International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 2, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v2n1.20
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Real-Time Detection of Drill Pipe Joints Using Improved YOLOv5x Model Applied to Drilling Operation Images

Fanyi Tang, Qizhi Zhang

School of Electronic Engineering, Xi'an Shiyou University, Xi'an, China

ABSTRACT

It is widely known that pipe assembly and disassembly still lacks automation. To enhance drilling efficiency, we propose an autonomous detection and positioning model based on improved YOLOv5x for drill pipe joints. Firstly, we use Activate or Not (ACON) activation function and Convolution Block Attention Module (CBAM) to enhance feature extraction and representation ability. Then, the loss function is changed from Complete IoU (CIOU) loss to Scale-Invariant IoU (SIOU) loss, which increases the accuracy of the bounding box regression. Finally, the predict heads are tailored to be more effective in detecting tiny targets. Following network training, the final drill pipe joint detection model, based on the improved YOLOv5x, achieved an average accuracy of 99.10%, a 3.5% improvement over the base YOLOv5x algorithm. The proposed method effectively promotes the intelligence of oil drilling equipment by quickly and accurately detecting drill pipe joints, consequently enhancing drilling operation efficiency.

KEYWORDS

YOLOv5x; Object detection; Deep learning

1. INTRODUCTION

The application of computer vision techniques for the detection of oil drilling equipment during operations constitutes a crucial component of numerous fully automated systems within the oil industry[1]. This sector is progressively seeking to enhance process efficiency through automation. Despite these advancements, the automation of pipe assembly and disassembly remains an unresolved challenge, necessitating manual intervention for certain procedural aspects[2]. Specifically, workers are required to manually verify the positioning of pipe sections prior to their attachment to or detachment from the drill string, a step critical for the subsequent actions of the iron roughneck[4]. Consequently, there is a growing interest in eliminating this manual verification step by employing computer vision systems capable of precisely estimating pipe positions, thereby facilitating the automated handling and assembly/disassembly of the pipe string.

Object detection technology, characterized by its rapid response, high accuracy, and non-contact nature, is well-suited for the dynamic and complex environments of drilling operations[5]. This approach employs image sensors to capture images of drill pipe joints, utilizes image processing algorithms to extract feature information, and then applies this information to classify and pinpoint the location of drill pipe joints. Unlike traditional detection algorithms, which depend on manually designed feature operators like SIFT[5], Harr[6] and HOG[7], and employ fixed-size image windows for detection—thereby enhancing accuracy—these conventional methods are sensitive to variations in lighting and environmental conditions, such as background interference. Consequently, traditional detection techniques require intricate threshold adjustments for defect recognition, with any environmental changes necessitating careful recalibration of these thresholds. This limitation

significantly undermines their adaptability and robustness, rendering them less effective in new or altered environments.

Deep learning-based object detection methods overcome the constraints of traditional algorithms through the utilization of Convolutional Neural Networks (CNNs), establishing deep learning as the predominant approach in image processing[8]. These algorithms exploit CNNs to extract sophisticated and abstract features, significantly enhancing representational capabilities[9]. Predominantly, two types of deep learning architectures are employed: the One-Stage and the Two-Stage models. One-Stage detectors, such as YOLO[10] and SSD[11], offer real-time object localization and classification within a singular framework, prioritizing speed albeit at the expense of lower accuracy compared to their Two-Stage counterparts. Conversely, Two-Stage detectors, exemplified by RCNN[11] and Faster-RCNN[12], first generate region proposals before classifying and refining the positioning of each proposal. This approach, while computationally more demanding, yields superior accuracy due to its detailed processing methodology.

Several object detection methods have been applied in drilling filed. Liang B [14] enhanced the Faster R-CNN model by incorporating multi-scale image inputs and multi-layer feature fusion within an improved SqueezeNet framework, achieving significant advancements in the detection of small-sized targets. This model demonstrated superior performance with an accuracy of 90.09% and a recall rate of 98.32% in evaluations conducted using data from underground rockburst relief zones in coal mines, albeit at the cost of increased training and recognition durations. In another development, Alexander[15] introduced Tiny SSD, a real-time, single-shot deep neural network that combines a non-uniform, optimized Fire subnetwork stack with SSD-based auxiliary convolutional layers. This innovation resulted in a remarkably compact model size of 2.3MB, which is approximately 26 times smaller than Tiny YOLO, without compromising on object detection efficacy. Furthermore, Xiaofeng Ji[16] proposed the RFA-YOLO, a modified YOLOv4-based framework, which has been proven to deliver high-precision and high-speed Personal Protective Equipment (PPE) detection on offshore drilling platforms. This was substantiated through comparative analyses, where the model achieved a 93.1% accuracy and a processing speed of 13 FPS. When juxtaposed with the Faster R-CNN and SSD models, the YOLO variant exhibited a notable enhancement in accuracy, alongside reductions in both training and recognition times, underscoring its efficacy and efficiency in practical applications.

During the assembly and disassembly of drill pipe joints, the targets often exhibit specific characteristics that complicate detection efforts. These include the mobility of the targets, which can be obscured by the dense arrangement of pipelines on drilling platforms, and suboptimal camera placements that result in the targets appearing exceedingly small in captured images. These factors necessitate enhanced feature extraction and representation capabilities within the YOLO algorithm, necessitating adjustments to the network to meet various detection criteria effectively.

This study introduces an augmented YOLOv5x model tailored for the automated detection and localization of drill pipe joints, incorporating several advancements to address these challenges. Firstly, the adoption of the ACON activation function enhances the model's feature extraction capabilities. Additionally, the use of the CBAM mechanism serves as a spatial and channel attention method, augmenting the model's capacity for feature representation. Subsequently, the transition to an efficient SIOU loss function from the conventional CIOU loss enhances bounding box regression accuracy. Lastly, modifications to the prediction head are specifically designed to improve the detection of minuscule drill pipe connections, collectively bolstering the model's performance in challenging detection scenarios.

2. METHODOLOGIES

2.1. YOLOv5

Compared with most previous models, YOLOv5 has a stronger detection effect, with the Ultralytics team releasing version 7.0 in Nov 22,2022. This version further improved the detection performance and speed by using the Conv module instead of the Focus module, replacing the SPPF module with the SPP module and reducing the number of repetitions of the C3 layer in the backbone network from 9 to 6.

The YOLOv5 7.0 has five main models, classified by size and complexity as v5n, v5s, v5m, v5l and v5x. Among them, YOLOv5x exhibits the highest detection performance, enabling fast identification of drill pipe joints in complex drilling operations. The network structure of YOLOv5x is shown in Fig.1, including four parts: Input, Backbone, Neck and Prediction.

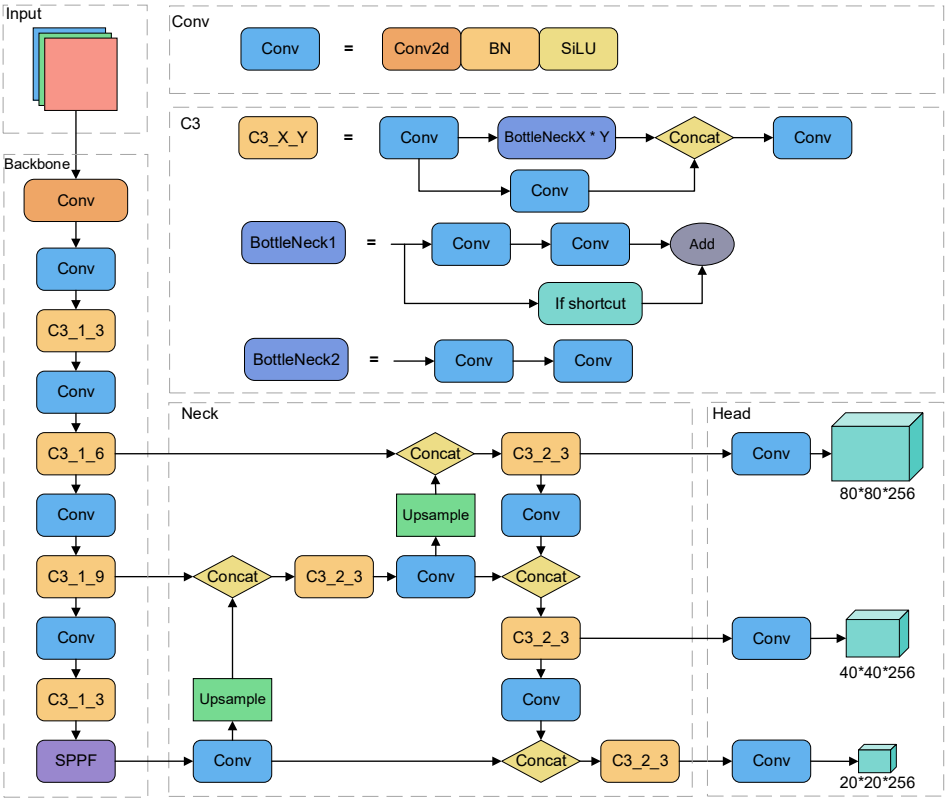


Fig.1 Network structure of YOLOv5x algorithm

2.2. Input

The input preprocesses the image, which includes the mosaic data enhancement, the adaptive anchor frame computation and the adaptive image scaling. The mosaic data augmentation combines four randomly selected images, applying random scaling, cropping, and arrangement to create a larger image. This part enhances the sample diversity and improves the generalization capability of the model. The adaptive anchor box calculation automatically determines suitable anchor box sizes based on object dimensions, which prevents the use of unreasonable anchor boxes that would increase the training time. The adaptive image scaling adjusts the input image to the network's required size (usually 640*640). The relatively consistent dimensions and aspect ratios could avoid distortion.

2.3. Backbone

The Backbone is responsible for extracting and representing image features, which includes the Conv module, C3 module, and SPPF (Spatial Pyramid Pooling Faster) module.

The Conv module performs convolution operations on input features, consisting of the convolutional layers, the batch normalization layers, and the activation function.

The convolution layers extract the local spatial information of the input features. The batch normalization layers normalize feature value distributions in the neural network. The activation function can balance computational efficiency and model performance by introducing nonlinear transformations. In YOLOv5x, the activation function used is SiLU.

The C3 module performs feature extraction and fusion, and it promotes information flow across different-level feature maps through partial connections, thus enhancing detection speed. The C3 module consists of multiple Bottleneck structures. Each Bottleneck structure contains two consecutive Conv modules. The first Conv reduces the number of channels of the feature map by half, aiming to increase the perceptual field of the network and reduce the computational effort. The second Conv is used to extract features and double the number of channels, ensuring that the inputs and outputs have the same number of channels. In addition, the output feature maps obtained from the two Conv modules are connected to the input feature maps by shortcutting the residuals to achieve feature fusion. It should be noted that Bottleneck in Neck does not use shortcut. The SPPF module performs pooling operations on features maps at various scales. SPPF is an improved spatial pyramid pooling based on SPP. The structure of both is shown in Fig.2. Different from the a single large-sized pooling kernel used by SPP, SPPF uses multiple cascades of small-sized pooling kernels. Compared with the SPP, SPPF has a lower computational costs and higher efficiency.

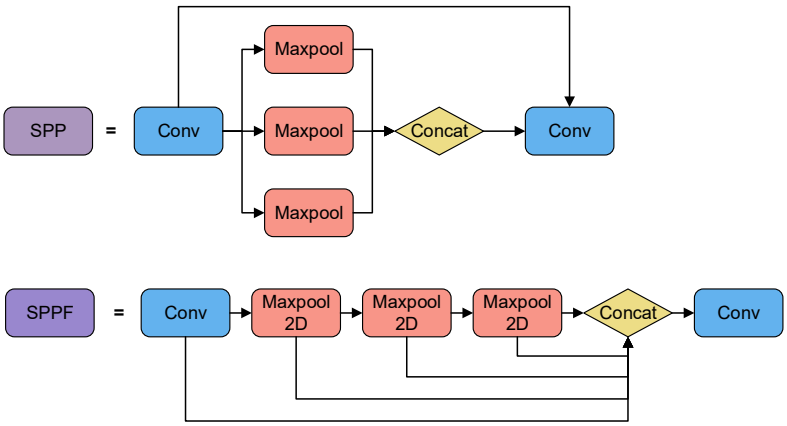


Fig.2 Structure of SPP and SPPF

2.4. Neck

The neck network performs feature fusion, where feature maps from different scales of the backbone network are fused and processed for target detection at different scales. In this case, shallow feature maps have high resolution but weak semantic information, and deep feature maps have stronger semantic information but lower resolution. Neck network contains feature pyramid structure and path aggregation network PANet (Path Aggregation Network). The feature pyramid helps to detect targets in different parts of the image and at different scales, improving the multi-scale performance of the model. PANet gives consistent resolution to the feature maps at different scales through up-sampling and down-sampling operations, improving the accuracy of the model detection.

2.5. Prediction

The Prediction module completes the final regression prediction to obtain the target detection result, which includes the bounding boxes and the associated class labels. This process adopts the grid-based anchors for target detection on feature maps with different scales. The original YOLOv5x includes three predict heads, and each head corresponds to feature maps of three different sizes obtained in the Neck network.

3. PROPOSED METHOD

This paper proposes an improved YOLOv5x model for automatically detecting and positioning drill pipe joints. The improved network architecture is illustrated in Fig. 6. The study focuses on enhancing the Backbone and Prediction networks. In the Backbone network, the SiLU activation function in the Conv module is replaced with ACON to improve feature capture, and representation, and reduce information loss. The integration of the C3CBAM module, incorporating the CBAM attention mechanism, before the SPPF module, emphasizes drill pipe joint identification and suppresses irrelevant features. In the Prediction network, the CIoU loss function is replaced with SIOU to address bounding box wandering and improve model stability. Additionally, a small object detection head is added to enhance the model's capability in detecting object targets.

3.1. ACON activation function

The common activation functions in deep learning like the ReLU (Rectified Linear Unit) and SiLU (Sigmoid Linear Unit) 错误!未找到引用源。perform well in non-saturation, sparsity, and alleviating gradient vanishing. However, ReLU has the phenomenon called "dead neurons," so the SiLU can induce the NAN (Not-a-Number) errors and negative neurons due to its non-monotonic increments and unbounded nature.

To solve this, Ma et al. integrated the smooth characteristics of ReLU and SiLU into the Maxout activation function, and the ACON family of activation functions are generated: ACON-A, ACON-B, and ACON-C, as follows:

$$\text{ACON-A:} \quad x \cdot \sigma(\beta x) \tag{1}$$

$$\text{ACON-B:} \quad (1-p)x \cdot \sigma(\beta(1-p)x) + px \tag{2}$$

$$\text{ACON-C:} \quad (p_1 - p_2)x \cdot \sigma(\beta(p_1 - p_2)x) + p_2 x \tag{3}$$

where σ denotes Sigmoid, β is a conversion factor, p is a learnable parameter, p1 and p2 are channel-wise.

Experiments proved the ACON family outperforms ReLU and SiLU in tasks like classification and detection. In this paper, the ACON-C function is applied to improve feature capture and representation, and the information loss can be minimized.

3.2. CBAM attention mechanism

The CBAM Attention Mechanism is proposed by Woo et al. in 2018. As shown in Fig.3, CBAM includes the Channel Attention Module (CAM) and Spatial Attention Module (SAM).

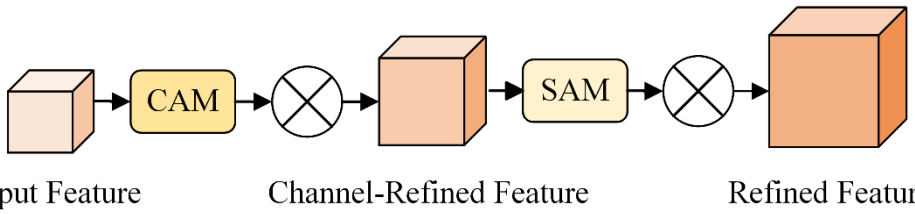


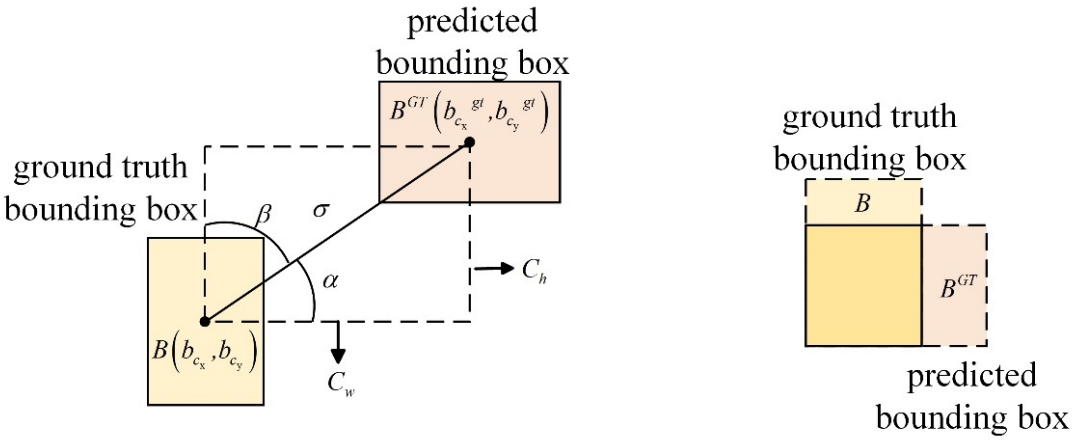
Fig.3 Structure of CBAM

The CAM can automatically adjust the importance of each channel in the feature map to enhance critical features and downplays irrelevant features. The SAM can adaptively focus on the importance of different locations in the feature map to improve the representation of key features. The CBAM combines the advantages of the CAM and SAM, enhancing the feature representation ability of the detection model.

3.3. SIOU_LOSS

The loss function in the detection model is used to measure the training errors and aids parameter adjustment. The YOLOv5x incorporates the CIoU loss function to bound the box prediction. However, the enhanced drill pipe joint detection model utilizes the SIoU loss function, which considers the distance from the center of the bounding box, width-to-height ratio, diagonal distance from the smallest outer rectangle, and angle matching problem of the object box. This modification enhances the training accuracy of the model. The improved drill pipe joint detection model replaces CIoU with SIoU to measure differences more comprehensively and improve training accuracy.

The following section elaborates on the four components of the SIoU loss function, namely the angle cost, the distance cost, the shape cost, and the IoU cost. In the other word, the positional relationship between the predicted bounding box and the ground truth bounding box is considered, as illustrated in Fig.4.



(a)When they do not intersect (b)When the predicted box intersects the real box

Fig.4 Position of the prediction box in relation to the real box

In Fig.4, B (bcx, bcy) and BGT (bcx, bcy) denote the center points of the predicted and ground truth bounding boxes, respectively. σ represents their Euclidean distance, CW and CH indicate the height

and width of the inner rectangle, α is the angle between the center line and the horizontal line, and β is its supplementary angle.

3.3.1. Angle cost

The Angle cost establishes the position correlation between the predicted bounding box and the ground truth bounding box. The Angle cost is calculated as follows:

$$\Lambda = 1 - 2 * sin^2\left(arcsinx - \frac{\pi}{4} \right) \tag{4}$$

$$x = \frac{C_h}{\sigma} = sin\alpha \tag{5}$$

$$\sigma = \sqrt{\left(b_{c_x}^{gt} - b_{c_x} \right)^2 + \left(b_{c_y}^{gt} - b_{c_y} \right)^2} \tag{6}$$

$$C_h = max\left(b_{c_y}^{gt}, b_{c_y} \right) - min\left(b_{c_x}^{gt}, b_{c_x} \right) \tag{7}$$

3.3.2. Distance cost

The Distance cost in the SIoU loss function is recalibrated based on the distance between the center points of the predicted and ground truth bounding boxes. As α approaches 0, aligning the line between the centers with the coordinate axis, the distance cost decreases significantly. Conversely, as α approaches π/4, the distance cost increases. The Distance cost is calculated as follows:

$$\Delta = \sum_{i=x,y} \left(1 - e^{-\gamma \rho_i} \right) \tag{8}$$

$$\rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w} \right)^2 \tag{9}$$

$$\rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h} \right) \tag{10}$$

$$\gamma = 2 - \Lambda \tag{11}$$

3.3.3. Shape cost

The Shape cost component assesses the shape correlation between the predicted and ground truth bounding boxes, encouraging the alignment of their size and shape. The Shape cost is calculated as follows:

$$\Omega = \sum_{t=w,h} \left(1 - e^{-\omega_t} \right)^{\theta} \tag{12}$$

$$\omega_w = \frac{\left| w - w^{gt} \right|}{max\left(w, w^{gt} \right)} \tag{13}$$

$$\omega_h = \frac{\left| h - h^{gt} \right|}{max\left(h, h^{gt} \right)} \tag{14}$$

where θ is equal to 4, ω and h represent the width and height of the ground truth bounding box, while ωgt and hgt correspond to the width and height of the predicted bounding box.

3.3.4. IoU cost

When the predicted bounding box intersects the ground truth bounding box, it must be improved according to their IoU cost to better approximate the ground truth bounding box. The IoU cost is calculated as follows:

$$IoU = \frac{|B \cap B^{GT}|}{|B \cup B^{GT}|} \tag{15}$$

In summary, by minimizing the SIoU loss function, the model parameters can be optimized, resulting in improved proximity between the predicted and ground truth bounding boxes and enhancing the model's flexibility in representing bounding boxes.

3.4. Detection Head for Tiny Targets

In some images of the drill pipes, the targets vary in number and size, and some drill pipe joint targets are small and densely distributed. This small target information can easily be lost during the convolution and downsampling process, which would reduce YOLOv5x's detection ability. Traditional detection heads have some limitations in performance. To address this problem, a larger-scale feature map is introduced as a small target detection head in the drill pipe joints detection model. So, the improved YOLOv5x model contains four different-sized detection heads that are used to output classification and regression results independently.

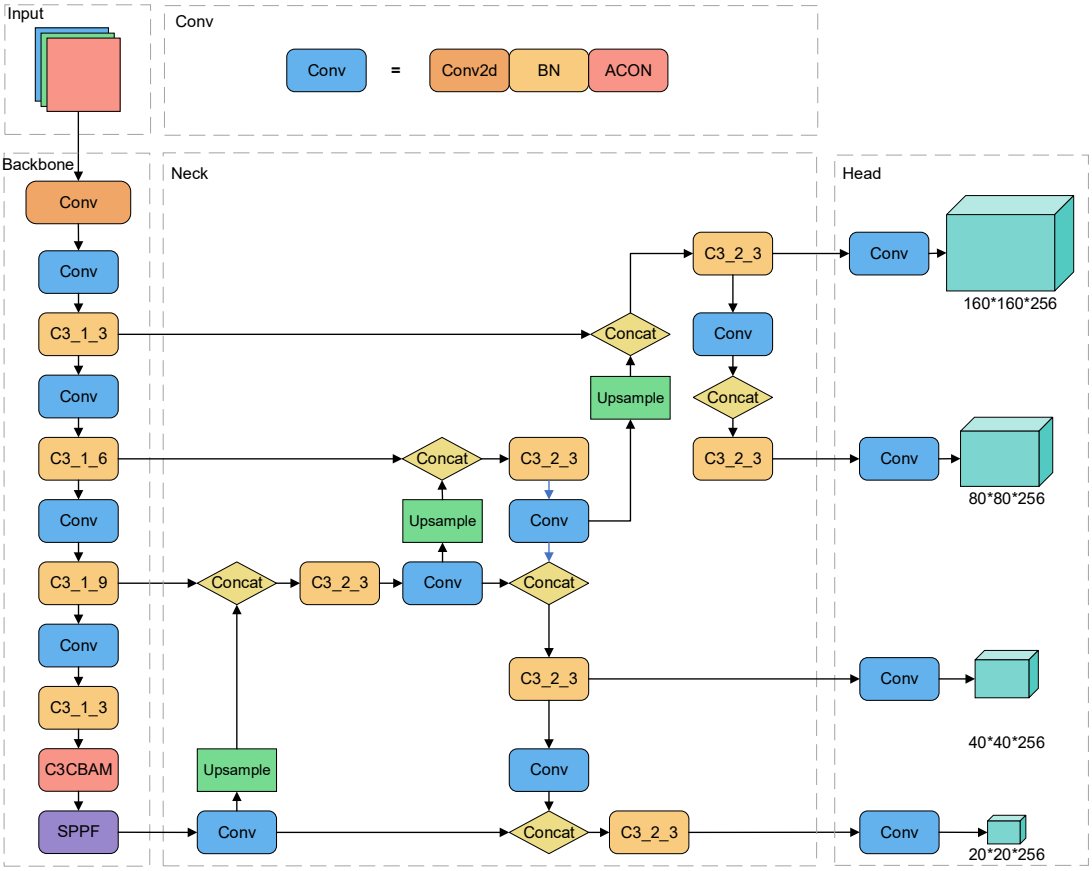


Fig.5 Network structure of Advanced YOLOv5x algorithm

4. EXPERIMENT AND ANALYSIS

4.1. Experimental platform and experimental environment

The experimental setup, as depicted in Fig.6, primarily comprises a desktop computer and the RealSense D435 depth camera. The desktop computer is equipped with the Anaconda software library and PyCharm compiler, facilitating the training and development of the object detection model. The

D435 depth camera is utilized to capture binocular images for conducting simulation experiments on the positioning of drill pipe joints.

Fig.6 Experimental platform

The environment configuration for the experiments is shown in Table 1:

Table 1. Experimental environment configuration

Name	Parameters
CPU	Inter（R）Core（TM）i9-10900XCPU@
RAM	64GB
GPU	NVIDIA GeForce RTX 2080Ti
Operating System	Windows10
Deep Learning Framework	Pytorch2.0.0 + GPU
Parallel Software Library	CUDA11.7 +cuDNN8.9
Programming Language	Python3.8

4.2. Detection experiment of drill pipe joints

4.2.1. Introduction of dataset

I personally curated this dataset comprising drill pipe joints, primarily sourced from field-captured images and downloaded from online platforms. The dataset encompasses drill pipe images of varying sizes, scenes, and orientations. It consists of 1300 static images, segregated into a training set and a test set with a ratio of 9:1, facilitating model training and evaluation. A selection of these datasets is illustrated in Fig.7.

Fig.7 Selection of dataset

The labelme labeling tool was used to label the data set, which contains more than 18,000 instances of manual labeling, including two categories of drill pipe upper and lower joints, which were labeled using the "inner joint" and "outer joint" labels, as shown in Fig.8.

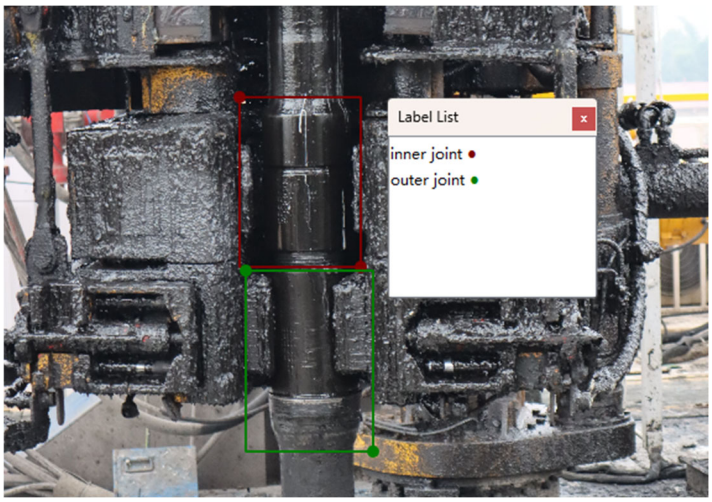


Fig.8 Examples of joint set

4.2.2. Algorithm Training

After building the drill pipe joint dataset, the drill pipe joint detection model was trained on the experimental platform, and the parameters of training were set as Table 2.

Table 2. Parameters of training

Parameters	Value
Learning rate	0.01
Batch size	8
Total epoch	300
Image size	640x640

4.2.3. The detection results

The trained drill pipe joint detection model was used to detect the video of the upper unbuckling operation being performed, and the results are shown in Fig.9. It can be clearly seen that the model is able to detect the drill pipe joints in real time and achieves the desired results.

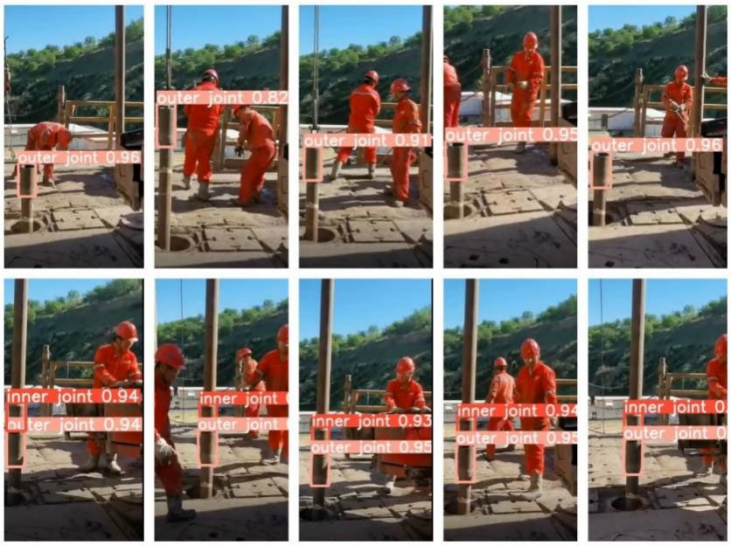


Fig.9 The detection results

4.3. Typical Evaluation Metrics

The main evaluation metrics selected for the experiments in this paper are Precision (P), Recall (R), Average Precision (AP) and mean Average Precision (mAP), with the formulas (16)-(19).

$$P = \frac{TP}{TP + FP} \tag{16}$$

$$R = \frac{TP}{TP + FN} \tag{17}$$

$$AP = \sum_{i=1}^{n} P(i)\Delta r(i) = \int_0^1 p(r)dr \tag{18}$$

$$mAP = \frac{\sum_{n=1}^{N} AP(n)}{N} \tag{19}$$

In the formula, TP (Ture Positive) represents correctly detected drill pipe joints, TN (Ture Negative) represents non-misdetected drill pipe joints, FP (False Positive) represents misdetected drill pipe joints, and FN (False Negative) represents non-detected drill pipe joints.

The PR curve shows the relationship between Precision and Recall of the model under different classification thresholds. The Area Under the Curve (AUC) of the PR curve can evaluate the overall performance of the model. Larger AUC values indicate that the model achieves higher Precision and Recall with better performance under all classification thresholds.

The mAP is the AP of all categories taken as an average, and the larger the mAP value, the better the result.

4.4. Analysis of results

4.4.1. Comparison of different algorithms

In order to more intuitively reflect the performance of improved YOLOv5x, the algorithms in this paper are compared with Faster R-CNN, SSD, YOLOv3, YOLOv4, YOLOv5 series and improved

YOLOv5x algorithms. 10 detection algorithms are trained and tested on the same experimental platform, and the comparison of algorithm results is shown in Table 3.

Table 3. Comparison of the detection effect of different algorithms

Model	mAP0.5	P	R	FPS / (frame·s^{-1})
Faster-RCNN	0.855	0.842	0.799	24.0
SSD	0.846	0.837	0.786	45.1
YOLOv3s	0.874	0.855	0.803	61.2
YOLO v4s	0.912	0.901	0.844	56.7
YOLO v5n	0.917	0.905	0.857	77.2
YOLO v5s	0.925	0.910	0.862	68.2
YOLO v5m	0.933	0.916	0.873	62.9
YOLO v5l	0.948	0.929	0.892	57.8
YOLO v5x	0.956	0.941	0.911	52.5
ImprovingYOLOv5	0.991	0.976	0.936	30.6

From the table, we can see that the improved YOLOv5 proposed in this paper has the highest mAP, which is 13.6% better than Faster-RCNN, 14.5% better than SSD, 11.7% better than YOLOv3, and 7.9% better than YOLOv4, which can verify the effectiveness of the improved YOLOv5 proposed in this paper.

4.4.2. Ablation experiments

To specifically verify the effect of each improvement step on the drill pipe joint identification model, ablation experiments were conducted and the results are shown in Table 4.

Table 4. Comparison of ablation experiment results

Model	Map0.5	P	R
YOLOv5x	0.956	0.956	0.887
YOLOv5x+ACON	0.959	0.934	0.910
YOLOv5x+CBAM	0.974	0.972	0.925
YOLOv5x+SIoU	0.960	0.958	0.896
YOLOv5x+ tiny targets predict head	0.976	0.974	0.928
YOLOv5x+ACON+CBAM	0.979	0.977	0.928
YOLOv5x+ACON+SIoU	0.980	0.976	0.936
YOLOv5x+ACON+ tiny targets predict head	0.983	0.982	0.941
YOLOv5x+CBAM+SIoU	0.981	0.980	0.938
YOLOv5x+CBAM+ tiny targets predict head	0.986	0.984	0.940
YOLOv5x+SIoU+ tiny targets predict head	0.984	0.982	0.943
YOLOv5x+all	0.991	0.989	0.951

The table shows the effect of each improvement method individually as well as their combined effect, with different improvement schemes improving the model metrics to different degrees. Among these improvement methods, adding a small target detection layer proved to be the most effective, which led to a 2% improvement in the model from the original one. Ultimately, the mAP of the model improved from 95.6% to 99.1%, an overall improvement of 3.5%.

4.4.3. Comparison of algorithms before and after improvement

4.4.3.1. Comparison of mAP

In this experiment, the YOLOv5x drill pipe joint recognition model before and after the improvement was trained for 300 rounds, and the mAP pairs are shown in Figure 10.

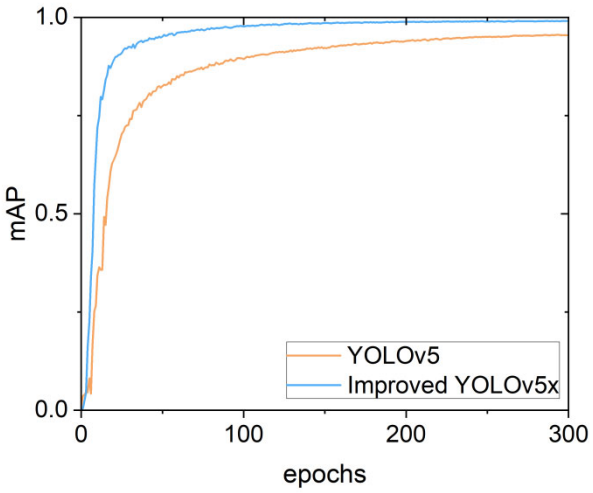
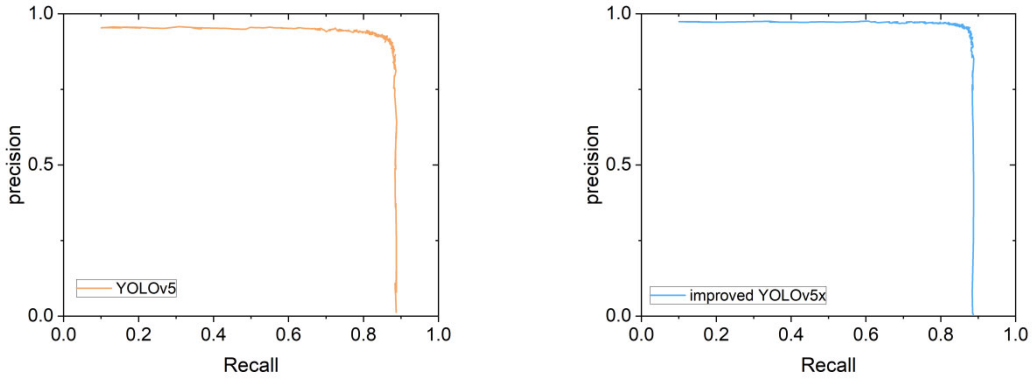


Fig.10 Comparison of mAP

The curves in the figure represent the changes of the average accuracy during 300 rounds of training. By comparison, the trends of the mAP curves of the two models are basically the same, and the curves rise rapidly in the first 50 rounds, and then gradually level off and finally reach the best mAP values. The mAP of the improved model is improved, indicating that the improvement scheme proposed in the paper is effective and feasible.

4.4.3.2. Comparison of PR Curves

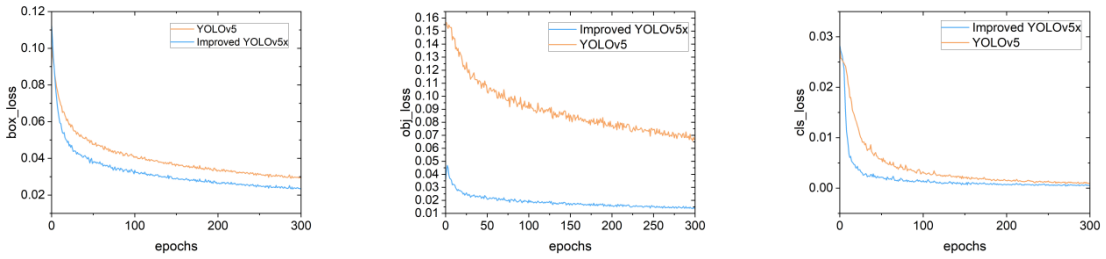


(a) YOLOv5 PR Curve (b) Improved YOLOv5 PR Curve

Fig.11 Comparison of PR curves

Figure 11 shows the PR curves of the YOLOv5x and the improved YOLOv5x drill pipe articulation head identification models. The comparison shows that the improved model has a larger AUC under the PR curve and has better performance.

4.4.3.3. Loss function graph



(a)box_loss (b)obj_loss (c)cls_loss

Fig.12 Comparison of loss function

Figure 12 shows the changes of the three types of loss functions for the YOLOv5x and the improved YOLOv5x drill pipe joint identification models over 300 training rounds. After the comparison, we found that the improved model has improved in both loss values. The final bounding box loss value decreased from 2.95e-2 before improvement to 2.34e-2, the target confidence loss decreased from 6.68e-2 before improvement to 1.39e-2, and the classification loss decreased from 9.58 e-4 before improvement to 5.54e-4. This indicates that the improvement to the YOLOv5x model is effective and the model's binary classification capability is improved.

5. CONCLUSIONS

In this paper, we propose an intelligent model for drill pipe joints detection and position, which is based on YOLOv5x with improvements to the backbone and head. The improved YOLOv5x can achieve great performance in both detection accuracy and speed. The proposed modifications are focused on three aspects: feature extraction, feature representation, and prediction output. The ACON activation function and the CBAM attention mechanism are employed to improve the learnt features. The SIOU function improves the flexibility of the model in representing bounding boxes. To increase the detection impact for tiny targets, the detection head is customized. Experimental results show that the proposed intelligent model performs admirably in terms of detection accuracy and location accuracy. The accuracy of the improved YOLOv5x model achieves 99.10%, which is 3.5% higher than the original YOLOv5x. The proposed model improves the efficiency of making up two joints of drill pipe and breaking out this connection by increasing the detection speed and positioning accuracy.

Future works can explore how to locate drill pipe joints in complex situations and improve technique performance such as offshore drilling platforms. The method can also be extended to other related tasks in the field of intelligent oil drilling equipment, such as the detection and tracking of other drilling components or the prediction of drilling performance.

REFERENCES

[1] Bello O, Holzmann J, Yaqoob T, et al. Application of artificial intelligence methods in drilling system design and operations: a review of the state of the art. Journal of Artificial Intelligence and Soft Computing Research 5(2) (2015): 121-139. https://doi.org/10.1515/jaiscr-2015-0024.

[2] Liu, Jia, and Yuqiang Wen. Analysis of the control strategy of the intelligent iron Roughneck's make-up and break-out fusion with fuzzy adaptive control algorithm. Journal of Physics: Conference Series 1992(3) (2021): 032099.

[3] Young, John. Iron roughnecks: Unmanned automation in the oil and gas industry. Chemistry in Australia (2020): 36-37.

[4] Z. Zou, K. Chen, Z. Shi, Y. Guo, et al. Object Detection in 20 Years: A Survey. Proceedings of the IEEE 111(3) (2023): 257-276.

[5] Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60, 91–110 (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94

[6] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," Proceedings. International Conference on Image Processing, Rochester, NY, USA, 2002, pp. I-I, doi: 10.1109/ICIP.2002.1038171.

[7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, doi: 10.1109/CVPR.2005.177.

[8] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. Deep learning. nature 521.7553 (2015): 436-444. https://doi.org/10.1038/nature14539.

[9] Mohammed, Ammar, and Rania Kora. A comprehensive review on ensemble deep learning: Opportunities and challenges. Journal of King Saud University-Computer and Information Sciences 35(2) (2023): 757-774. https://doi.org/10.1016/j.jksuci.2023.01.014.

[10] Redmon, Joseph, et al. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition (2016): 779-788.

[11] Liu, Wei, Anguelov, D., Erhan, et al. SSD: Single shot multibox detector. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 (2016): 21-37https://doi.org/10.1007/978-3-319-46448-0_2.

[12] Girshick, Ross, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition (2014): 580-587.

[13] Girshick, Ross. Fast R-CNN. Proceedings of the IEEE international conference on computer vision (2015): 1440-1448.

[14] Liang, B.; Wang, Z.; Si, L.; Wei, D.; Gu, J.; Dai, J. A Novel Pressure Relief Hole Recognition Method of Drilling Robot Based on SinGAN and Improved Faster R-CNN. Appl. Sci. 2023, 13, 513. https://doi.org/10.3390/app13010513

[15] A. Womg, M. J. Shafiee, F. Li and B. Chwyl, "Tiny SSD: A Tiny Single-Shot Detection Deep Convolutional Neural Network for Real-Time Embedded Object Detection," 2018 15th Conference on Computer and Robot Vision (CRV), Toronto, ON, Canada, 2018, pp. 95-101, doi: 10.1109/CRV.2018.00023.

[16] Ji, X., Gong, F., Yuan, X. et al. A high-performance framework for personal protective equipment detection on the offshore drilling platform. Complex Intell. Syst. 9, 5637–5652 (2023). https://doi.org/10.1007/s40747-023-01028-0