ITC 1/53

Information Technology
and Control
Vol. 53 / No. 1 / 2024
pp. 71-79
DOI 10.5755/j01.itc.53.1.35393

An OBB Detection Algorithm of Maintenance Components
Based on YOLOv5-OBB-CR

Received 2023/10/19 | Accepted after revision 2023/12/07

HOW TO CITE: Ge, Z., Yang, Y., Li, Q., Wang, F., Luo, X. (2024). An OBB Detection Algorithm of Maintenance Components Based on YOLOv5-OBB-CR. *Information Technology and Control*, 53(1), 71-79. https://doi.org/10.5755/j01.itc.53.1.35393

An OBB Detection Algorithm of Maintenance Components Based on YOLOv5-OBB-CR

Zhexue Ge, Yongmin Yang, Qiang Li, Fang Wang, Xu Luo

College of Intelligence Science and Technology, National University of Defense Technology, Changsha, China

Corresponding author: Yongmin Yang, yangyongmin@163.com

By detecting the position of maintenance components in real-time, maintenance guidance information can be superimposed and important operational guidance can be provided for maintenance personnel. The YOLOv5-OBB-CR real-time detection algorithm is proposed for maintenance component with orientation bounding box based on improved YOLOv5-OBB. The C3 module in the original network is improved to CReToNeXt, which can more effectively enhance the network's ability to learn image features. Considering that the network learning is the labeled rotation box information, the original Loss function CIoU is improved to SIoU with angle loss information, and the improved Loss function can more effectively describe the regression of the target box. The demonstration shows that the mAP@.5 0.95 of YOLOv5-OBB-CR-s (SIoU) is 85.6%, which is 6.7% higher than the original YOLOv5 OBB algorithm.

KEYWORDS: YOLOv5-OBB, maintenance component, YOLOv5-OBB-CR, SIoU, position real detection.

1. Introduction

In the process of complex product maintenance, due to the complexity of operating objects and process flow, maintenance personnel often make various maintenance mistakes. By detecting the name and location of maintenance targets in real-time, and using augmented reality technology to guide the maintenance process and operation process, the efficiency and accuracy of maintenance operations can be improved. For example, Castellanos et al. [2] developed an interactive electronic manual-based AR assisted maintenance and repair system for vertical centrifugal deep well pumps, which effectively reduced the workload and maintenance costs of operators. Chen et al. [3] developed the augmented reality detection and maintenance system BIM AR FSE for fire equipment based on the Building information modeling (BIM). This system combines Building information modeling and real objects through augmented reality technology, and can timely and effectively present maintenance and inspection information to staff. Due to the varying sizes of the maintenance targets, ranging from millimeter to meter scales, and the fact that the maintenance targets are easily obstructed by other external structures, there are significant technical challenges in ensuring both high accuracy and real-time performance.

Traditional maintenance object detection mainly uses image feature based methods, which require extracting Planar or three-dimensional features of the target image, and then performing template matching for recognition [10, 12]. These methods have played a good role in certain situations, but due to the complexity of the actual maintenance scene objects and the large number of feature points, real-time detection performance and robustness are often difficult to ensure.

With the development of deep learning technology and the improvement of computer hardware computing ability, Convolutional neural network (CNN) gradually occupies a dominant position in the field of computer vision. Researchers apply a large number of related technologies to target detection, which can extract more abstract and deeper feature information, make the model more generalized, and thus greatly improve robustness and accuracy. CNN has been widely applied in fields such as autonomous driving [18], facial recognition [13], and defect detection [15, 17].

The YOLO series is a typical representative of CNN object detection algorithms [1, 5, 14, 20]. YOLOv3 is improved on the basis of YOLOv1 and YOLOv2, and the basic classification network is improved to Darknet-53. The classifier abandons the original Softmax, and the classification loss function uses the binary Cross entropy loss. Compared with R-CNN and Fast R-CNN, its reasoning speed has been improved by leaps and bounds. The YOLOv5 algorithm integrates the advantages of previous versions, further improving detection accuracy and speed [8, 16, 21].

YOLOv5-OBB's real-time detection algorithm for maintenance parts is an Orientated Bounding Box detection algorithm. Compared with HBB (Horizontal Bounding Box) detection algorithm. it can not only capture the central point of the parts more accurately, but also adapt to the changes in the position and orientation of the target in space, so as to obtain a more compact and accurate bounding box. However, during the detection process, the generalization ability of the network is limited, in addition, there are still many issues such as missed detections and low feature extraction ability [4]. Therefore, a new real-time detection algorithm for maintenance components, YOLOv5-OBB-CR [9], is proposed. YOLOv5-OBB-CR [22] algorithm improves the original C3 module into CReToNeXt module, and improves the original Loss function to SioU [4], effectively enhancing the network feature extraction ability. Both of the improved algorithms are applied to the detection of maintenance components in the YN92 marine diesel engine, the experimental results show that YOLOv5-OBB-CR algorithm has stronger feature extraction ability and higher detection accuracy in maintenance component detection compared to the original YOLOv5-OBB algorithm [11].

2. YOLOv5-OBB-CR Algorithm

The YOLOv5-OBB-CR algorithm is improved on the basis of the rotating object detection network YOLOv5-OBB. Next, the OBB bounding box, network structure improvement, Loss function of YOLOv5-OBB-CR algorithm are introduced in detail [7].

2.1. The OBB Bounding Box

The bounding boxes for HBB and OBB detection are shown in Figure 1. The target box representation of the YOLOv5 algorithm is HBB (x, y, w, h), where (x, y) represents the center coordinate of HBB, w represents the width of HBB, and h represents the height of HBB. As shown in Figure 2, the target box of YOLOv5-OBB algorithm can be represented using the 5 parameter long edge representation D_{le}. The representation of D_{le} is $(x_{le}, y_{le}, w_{le}, h_{le}, \theta_{le})$, where $(x_{le}h, y_{le})$ is the center coordinate of the OBB, w_{le} is the longest edge of OBB, h_{le} is the adjacent edge of w_{le}, θ_{le} represents the x-axis and the longest edge w_{le}, where $\theta_{le} \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

Figure 1
HBB and OBB

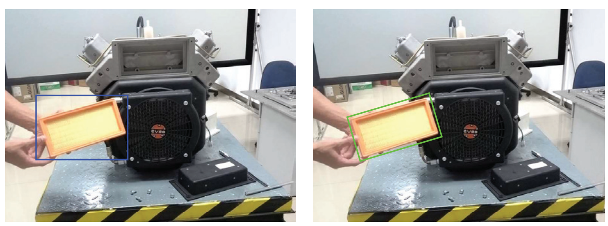


Figure 2

The 5 parameter long edge representation method of D_{le}

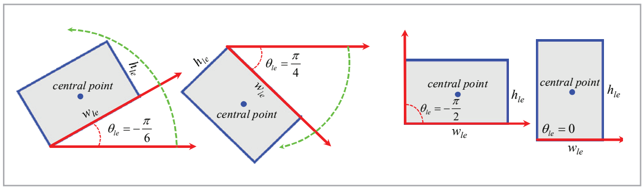


Figure 3

YOLOv5-OBB-CR overall network structure

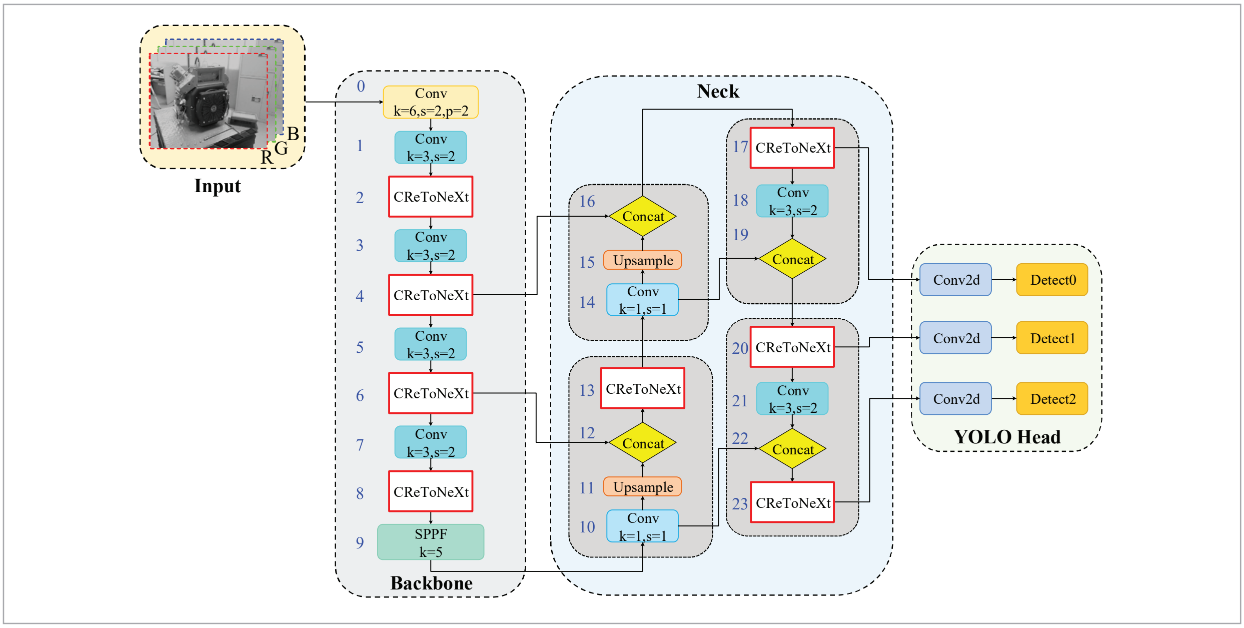
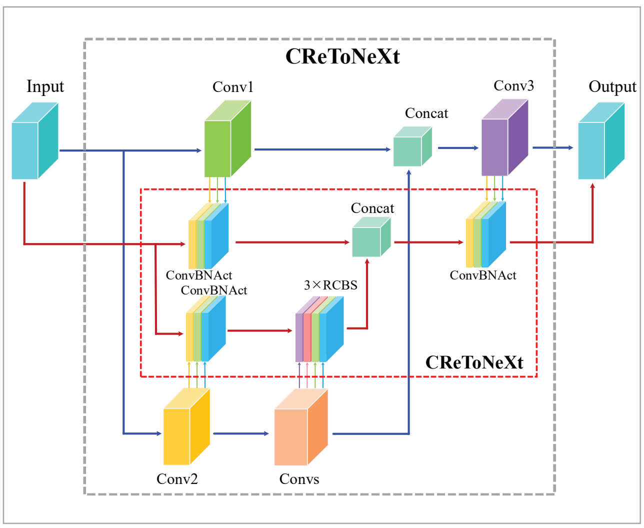


Figure 4

CReToNeXt structure



2.2. YOLOv5-OBB-CR Network Structure

The proposed overall network structure is shown in Figure 3. By improving the C3 module in the original network to CReToNeXt, it can more enhance the network's ability to learn image features effectively. Considering that the network learning is the labeled rotation box information, the original Loss function CIoU is improved to SIoU with angle loss information. The improved Loss function can more effectively describe the regression of the target box [19].

The structure of CReToNeXt is shown in Figure 4. The core structural blocks include Conv1, Conv2, Convs, and Conv3. Conv1, Conv2, and Conv3 are composed of Conv2d, BatchNorm2d, and Swish, and the module is named ConvBNAct (as shown in Figure 4). Convs consists of 3 RCBS (as shown in Figure 5), while RCBS consists of RepConv and one ConvBNAct. The structural detail of RepConv is shown in Figure 6. During the YOLOv5-OBB-CR network training process, RepConv consists of a 3×3 convolution with a BN layer, a 1×1 convolution with a BN layer and a separate BN layer when the number of input and output channels is the same. During the deployment phase, RepConv

Figure 5
ConvBNAct structure and RCBS structure

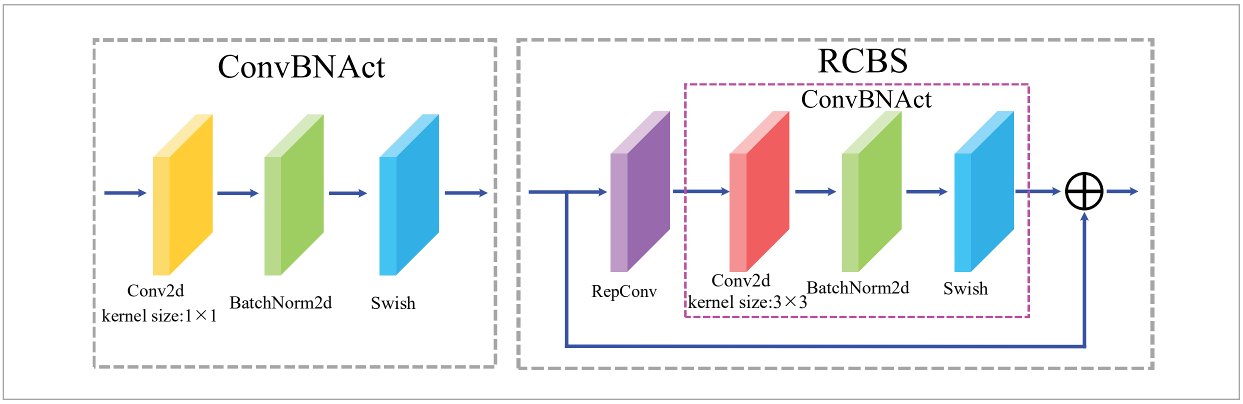
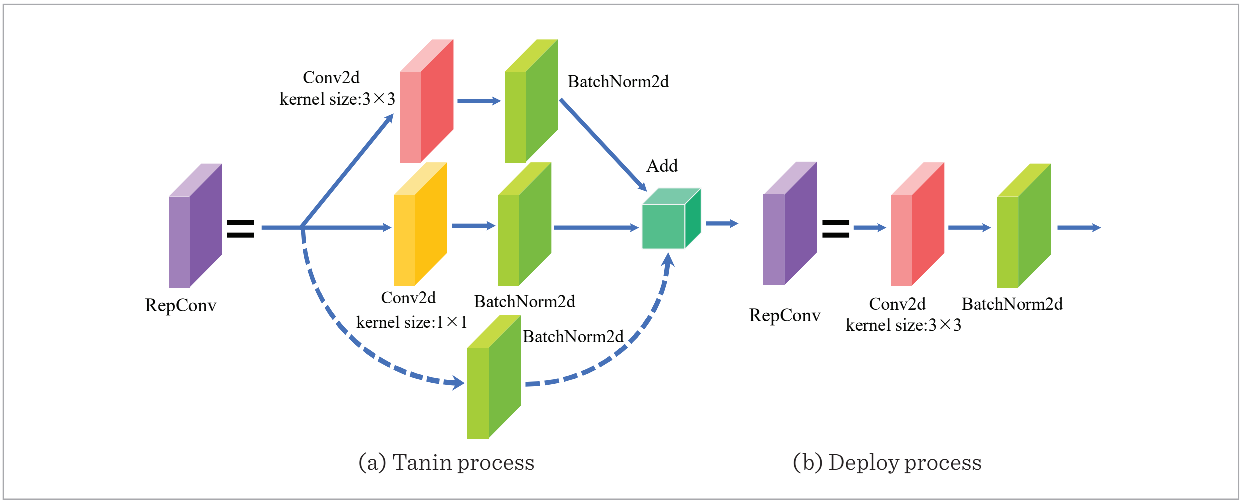


Figure 6
RepConv structure



consists only of a 3×3 convolution and a BN layer. This can reduce one Conv2d module and two BatchNorm2d modules, reducing computational complexity and improving the detection speed of maintenance components.

2.3. Positioning Loss function

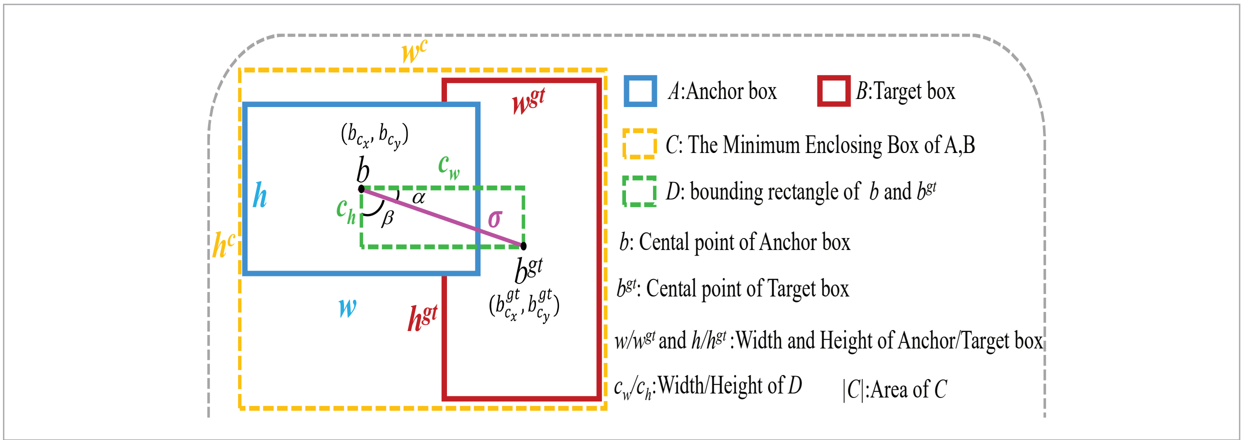
In network training, YOLOv5-OBB-CR algorithm abandons the original location Loss function and uses SIoU Loss function. It considers the vector angle between the real box and the prediction box, and redefines the penalty index. Compared with other Loss function, SIoU considers the vector angle between the expected regression. The relevant parameters are shown in Figure 7. The Loss function includes four parts: (1) Angle cost; (2) Distance cost; (3) Shape cost; (4) IoU cost.

1 Angle cost

$$\Lambda = 1 - 2 \times sin^2\left(arcsin(x) - \frac{\pi}{4}\right) \tag{1}$$

$$x = \frac{c_h}{\sigma} = \frac{max\left(b_{c_y}^{gt}, b_{c_y}\right) - min\left(b_{c_y}^{gt}, b_{c_y}\right)}{\sqrt{\left(b_{c_x}^{gt} - b_{c_x}\right)^2 + \left(b_{c_y}^{gt} - b_{c_y}\right)^2}} = sin\,(\alpha) \tag{2}$$

Figure 7
The parameters of SIoU location cost function



2 Distance cost

$$\Delta = \sum_{t=x,y} \left(1 - e^{(-\gamma\rho_t)}\right), \rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{w^c}\right)^2, \quad \rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{h^c}\right)^2, \gamma = 2 - \Lambda \tag{3}$$

3 Shape cost

$$\Omega = \sum_{t=w,h} \left(1 - e^{(-\omega_t)}\right)^\theta, \quad \omega_w = \frac{|w - w^{gt}|}{max(w, w^{gt})}, \omega_h = \frac{|h - h^{gt}|}{max(h, h^{gt})} \tag{4}$$

4 IoU cost

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{5}$$

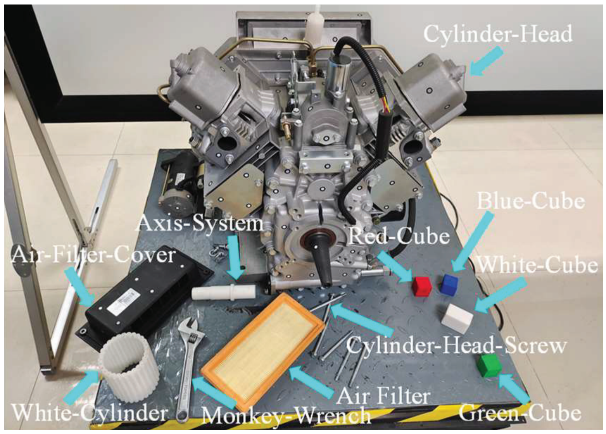
The total cost function is :

$$L_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \tag{6}$$

3. Demonstration

Taking some components of the YN92 marine diesel engine and some 3D printed models as the target category. As shown in Figure 8, there are 11 categories in the YN92 data OBB dataset, resulting in a training set of 440 images, a validation set of 49 images, and a test set that is recollected according to actual needs. Next, the improved algorithm is used for model training and obtain corresponding experimental results.

Figure 8
The maintenance target dataset YN92-data-OBB



3.1. The Train Parameters

The model is trained using the deep learning workstation shown in Table 1. Train models with three different width and depth factors: *s*, *m*, and *l*. Select the SGD optimizer with a learning momentum of 0.937 and a weight attenuation of 0.0005. In the first three generations of training (three epochs), the learning rate increases from 0.0033 to 0.01, and then decreases to 0.001 through cosine attenuation.

Table 1
Deep learning workstation configuration

Software configuration	System: Windows 11
	Frame: Pytorch 1.9.0
	Version: CUDA 11.1
	Develop environment: Pycharm 2022
Hardware configuration	CPU: Intel Core i7-10875H
	Memory: 32.0G
	Graphics card: RTX 2080 Super
	Graphics memory: 8G

Table 2
Different YOLOv5- OBB-CR and model factors

Model	depth_factor	width_ factor
YOLOv5-OBB-CR-s	0.33	0.50
YOLOv5-OBB-CR-m	0.67	0.75
YOLOv5-OBB-CR-l	1.00	1.00

The location Loss function uses SIoU, and the classification loss uses Cross entropy loss. For the geometric distortion, set the translation, scaling, fliplr, and flipup to 0.1, 0.25, 0.5, and 0.5, respectively. Image data augmentation uses Mosaic data augmentation and Cutmix data augmentation, with parameters of 0.75 and 0.1, respectively. The models are all trained using RTX2080 Super, with 8GB of GPU memory. The batch size is set reasonably based on the model size, and all models are trained with 300 epochs.

3.2. Results

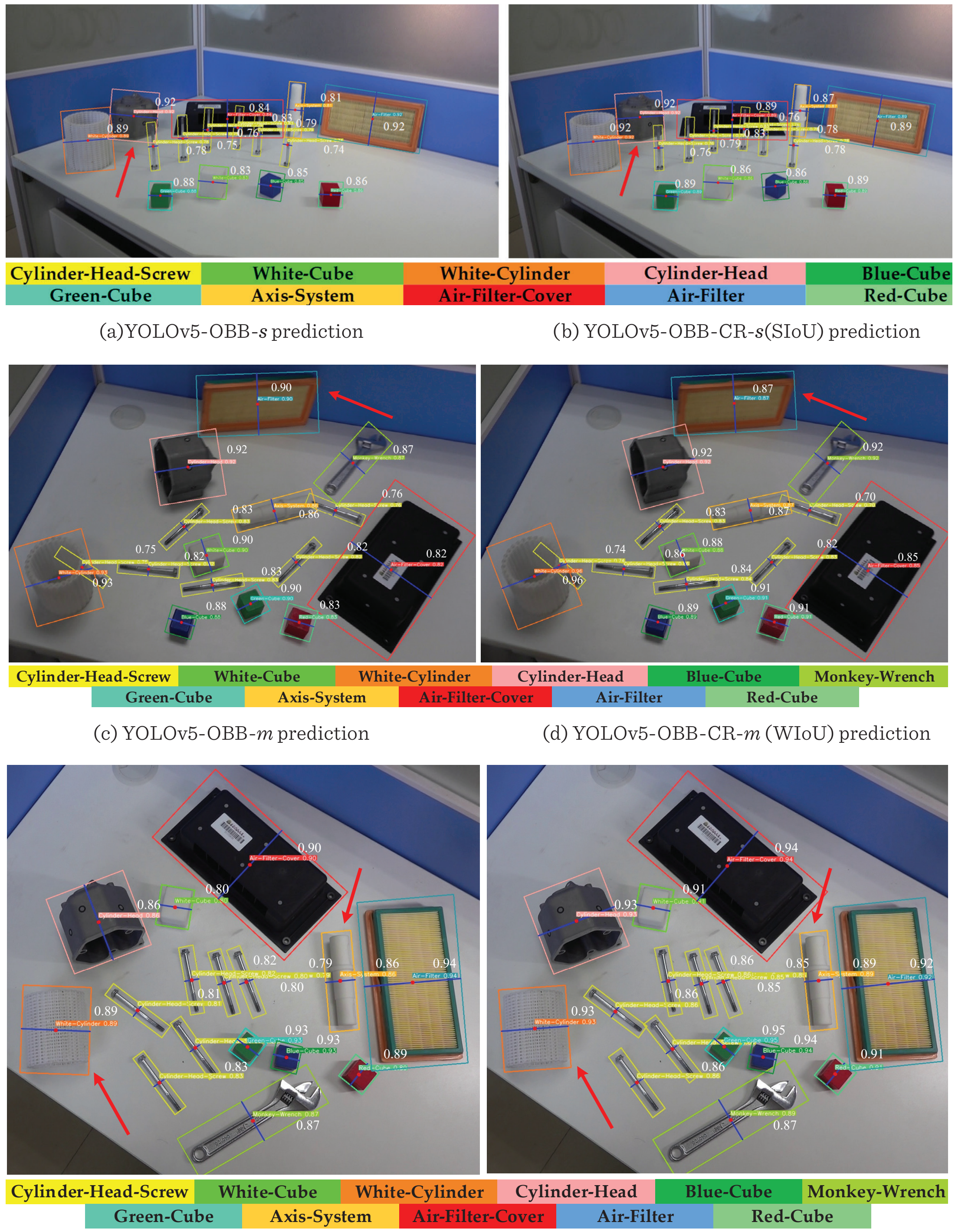
Table 3 presents the training results of the rotating target dataset YN92 data OBB. It can be seen that the improved YOLOv5-OBB-CR algorithm has different degrees of improvement compared with the original YOLOv5-OBB algorithm in the index mAP@.5:0.95. When SIoU is used to locate the Loss function mAP@.5, the performance of the 0.95 indicator is the best, in which YOLOv5-OBB-CR-*s* (SIoU) is 6.7% higher than YOLOv5-OBB-*s*, YOLOv5-OBB-CR-*m* (SIoU) is 2.8% higher than YOLOv5-OBB-*m*, and YOLOv5-OBB-CR-*l* (SIoU) is 3.1% higher than YOLOv5-OBB-*l*. The algorithm proposed in this article has more advantages in the accuracy of rotation detection of maintenance components.

Figures 9(a)-(b) use YOLOv5-OBB-s and YOLOv5-OBB-CR-s (SIoU) models for comparison. Figures 9(c)-(d) use YOLOv5-OBB-*m* and YOLOv5-OBB-CR-*m* (SIoU) models for comparison. Figures 9(e)-(f) use YOLOv5-OBB-*l* and YOLOv5-OBB-CR-*l* (SIoU) models for comparison. In terms of category detection confidence, except for a few categories, the improved algorithm is higher than the original YOLOv5-OBB algorithm, and the improved algorithm is closer to the real object box in the target detection box, which can better describe the rotation features of the real object (marked by the red arrow).

Table 3
YN92 data OBB for Rotating Dataset

Models	GFLOPs	mAP@.5	mAP@.5:0.95
YOLOv5-OBB-*s*	17.4	99.0	77.6
YOLOv5-OBB-CR-*s* (SIoU)	46.5	99.0	84.3(**+6.7%**)
YOLOv5-OBB-*m*	50.3	99.1	82.8
YOLOv5-OBB-CR-*m* (SIoU)	176.7	99.2	85.6(**+2.8%**)
YOLOv5-OBB-*l*	110.9	98.9	83.1
YOLOv5-OBB-CR-*l* (SIoU)	444.1	99.2	86.2(**+3.1%**)

Figure 9
The comparison of category detection confidence



(a)YOLOv5-OBB-*s* prediction

(b) YOLOv5-OBB-CR-*s*(SIoU) prediction

(c) YOLOv5-OBB-*m* prediction

(d) YOLOv5-OBB-CR-*m* (WIoU) prediction

(e) YOLOv5-OBB-*l* prediction

(f) YOLOv5-OBB-CR-*l* (WIoU v3) prediction

4. Conclusion

To achieve the OBB detection for maintenance components, an improved method based on YOLOv5-OBB is proposed. The C3 module in YOLOv5-OBB algorithm is improved to CReToNeXt, resulting in YOLOv5-OBB-CR that can more effectively enhance the network's ability to learn image features. Through experimental comparison, it is found that the OBB of YOLOv5-OBB-CR algorithm is more compact and accurate. Based on different depth_multiple and width_multiple, the mAP@.5:0.95 indicators of YOLOv5-OBB-CR-*s*, YOLOv5-OBB-CR-*m*, and YOLOv5-OBB-CR-*l* are higher than YOLOv5-OBB algorithm, among which the mAP@.5 0.95 indicator of YOLOv5-OBB-CR-s (SIoU) is 85.6%, which is 6.7% higher than the original algorithm.

The proposed method can identify the position of maintainability test objects and perform 3D reconstruction and collision detection in virtual space, thus achieving virtual and real fusion of physical equipment and virtual environment [6, 23]. In addition, this method can be used to identify maintenance objects and provide relevant maintenance knowledge to maintenance personnel, thereby reducing the threshold of maintenance technology.

Data Sharing Agreement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflict of Interests

The author(s) declared no potential conflicts of interest with respect to the research, author-ship, and/or publication of this article.

References

1. Bochkovskiy A., Wang C., Liao H. M. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934, 2020.
2. Castellanos, M. J., Navarro-Newball, A. A. Prototyping an Augmented Reality Maintenance and Repairing System for a Deep Well Vertical Turbine Pump. IEEE Xplore, 2019. https://doi.org/10.1109/CONIELECOMP.2019.8673254
3. Chen Y. A. Y. N., Lai Y. A., Lin Y. A. BIM-Based Augmented Reality Inspection and Maintenance of Fire Safety Equipment. Automation in Construction, 2020, 103041. https://doi.org/10.1016/j.autcon.2019.103041
4. Chen, Z. Y., Wang X. L., Zhang, W. W., Yao, G. D., Li, D. D., Zeng, L. Autonomous Parking Space Detection for Electric Vehicles Based on Improved YOLOv5-OBB Algorithm. World Electric Vehicle Journal, 2023. https://doi.org/10.3390/wevj14100276
5. Ge, Z. X., Qi, Z., Luo, X., Yang, Y. M., Zhang, Y. Multistage Bayesian Fusion Evaluation Technique Endorsing Immersive Virtual Maintenance, Measurement, 2021, 177, 109344. https://doi.org/10.1016/j.measurement.2021.109344
6. Ge, Z. X., Zhang, Y., Wang F., Luo X., Yang Y. M. Virtual-Real Fusion Maintainability Verification Based on Adaptive Weighting and Truncated Spot Method, Maintenance and Reliability, 2022. https://doi.org/10.17531/ein.2022.4.14
7. Gevorgyan, Z. SIoU Loss: More Powerful Learning for Bounding Box Regression. arXiv preprint arXiv:2205.12740, 2022.
8. Girshick, R., Donahue, J., Darrell, T., Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, 580-587. https://doi.org/10.1109/CVPR.2014.81
9. Girshick R. Fast R-CNN: Proceedings of the IEEE International Conference on Computer Vision, 2015. https://doi.org/10.1109/ICCV.2015.169
10. Jing O., Lai Z., Wan Q.H., Liu X., Li Y.J. Weld-Seam Identification and Model Reconstruction of Remanufacturing Blade Based on Three-Dimensional Vision. Advanced Engineering Informatics, 2021, 49: 101300. DOI: https://doi.org/10.1016/j.aei.2021.101300
11. Li X., Wu J. Developing a More Reliable Framework for Extracting Traffic Data from a UAV Video, IEEE Transactions on Intelligent Transportation Systems, 2023, 1-12. https://doi.org/10.1109/TITS.2023.3290827
12. Liu, Z. Y., Li, Z. X., Chen, J. Q., Li, S. P., Lu Q. B., Song, W. H. A Survey of Fault Diagnosis of Train Components Based on Machine Vision. Computer Applications and Software, 2022, 39(1): 1-9.

13. Najibi, M., Samangouei, P., Chellappa, R., Davis, L. S. SSH: Single Stage Headless Face Detector. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, 4885-4894, doi: https://doi.org/10.1109/ICCV.2017.522

14. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016. https://doi.org/10.1109/CVPR.2016.91

15. Redmon J., Farhadi A. YOLO9000: Better, Faster, Stronger. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. https://doi.org/10.1109/CVPR.2017.690

16. Redmon J., Farhadi A. YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767, 2018.

17. Wang, T., Chen, Y., Qiao, M., Snoussi, H. A Fast and Robust Convolutional Neural Network-Based Defect Detection Model in Product Quality Control. The International Journal of Advanced Manufacturing Technology, 2018, 94, 3465-3471. https://doi.org/10.1007/s00170-017-0882-0

18. Wu, D., Liao, M., Zhang, W., Wang, X., Bai, X., Cheng, W., Liu, W. YOLOP: You Only Look Once for Panoptic Driving Perception. 2022, ArXiv:2108.11250. https://arxiv.org/abs/2108.11250

19. Yun, S., Han, D., Chun, S., Seong Taek, O., Yoo, Y., Choe, J. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. International Conference on Computer Vision, 2019. https://doi.org/10.1109/ICCV.2019.00612

20. Zhang R., Wen C. SOD-YOLO: A Small Target Defect Detection Algorithm for Wind Turbine Blades Based on Improved YOLOv5. Advanced Theory and Simulations, 2022, 5(7), 2100631. https://doi.org/10.1002/adts.202100631

21. Zheng, J. C., Sun, S. D., Zhao, S. J. Fast Ship Detection Based on Lightweight YOLOv5 Network. IET Image Processing, 2022, 16(6), 1585-1593. https://doi.org/10.1049/ipr2.12432

22. Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(07), 12993-13000. https://doi.org/10.1609/aaai.v34i07.6999