ABSTRACT
We present an open-source toolbox, named MMRotate, which provides a coherent algorithm framework of training, inferring, and evaluation for the popular rotated object detection algorithm based on deep learning. MMRotate implements 18 state-of-the-art algorithms and supports the three most frequently used angle definition methods. To facilitate future research and industrial applications of rotated object detection-related problems, we also provide a large number of trained models and detailed benchmarks to give insights into the performance of rotated object detection. MMRotate is publicly released at https://github.com/open-mmlab/mmrotate.

KEYWORDS
open source; rotation detection; oriented object detection

1 INTRODUCTION
In recent years, deep learning has achieved tremendous success in fundamental computer vision applications such as image recognition [8], object detection [4, 16, 25, 26, 28] and image segmentation [7, 19]. In light of this, deep learning has also been applied to areas such as faces detection [27], text detection [12, 13, 17, 20, 45] and aerial images detection [32, 35, 38]. In these object detection tasks, oriented bounding boxes (OBBs) are widely used instead of horizontal bounding boxes (HBBs) because they can better align the objects for more accurate identification. This kind of special object detection is called rotated object detection, also known as arbitrary-oriented object detection. In addition to the three applications mentioned above, rotated object detection is also widely used in 3D objects detection [44] and retail scenes detection [2, 23].

Different approaches utilize different angle definition methods, optimization strategies (e.g., optimizers, learning rate schedules,
epoch numbers, and data augmentation pipelines), and CUDA operators (e.g. IoU and NMS for OBBs). To encompass the diversity of components used in various models, we have proposed the MM-Rotate toolbox covering recent popular rotated object detection approaches in a unified framework. The toolbox now implements 18 rotated object detection methods, 10 CUDA speed-up operators, and 12 losses. Integrating various algorithms allows code reusability and therefore dramatically simplifies the implementation of algorithms. Moreover, the unified framework allows different approaches to be compared against each other fairly and that their key effective components can be easily investigated.

MMRotate is hosted on GitHub under the Apache-2.0 License. The repository contains the compressed archive file of software and documentation, including installation instructions, dataset preparation scripts, API documentation, model zoo, tutorials and user manual. MMRotate re-implements 18 state-of-the-art rotated object detection algorithms and provides extensive benchmarks and models trained on popular academic datasets. In addition to (distributed) training and testing scripts, it offers a rich set of utility tools covering visualization and demonstration.

2 RELATED WORK

Text detection. Text detection aims to localize the bounding boxes of text instances. Recent research focus has shifted to challenging arbitrary-shaped text detection [7, 12, 14, 19, 20]. R2CNN [12] simultaneously predicts the axis-aligned and inclined boxes by adding an inclined box branch and uses an inclined NMS to obtain the detection results. While Mask R-CNN [7] can be used to detect texts, it might fail to detect curved and dense texts due to the rectangle-based ROI proposals. On the other hand, RPN [20] proposes a Rotation RPN to generate inclined proposals with text orientation angle information and project arbitrary-oriented proposals to the feature map with Rotation ROI pooling. TextSnake [19] describes text instances with a series of ordered, overlapping disks.

Aerial image detection. Aerial image detection plays a vital role in the military and attracts more and more attention in civilian field [3, 33, 34, 37]. It aims to predict more accurate bounding boxes and preserve the direction information of the object on aerial images (including ship, plane, vehicles, bridge, etc.). Although rotated object detection provides more accurate prediction results than horizontal detection, it requires defining a new bounding box representation. The most common is the $\theta$-based representation $(x, y, w, h, \theta)$, and it adds an extra angle parameter based on the horizontal box. Depending on the angle range, it can be divided into OpenCV definition $(D_{oc}, \theta \in [-\pi/2, 0])$ [32, 36, 38], long edge 90$°$ definition $(D_{le90}, \theta \in [-\pi/2, \pi/2])$ [3, 6], and long edge 135$°$ definition $(D_{le135}, \theta \in [-\pi/4, 3\pi/4])$ [5]. Recent works used two-dimensional Gaussian distribution [37, 39] and point sets [22–24, 31] to represent objects, which have achieved excellent results. Feature alignment is another research direction of lifting rotated object detection performance. R3Det [35] proposes a feature refinement module to re-construct the feature map based on the refined bounding box output from the previous stage. S2A-Net [5] proposes an alignment convolution to alleviate the misalignment between axis-aligned convolutional features and arbitrary oriented objects. Recently, ReDet [6] began to study a novel rotation-equivariant RoI Align to produce rotation-equivariant features. Label Assignment is also a research hotspot. DAL [22] reconsideres whether IoU is a truly credible division basis and defines a new matching degree. SLA [21] proposes a sparse label assignment strategy to achieve training sample selection based on posterior IoU distribution. SASM [9] proposes two novel shape-adaptive strategies which can dynamically select samples and measure the quality of positive samples.

Open source toolbox. Several open source rotated object detection toolboxes have been developed over the years to meet the increasing demand from both academia and industry.

AerialDetection is the pioneer of deep learning-based open source rotated object detection toolbox. It was publicly released in 2019, and provides evaluation tools for DOTA data set [29]. It supports five rotated object detection methods, e.g., RetinaNet [15], Faster R-CNN [26] and RoI Trans [3]. However, it has not been maintained anymore. OBBDetection [40] has been released recently. JDet is another open source aerial image object detection toolbox based on TensorFlow. Nevertheless, they cannot enjoy the latest technology provided by MM Detection, since they rely on a specific version of MM Detection. JDet is an open source aerial image object detection toolbox based on a high-performance deep learning library [11]. It can be deployed on multiple platforms such as Linux and Windows, and the 8 detectors it reproduces have faster inference speed than PyTorch. TensorFlow-based rotated object detection toolbox AlphaRotate [2] [40] has been released recently. It currently implements 18 rotated object detection methods, including the algorithms that PyTorch and Jitter do not support, e.g., GWD [37], KLD [39], and R3Det [35]. Comprehensive comparisons among these open source toolboxes are given in Table 1.

3 ROTATED OBJECT DETECTION STUDIES

Many important factors can affect the performance of deep learning-based detectors. This section investigates the angle definition method, backbones, and loss of network architectures. We exchange the

---

1. https://github.com/dingjianw/101/AerialDetection
2. https://github.com/jwanger997/Observer
3. https://github.com/open-mmlab/mmdetection
4. https://github.com/Jittor/JDet
5. https://github.com/yangxue0827/RotationDetection

| Benchmark | AerialDet | JDet | OBBDet | AlphaRotate | MMRotate |
|-----------|-----------|------|--------|-------------|----------|
| DL library | PyTorch   | Jitter | PyTorch | TensorFlow | PyTorch |
| Inference engine | PyTorch   | Jitter | PyTorch | TensorFlow | PyTorch |
| OS          | Linux     | Windows | Windows | Linux       | Windows |
| Algorithm   | 5         | 8     | 9      | 16          | 18       |
| Dataset     | 1         | 4     | 5      | 11          | 4        |
| Doc.        | -         | -     | -      | -           | ✓        |
| Easy install| -         | -     | -      | -           | ✓        |
| Maint        | -         | ✓     | ✓      | ✓           | ✓        |

Table 1: Open source rotated object detection benchmarks.
Table 2: Accuracy comparison of rotated object detectors on DOTA v1.0. MS means multiple scale image split. RR means random rotation. All models are inferred with one 2080Ti GPU.

| Baseline | Technique | fp16 | Box Def. | Lr sched. | Mem.(GB) | Inf. time (fps) | Aug. | mAP |
|----------|-----------|------|----------|-----------|----------|----------------|------|-----|
| RetinaNet-H [15] | - | oc | 1x | 3.38 | 15.7 | - | 64.55 |
| RetinaNet-H [15] | GWD [37] | - | oc | 1x | 3.39 | 15.5 | - | 69.55 |
| RetinaNet-H [15] | KFlou [41] | - | le90 | 1x | 3.38 | 15.1 | - | 69.60 |
| RetinaNet-H [15] | KFlou [41] | - | oc | 1x | 3.39 | 15.6 | - | 69.76 |
| RetinaNet-H [15] | KFlou [41] | - | le135 | 1x | 3.38 | 15.3 | - | 69.77 |
| RetinaNet-H [15] | KLD [39] | - | oc | 1x | 3.39 | 15.6 | - | 69.94 |
| RetinaNet-O [15] | - | le90 | 1x | 3.38 | 16.9 | - | 68.42 |
| RetinaNet-O [15] | GSL [33] | ✓ | le90 | 1x | 2.36 | 22.4 | - | 68.79 |
| RetinaNet-O [15] | KLD [39] | ✓ | le90 | 1x | 3.35 | 16.9 | - | 70.22 |
| RetinaNet-O [15] | KLD [39] | ✓ | le135 | 1x | 3.38 | 17.2 | - | 69.79 |
| RetinaNet-O [15] | ATSS [45] | - | le90 | 1x | 3.12 | 18.2 | - | 70.64 |
| RetinaNet-O [15] | ATSS [45] | - | le135 | 1x | 3.19 | 18.8 | - | 72.29 |
| RepPoints [42] | - | oc | 1x | 3.45 | 15.6 | - | 59.44 |
| RepPoints [42] | SASM [9] | - | oc | 1x | 3.53 | 15.7 | - | 66.45 |
| RepPoints [42] | G-Rep [10] | - | le135 | 1x | 4.05 | 8.6 | - | 69.49 |
| RepPoints [42] | G-Rep [10] | - | le135 | 1x | 3.45 | 16.1 | - | 69.63 |
| RepPoints [42] | G-Rep [10] | - | oc | 40e | 3.45 | 16.1 | - | 73.45 |
| FCOS [28] | - | le90 | 1x | 4.18 | 20.9 | - | 70.70 |
| FCOS [28] | CSL [33] | - | le90 | 1x | 4.23 | 20.7 | - | 71.76 |
| FCOS [28] | CSL [33] | - | le90 | 1x | 4.28 | 20.7 | - | 71.89 |
| FCOS [28] | KLD [39] | - | le90 | 1x | 3.54 | 12.4 | - | 69.80 |
| FCOS [28] | KLD [39] | - | oc | 1x | 3.65 | 13.6 | - | 70.54 |
| FCOS [28] | KLD [39] | - | oc | 1x | 3.54 | 12.4 | - | 71.83 |
| FCOS [28] | KLD [39] | - | oc | 1x | 3.62 | 12.2 | - | 72.68 |
| R3Det [35] | ATSS [45] | - | oc | 1x | 3.54 | 12.4 | - | 69.80 |
| R3Det [35] | KLD [39] | - | oc | 1x | 3.54 | 12.4 | - | 71.83 |
| R3Det [35] | KLD [39] | - | oc | 1x | 3.62 | 12.2 | - | 72.68 |
| R3Det [35] | S2ANet [5] | - | le135 | 1x | 3.14 | 15.5 | - | 73.91 |
| R3Det [35] | S2ANet [5] | - | ✓ | le135 | 1x | 2.17 | 17.4 | - | 74.19 |
| Faster RCNN [26] | Gliding Vertex [31] | - | le90 | 1x | 8.46 | 16.1 | - | 73.40 |
| Faster RCNN [26] | Oriented RCNN [30] | - | le90 | 1x | 8.46 | 16.4 | - | 73.23 |
| Faster RCNN [26] | RoI Trans. [3] | ✓ | le90 | 1x | 7.71 | 13.3 | - | 75.99 |
| Faster RCNN [26] | RoI Trans. [3] | ✓ | le90 | 1x | 8.67 | 14.4 | - | 76.08 |
| Faster RCNN [26] | RoI Trans. [3] + Swin-T [18] | - | le90 | 1x | 9.32 | 10.9 | - | 76.68 |
| Faster RCNN [26] | RoI Trans. [3] + Swin-T [18] | - | le90 | 1x | 9.23 | 10.9 | - | 77.51 |
| Faster RCNN [26] | RoI Trans. [3] + KLD [39] + Swin-T [18] | - | le90 | 1x | 8.96 | 14.4 | - | 79.66 |

Angle definition method. OpenCV definition, long edge 90° definition, and long edge 135° definition are all supported by MMRotate. All rotated object detection algorithms can easily switch between these three angle definition methods by modifying the configuration file. Meanwhile, we established an angle conversion API to facilitate other angle definition methods.

Backbone. ResNet50 [8] are commonly used in object detection approaches. To improve the accuracy, we also introduce a transformer-based backbone Swin transformer [18]. Table 2 compares ResNet50 and Swin-T in terms of memory, inference time and mAP by plugging them in RoI Trans. It has been shown that Swin-T significantly outperforms ResNet50, although its inference speed is 21% slower than that of ResNet50.

Loss. GWG, KLD, and KFlou propose different loss to train the rotated object detector. Our experimental results in Table 2 show that the KLD loss achieved the best mAP in the OpenCV definition method when using RetinaNet as the baseline. However, when using the R3Det as the baseline, the KFlou loss achieved the best mAP in the OpenCV definition method.

Mixed precision training and Useful tools. All detectors in MMRotate support mixed precision training. Our experimental results in Table 2 show that the model trained with fp16 has a similar mAP as the original model. MMRotate also provides a range of efficient and convenient tools (including visualization, confusion matrix analysis, huge image inference), allowing researchers to focus on the rotated object detection algorithm itself.

4 CONCLUSIONS

With the practical importance and academic emergence for visual rotation detection, MMRotate is a deep learning benchmark for visual object rotation detection in PyTorch under the Apache-2.0 license. The architecture is designated for flexibility and ease of use to facilitate the deployment of rotated object detection in diverse domains, both in industrial applications and academic research. We will continue to improve the entire optimized benchmark and support representative detection methods in the future. We also welcome the community to participate in the development.

above-mentioned components between different rotated object detection approaches to measure the performance, memory usage, and inference time.

Angle definition method. OpenCV definition, long edge 90° definition, and long edge 135° definition are all supported by MMRotate. All rotated object detection algorithms can easily switch between these three angle definition methods by modifying the configuration file. Meanwhile, we established an angle conversion API to facilitate other angle definition methods.

Backbone. ResNet50 [8] are commonly used in object detection approaches. To improve the accuracy, we also introduce a transformer-based backbone Swin transformer [18]. Table 2 compares ResNet50 and Swin-T in terms of memory, inference time and mAP by plugging them in RoI Trans. It has been shown that Swin-T significantly outperforms ResNet50, although its inference speed is 21% slower than that of ResNet50.

Loss. GWG, KLD, and KFlou propose different loss to train the rotated object detector. Our experimental results in Table 2 show that the KLD loss achieved the best mAP in the OpenCV definition method when using RetinaNet as the baseline. However, when using the R3Det as the baseline, the KFlou loss achieved the best mAP in the OpenCV definition method.

Mixed precision training and Useful tools. All detectors in MMRotate support mixed precision training. Our experimental results in Table 2 show that the model trained with fp16 has a similar mAP as the original model. MMRotate also provides a range of efficient and convenient tools (including visualization, confusion matrix analysis, huge image inference), allowing researchers to focus on the rotated object detection algorithm itself.

4 CONCLUSIONS

With the practical importance and academic emergence for visual rotation detection, MMRotate is a deep learning benchmark for visual object rotation detection in PyTorch under the Apache-2.0 license. The architecture is designated for flexibility and ease of use to facilitate the deployment of rotated object detection in diverse domains, both in industrial applications and academic research. We will continue to improve the entire optimized benchmark and support representative detection methods in the future. We also welcome the community to participate in the development.
ACKNOWLEDGMENTS

This work was supported by National Key Research and Development Program of China (2020AAA0107600), and the National Natural Science Foundation of China (61971279, 61972250, 62022054).

REFERENCES

[1] Kai Chen, Jiaqi Wang, Jiawei Shang, Mingtao Pang, Youyang Cao, Yu Xiong, Xiaoxiao Li, Shiyuan Sun, Wanqing Feng, Zowei Liu, Junrui Cheng, Dacheng Chen, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyi Li, Xin Lu, Rui Zhu, Yue Wu, Jiefeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. 2018. MM: Detection Open MMLab Detection Toolbox and Benchmark. arXiv preprint arXiv:1806.07155 (2019).

[2] Zhiheng Chen, Kean Chen, Weiyao Lin, John See, Hui Yu, Yan Ke, and Cong Yang. 2020. Piou loss: Towards accurate oriented object detection in complex environments. In European Conference on Computer Vision. Springer, 195–211.

[3] Jian Ding, Nan Xue, Yang Long, Guo-Song Xia, and Qikai Lu. 2019. Learning Rot Transformer for Oriented Object Detection in Aerial Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2849–2858.

[4] Ross Girshick. 2015. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision. 1440–1448.

[5] Jianming Han, Jian Ding, Jie Li, and Gui-Song Xia. 2021. Align deep features for oriented object detection. IEEE Transactions on Geoscience and Remote Sensing 59, 11 (2021), 7910–7923.

[6] Jianming Han, Jian Ding, Nan Xue, and Gui-Song Xia. 2021. Retd: A rotation-equivariant detector for aerial object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2786–2795.

[7] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision. 2961–2969.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 770–778.

[9] Liping Hou, Ke Lu, Jian Xue, and Yuqiu Li. 2022. Shape-Adaptive Selection and Measurement for Oriented Object Detection. In Proceedings of the AAAI Conference on Artificial Intelligence.

[10] Liping Hou, Ke Lu, Xue Yang, Yuqiu Li, and Jian Xue. 2022. G-Rep Gaussian Representation for Arbitrary-Oriented Object Detection. arXiv preprint arXiv:2205.11796 (2022).

[11] Shi-Min Hu, Dun Liang, Guo-Ye Yang, Gao-Wei Yang, and Wen-Yang Zhou. 2020. Jittor: a novel deep learning framework with meta-operators and unified graph execution. Science China Information Sciences 63, 22 (2020), 1–21.

[12] Yingying Jiang, Xiangyu Zhu, Xiaobing Wang, Shuli Yang, Wei Li, Hua Wang, Po Fu, and Zhenhao Luo. 2017. R2CNN:rotational region CNN for orientation robust scene text detection. arXiv preprint arXiv:1706.09579 (2017).

[13] Minghui Liao, Baoguang Shi, and Xiang Bai. 2018. Textboxes++: A single-shot oriented scene text detector. IEEE Transactions on Pattern Analysis and Machine Intelligence 41, 4 (2019), 1452–1459.

[14] Xue Yang, Hao Sun, Kun Fu, Jirui Yang, Xian Sun, Menglong Yan, and Zhi Guo. 2018. Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks. Remote Sensing 10, 1 (2018), 132.

[15] Xue Yang and Junchi Yan. 2020. Arbitrary-Oriented Object Detection with Circular Smooth Label. In European Conference on Computer Vision. Springer, 677–694.

[16] Xue Yang and Junchi Yan. 2022. On the arbitrary-oriented object detection: Classification based approaches revisited. International Journal of Computer Vision (2022), 1–26.

[17] Xue Yang, Junchi Yan, Ziming Feng, and Tao He. 2021. R3Det: Refine Single-Stage Detector with Feature Refinement for Rotating Object. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 35. 3163–3171.

[18] Xue Yang, Junchi Yan, Wenlong Liao, Xiaokang Yang, Jing Tao, and Tao He. 2022. SCRDet-v: Detecting Small, Cluttered and Rotated Objects via Instance-Level Feature Denoising and Rotation Loss Smoothing. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022).

[19] Xue Yang, Junchi Yan, Qi Ming, Wentao Wang, Xiaoping Zhang, and Qi Tian. 2021. Rethinking rotated object detection with gaussian wasserstein distance loss. In International Conference on Machine Learning. PMLR, 11830–11841.

[20] Xue Yang, Jirui Yang, Junchi Yan, Yue Zhang, Tengfei Zhang, Zhi Guo, Xian Sun, and Kun Fu. 2019. Scriet: Towards more robust detection for small, cluttered and rotated objects. In Proceedings of the IEEE International Conference on Computer Vision. 8232–8241.

[21] Xue Yang, Xiaojiang Yang, Jirui Yang, Qi Ming, Wentao Wang, Qin Tian, and Junchi Yan. 2021. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence. Advances in Neural Information Processing Systems 34 (2021).

[22] Xue Yang, Yue Zhou, and Junchi Yan. 2021. AlphaRotator: A Rotation Detection Benchmark using TensorFlow. arXiv preprint arXiv:2111.06677 (2021).

[23] Xue Yang, Yue Zhou, Gefan Zhang, Jitun Yang, Wentao Wang, Junchi Yan, Xiaopeng Zhang, and Qi Tian. 2022. KFBox Loss for Rotated Object Detection. arXiv preprint arXiv:2201.12558 (2022).

[24] Ze You, Shouhui Liu, Han Hu, Liwei Wang, and Stephen Lin. 2019. Reppoints: Point set representation for object detection. In Proceedings of the IEEE International Conference on Computer Vision. 9657–9666.

[25] Shufeng Zhang, Cheng Chi, Yongjiaqng Yao, Zhen Lei, and Stan Z. Li. 2020. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In IEEE Conference on Computer Vision and Pattern Recognition. 9759–9768.

[26] Yu Zheng, Dan Yang, Zinan Xie, Jiwon Lu, and Jie Zhou. 2020. Rotation Robust Intersection over Union for 3D Object Detection. In European Conference on Computer Vision. Springer, 464–480.

[27] Xinyu Zhou, Cong Yao, He Wen, Yuzhi Wang, Shuchang Zhou, Weiran He, and Jiajun Liang. 2017. East: an efficient and accurate scene text detector. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 5551–5560.