.; Zhang, Y.; Zhang, T.; Guo, Z.; Sun, X.; and Fu, K. 2019. SCRDet: Towards more robust detection for small, cluttered and rotated objects. In *Proceedings of the IEEE International Conference on Computer Vision*, 8232–8241.

Yang, X.; Zhou, Y.; Zhang, G.; Yang, J.; Wang, W.; Yan, J.; Zhang, X.; and Tian, Q. 2022. The KFIoU Loss for Rotated Object Detection. *arXiv preprint arXiv:2201.12558*.

Zhang, G.; Lu, S.; Zhang, W.; and Liu, X. 2019. CAD-Net: A context-aware detection network for objects in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12): 10015–10024.

Zoph, B.; Cubuk, E. D.; Ghiasi, G.; Lin, T.-Y.; Shlens, J.; and Le, Q. V. 2020. Learning data augmentation strategies for object detection. In *European conference on computer vision*, 566–583. Springer.