Research and Design of Industrial Transportation Oxygen Cylinder Detection and Tracking Based on Deep Learning

Minghao Ma, Shaochuan Xu and Dehua Liu

***Abstract*—To enhance the detection capabilities for identifying illegal operations by oxygen factory workers and abnormal conditions of oxygen cylinders, this paper proposes an algorithm based on YOLOv7 to improve target detection. The algorithm specifically addresses the need to check whether sufficient rubber rings are installed on the oxygen cylinders and to detect whether the oxygen cylinders are rolling on the ground. A small object detection layer was introduced to enhance the capture of shallow features. Additionally, the accuracy was improved by incorporating the CBAM Attention Mechanism module. The actual field dataset was utilized for ablation and comparison tests. Safety inspections were conducted through target detection and tracking to timely identify illegal and abnormal operations, thereby avoiding repeated alarms for the same target. The issues of misjudgment and missed detection of small targets, such as rubber rings, were resolved. This significantly improved detection accuracy, reduced labor costs, and provided assurance for the safety of the factory production process.**

***Index Terms*—Oxygen cylinder, Rubber ring, Target detection, Target tracking**

I. Introduction

The shock-proof ring on the oxygen cylinder, made of elastic rubber, is essential for cushioning impacts and protecting the cylinder's paint layer. Each oxygen cylinder needs to be fitted with two rubber rings. In actual production processes, some people often remove safety devices such as shock absorbers and caps from oxygen cylinders for transportation convenience, which poses significant safety hazards. Oxygen cylinders are high-pressure containers; without shock absorbers, they are vulnerable to external impacts or vibrations during transportation or use. This can cause deformation, rupture, or leakage of the cylinder, rendering it unusable and potentially leading to serious safety accidents. Additionally, in oxygen cylinder factories, some workers transport oxygen cylinders upside down to save effort, which is another unsafe practice that needs real-time detection and warnings. Given their critical role in both daily life and medical care, the loading and unloading of oxygen cylinders must be strictly controlled to prevent safety hazards.

Manuscript received January 23, 2024; revised January 9,2025.

Minghao Ma is a Postgraduate Student of School of Electronic Information, University of Science and Technology Liaoning, Anshan, 114051 China. (e-mail: mmh13591920408@163.com).

Shaochuan Xu is a Professor of School of Control Science and Engineering, University of Science and Technology Liaoning, Anshan, 114051 China. (Corresponding author, phone: 86-0412-5929747; e-mail: shaochuanxu1@163.com).

Dehua Liu is a Bachelor's graduate from Anshan Liaoning University of Science and Technology in automation in 2010, 114051 China. (e-mail: liudehua0527@163.com).

II. Analysis of Cylinder Transportation

By adjusting the appropriate camera angle, we can see the entire process of workers transporting oxygen cylinders to the workshop [1]. The standard oxygen cylinder body should have two rubber rings to prevent collisions, impacts and other situations during transportation, in order to reduce safety hazard. A qualified transportation method for oxygen cylinders should be to hold them with one hand and let the bottom of the oxygen cylinder touch the ground, continuously rotating them for transportation as shown in Fig. 1. Overturning oxygen cylinders for transportation is a violation of regulations and poses significant safety hazards, as shown in Fig. 2.

Fig. 1. Standard transport oxygen cylinders

Fig. 2. Oxygen cylinder rolling on the ground

This article primarily focuses on detecting illegal operations and abnormal conditions of workers in an oxygen factory. This includes checking whether sufficient rubber rings are installed on oxygen cylinders and whether the cylinders are rolling on the ground. Difficulty analysis: To determine if a sufficient number of rubber rings are installed on an oxygen cylinder, it is necessary to identify and assess

the positions of both the rubber rings and the oxygen cylinders.

Another issue is that rubber rings are small targets, posing specific challenges for detection. Below is a detailed introduction to the main challenges faced in small object detection:

(1) Low Resolution and Limited Feature Information: Small objects often have low resolution and limited feature information, compounded by their small size, low contrast, and complex backgrounds. These factors make it difficult to accurately detect and locate small targets.

(2) Target and Background Similarity: Small targets frequently resemble the background or other objects, making it challenging to distinguish them. Similarities in color, texture, shape, and other features between the target and the background can lead to false detections or missed detections. For example, in an oxygen factory, rubber rings appear as black strips on camera, which can be easily confused with other similar objects such as black railings, leading to misjudgment.

(3) Scale Variation: The size of small targets in an image can be affected by factors such as perspective and distance, resulting in significant scale differences. This makes it challenging to detect small targets of varying scales. When detecting rubber rings, the distance between the ring and the camera can vary, causing significant size variations in the captured image. Consequently, the rubber ring may become undetectable from certain distances or angles.

This results in significant differences in the scale of various small targets, complicating their detection. When detecting rubber rings, the distance between the ring and the camera can vary, with some being far away and others being close. Consequently, the size of the rubber ring in the camera's view may vary significantly, potentially making the rubber ring undetectable.

III. Model Selection and Improvement

Due to the varying positions of oxygen cylinders in each frame of the video and the presence of multiple cylinders in the image, an object detection algorithm is well-suited for these scenarios where the position is not fixed and the target needs classification. Mainstream object detection algorithms include the R-CNN series and YOLO series. The R-CNN series consists of R-CNN, Fast R-CNN, and Faster R-CNN, while the YOLO series includes YOLOv3, YOLOv4, YOLOv5, YOLOv7, among others.

This article selects the YOLO series single-stage object detection model to detect oxygen cylinders and rubber rings, and selects the appropriate model by comparing the accuracy and recall parameters of each model [2]. First, make a data set of 2500 pictures for the evaluation of the model, and mark the oxygen cylinder and ring. Train the dataset using YOLOv4, YOLOv5 and YOLOv7 respectively. By comparing the mAP@0.5 of these three models, it is evident that YOLOv7 performs better [3]. In this article, it is necessary to detect the rubber rings on the oxygen cylinder. If there are omissions and misjudgments, there is a high possibility of causing missed and erroneous alarms. Fig. 3. and Fig. 4. and Fig. 5. show the P-R curves for YOLOv4, YOLOv5, and YOLOv7, respectively, and the precision and recall corresponding to the points on these three P-R curves are the average of all kinds of precision and the average of all kinds of recall, respectively.

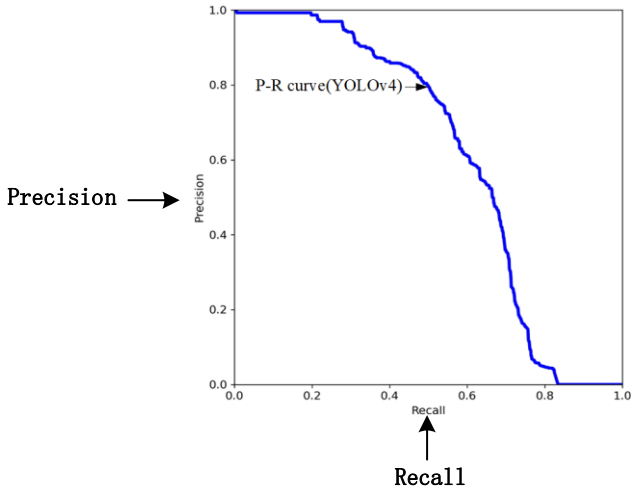

Fig. 3. P-R curve of YOLOv4

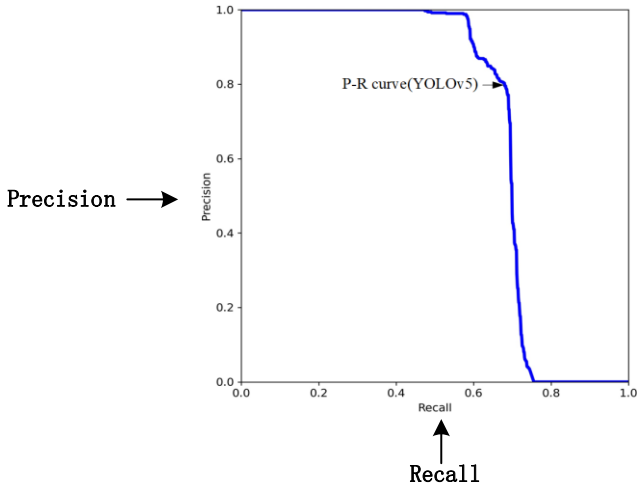

Fig .4. P-R curve of YOLOv5

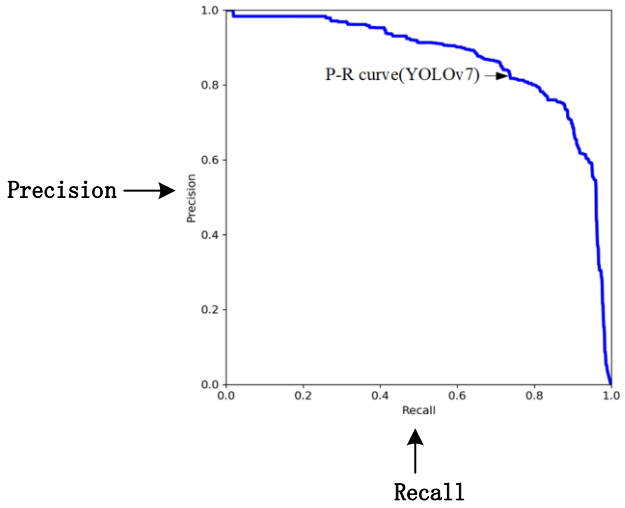

Fig .5. P-R curve of YOLOv7

A. YOLOv7 object detection algorithm with added small object detection layer

The YOLO algorithm belongs to the one stage object detection algorithm, it is based on deep neural networks for object classification and localization, it is divided into three

parts: Backbone, FPN and YOLO Head. Among them, the role of Backbone is to extract the features of the input image, and Backbone will output three feature layers of different sizes. The FPN network is used to enhance feature extraction. It partially fuses the three feature layers extracted by Backbone. Up-sampling and down-sampling are performed during the fusion process. YOLO Head is used for classification and regression, outputting the final prediction box [4].

Due to the rubber ring is a small object, it occupies a relatively small number of pixels in the image. After continuous convolutional down-sampling, the semantic features of the rubber ring in the image will become fewer. When the oxygen cylinder is far from the camera, the already small rubber ring will become smaller in the surveillance video. Using the YOLOv7 to detect the rubber rings on oxygen cylinders, they are far from the camera may result in undetectable or misjudgment situations, and the confidence level of the rubber rings that can be detected is also very low. When the confidence threshold is set high, more rubber rings will not be detected. Therefore, this section improves the YOLOv7 deep learning model [5], which draws on the structure of feature pyramids [6].

Introduction to Feature Pyramid: In a target detection task, the target may be anywhere in the image and may vary in size. By using a feature pyramid, the target can be detected at different scales, thus improving the accuracy and robustness of the detection. Due to the small proportion of pixels occupied by small objects in the image, a lot of pixel information of small objects will be lost during the continuous down-sampling process. The feature pyramid integrates high-level features with low-level features and independently predicts multi-scale features, significantly improving the detection performance of small objects.

In summary, this section will make small-scale modifications to the YOLOv7 model by adding a detection layer for small targets [7]. The modified model is shown in Fig.6. and the dashed box represents the changes made to the model.

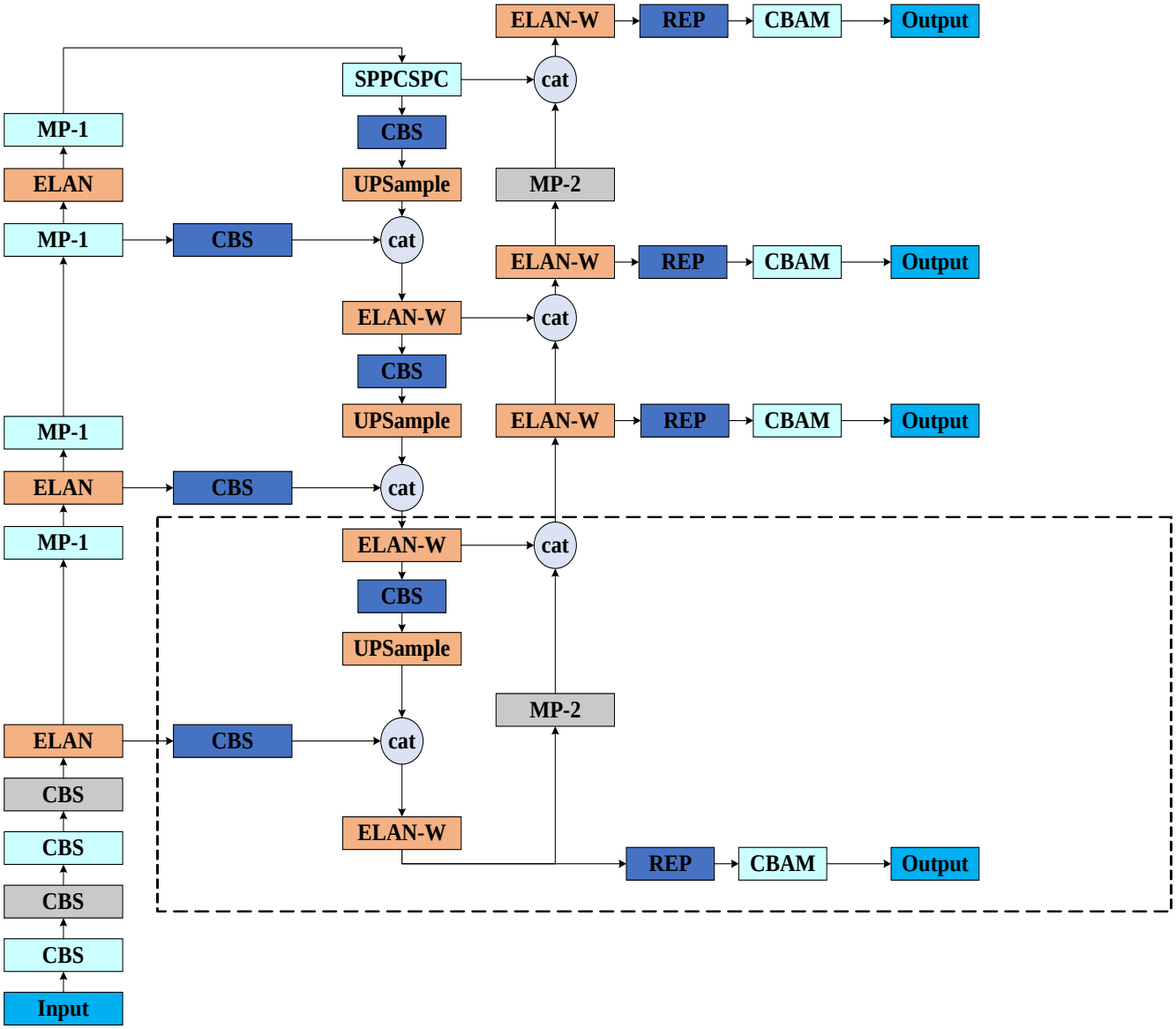

Fig. 6. Structure diagram of YOLOv7 after adding a small target detection layer

B. Training Process and Result Analysis

Before incorporating the training images into the YOLOv7 model, a series of data augmentation processes were performed on the images, including image scaling, Mosaic processing, deformation processing and random flipping. Mosaic processing is the process of concatenating multiple images to avoid the object being detected being located in the center area of the image and to reduce the impact of the target's background on the detection results. Deformation processing includes processing methods such as rotation, translation, scaling and shearing. The above data augmentation methods incorporate perturbations into the images, enabling the trained model to better adapt to various scenarios and changes in the actual production process. By generating target samples at different scales, angles and environmental conditions, we can increase data diversity and effectively prevent overfitting.

Train the divided dataset into 300 epochs, with a batch size of 32, an input image size of 640x640, a learning rate of 0.001 and pre training weights set to the weight of the coco dataset. Setting pre training weights can accelerate the convergence speed of the model.

Train the YOLOv7 model and test it on the test set. The results are shown in TABLE I, where the true values of the oxygen cylinders and rubber rings refer to the total number of labeled oxygen cylinders and rubber rings on the oxygen cylinder in the test set image. The detection value of oxygen cylinders and rubber rings refers to the total number of rubber rings on the oxygen cylinders detected using the YOLOv7 model. The recognition rate is the ratio of the detected value of a class of samples to the true value of that class of samples. From the table, it can be seen that the recognition rate of oxygen cylinders is very high, but the recognition rate of rubber rings is not high, and many rubber rings cannot be detected.

TABLE I
TEST RESULTS

	Oxygen cylinder	Rubber ring
True Value	4385	7216
Detection Value	4006	6018
Recognition Rate	91.35%	83.39%

The oxygen cylinder is inclination in the image, which causes the prediction box to contain other oxygen cylinders. In this case, it is difficult to determine whether there is a rubber ring on each oxygen cylinder. This leads to misjudgment when detecting the rubber ring, and many rubber rings have a low confidence level. This phenomenon indicates that YOLOv7 has poor detection performance for small targets such as rubber rings. Therefore, if a higher confidence threshold is set, it may cause some rubber rings to be undetectable, resulting in an error alarm.

As shown in Fig. 5. During the oxygen cylinder detection process, because the oxygen cylinders are placed adjacent to each other and some of the oxygen cylinders are tilted in the image, the oxygen cylinder prediction box contains other oxygen cylinders. There may be misjudgments and omissions when using YOLOv7 to detect oxygen cylinders. Therefore, a new data set is remade for the scattered oxygen cylinders, as shown in Figure 4, and the fully displayed oxygen cylinders are marked.

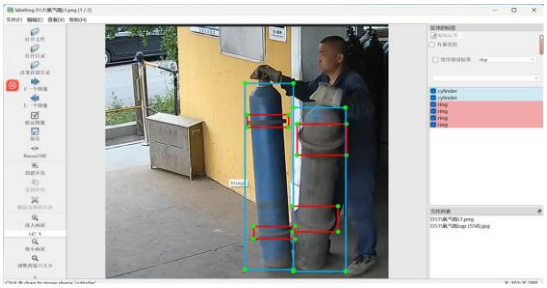

Fig. 5. Schematic diagram of oxygen cylinder

C. Target Detection of Rubber Rings and Oxygen Cylinders

After adjusting the dataset annotation and adding a small object detection layer [7], the newly established dataset was trained and the results are shown in Fig. 6.

Fig. 6. Detection results of the trained model on the test set

D. Falling Oxygen Cylinder Detection

An oxygen cylinder weighs approximately 60-70 kilograms, so workers sometimes choose a labor-saving way to roll the cylinder onto the ground and into the oxygen workshop during transportation. When an oxygen cylinder rolls on the ground, it is easy to collide with other objects, leading to oxygen leakage and safety accidents. we need to promptly detect and alarm the oxygen cylinders transported upside down to alert workers and correct them in a timely manner [8].

After solving the problem of oxygen cylinder detection, we began to check whether the oxygen cylinder was rolling on the ground. The methodology used in this paper is to separate oxygen cylinders rolling on the ground into a separate category when creating the dataset. The YOLOv7 model is then used for training. The detection of the trained model on the test set is shown in Fig. 7. It can detect oxygen cylinders rolling on the ground [9].

Fig. 7. Detection of fallen oxygen cylinders

After adding a small object detection layer, train the newly established dataset and use the test set for testing. The detection results are shown in TABLE II. Compared with the detection results in TABLE II, the recognition rate of the rubber ring has been greatly improved, which can greatly reduce the probability of false alarms.

TABLE II
DETECTION RESULTS AFTER ADDING SMALL TARGETS

	Oxygen cylinder	Rubber ring
True Value	4385	7216
Detection Value	4011	6572
Recognition Rate	91.47%	91.07%

The detection results of the trained model for the images are shown in Fig. 6. There was no misclassification and the confidence level was high [10]. This achieves the goal of effectively detecting small targets such as rubber rings. The current model can be used to detect oxygen cylinders and rubber rings and make subsequent judgments.

E. Introduction of Attention Mechanism

When performing deep clustering, the original YOLOv7 network loses information about the rubber rings of small target. Thus, by introducing the CBAM attention mechanism, The link between the location information in the image and the location information of the rubber ring features is further strengthened, and the feature expression of important information in the network is enriched. The network structure is shown in Fig. 8. The CBAM attention mechanism combines the channel attention module and the spatial attention module to process the input character layer of the channel attention module. The input character layer of the channel attention module and the spatial attention module are processed separately.

As shown in Fig. 9. the channel attention module first uses the spatial information of the average pooling and max pooling operating polymers before processing the input character layer. the spatial information of the maximum pooling operating polymer character to generate the average pooling characteristics F_{avg}^{C} and F_{max}^{C}. These two spatial descriptors are then forwarded to the shared network to generate the Channel Attention Map $M_c \in R^{C\times1\times1}$.. The shared network consists of a multi-sensor (MLP) containing a hidden layer, and the hidden activation size is to reduce parameter calls, the hidden activation size is set to $R^{C/r\times1\times1}$ where r is the reduction rate. The computation process is shown in equation 1, where b stands for the Sigmoid function, $W_0 \in R^{C\times C/r}$, $W_1 \in R^{C\times C/r}$ and the MLP is weights are shared by $W_0$and W_1, and W_0 is activated by the previous ReLU function [10].

$$M_C(F) = \sigma\left(MLP\left(AvgPool(F)\right) + MLP\left(MaxPool(F)\right)\right) = \sigma(W_1(W_0F_{avg}^{c})) + W_1(W_0(F_{max}^{c}))) \quad (1)$$

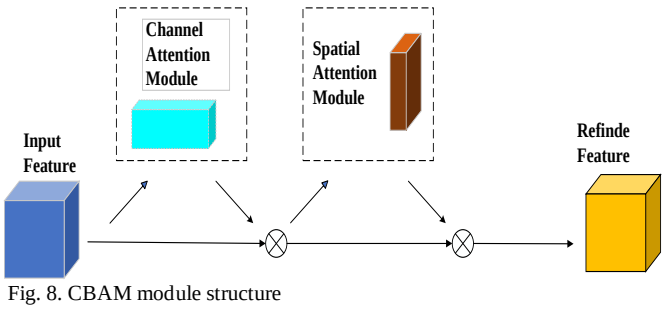

Fig. 8. CBAM module structure

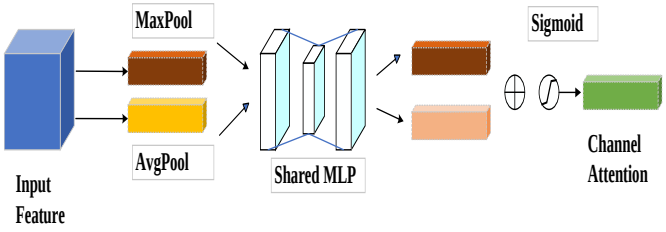

Fig. 9. Channels attention module structure

Unlike channel attention modules, the space attention module focuses on which part of the input information is more important, and its structure is as shown in Fig. 10. First, through two pooling operations, the channel information of a characteristic chart is aggregated and two mappings are generated: $F_{avg}^{s} \in R^{1\times H\times W}$ and $F_{max}^{s} \in R^{1\times H\times W}$, which represent the average pooling characteristics of the channels and the max pooling characteristics respectively. Finally, standardization is done with the Sigmoid function [10]. The specific calculation equation is as shown in the equation 2, in which b represents the Sigmoid function, and $f^{7\times7}$ represents a compact nucleus with a size of 7 × 7.

$$M_s(F) = \sigma\left(f^{7\times7}([AvgPool(F); MaxPool(F)])\right) = \sigma(f^{7\times7}([F_{avg}^{s}; F_{max}^{s}])) \quad (2)$$

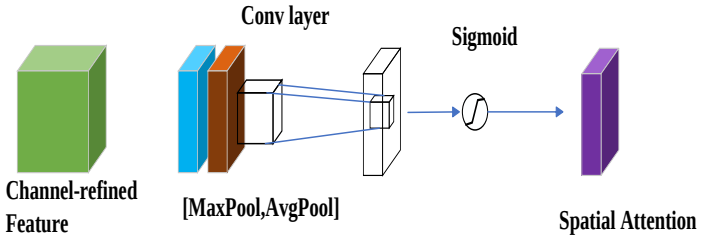

Fig. 10. Spatial attention module structure

F. Indicators of Evaluation

In order to measure the performance of the algorithm, the detection precision, recall rate, average precision mAP is used as the criterion for measuring the performance, $mAP^{0.5:0.95}$ represents the total average value of mAP when the threshold is between 0.5 and 0.95, the calculation can be calculated using the following equations.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$mAP = \frac{1}{C}\sum_{i=1}^{c} AP_i \quad (5)$$

In the equations, the TP (True Positives) represents the positive sample has been correctly identified; the FN (False Negatives) represents the negative sample is erroneously recognized as negative; and the FP (False Positive) represents negative samples have been incorrectly recognized as positive. AP represents the average precision on a single category; C represents the number of categories.

G.Ablation Experiment

Ablation experiments are conducted on the same dataset for each improved module of YOLOv7 to test the effect of each improved module. The experiments are based on YOLOv7 with the introduction of small target detection layer, and the introduction of the CBAM attention mechanism, and "+" indicates that an improvement of a module is added, as shown in TABLE III.

TABLE III
ABLATION EXPERIMENT RESULTS

	mAP@0.5	Cylinder precision	Rubber ring precision
YOLOv7	87.2%	90.1%	83.4%
YOLOv7+Detection Layer	89.7%	90.7%	90.5%
YOLOv7+CBAM	90.8	94.2%	87.3%
Ours	93.5%	94.3%	93.1%

The P-R curve for the final improved model training is shown below. The mAP value of the model at this point is 0.935. The YOLOv7 deep learning model is used to train the divided dataset, the training is divided into 200 epochs, the batchsize is set to 32, the input image size is 640x640, the learning rate is 0.001, and the pre-training weights are set to the weights of the coco dataset. Then the training starts, after the model is trained to 150 epochs, the weight file is saved every 10 epochs of training, while the optimal weight file is constantly updated.

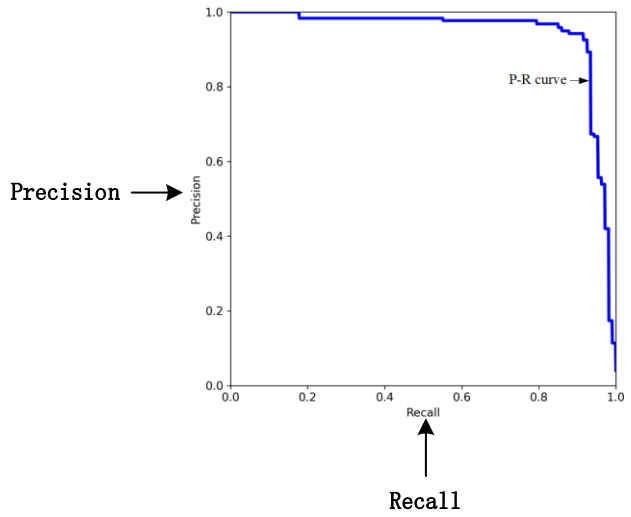

Fig. 11. P-R curve

IV. Solutions for Duplicate Alarm Problems

This article is to detect the oxygen cylinder and rubber ring in the video, an oxygen cylinder will be detected many times from entering the camera range to leaving the camera range. If sufficient rubber rings are not installed on this oxygen cylinder, there will be a phenomenon of repeated alarms for one oxygen cylinder, so it is necessary to propose solutions to this phenomenon of repeated alarms. This article uses a target tracking algorithm to solve this problem. The target tracking algorithm generates the same ID for the same object in different frames, which is referred to as the tracking ID in this article. By judging the tracking ID, the same oxygen cylinder that needs to be alerted in different frames will only be alerted once.

This article combines the DeepSORT object tracking algorithm with YOLOv7 to solve the problem of repeated alarms [11]. DeepSORT is a deep learning based on object tracking algorithm. It can accurately track multiple targets and perform identity association in videos. The core idea of DeepSORT is to use deep learning models to extract features from targets and use Kalman filters to predict the position and velocity of targets. By matching feature vectors with historical trajectories, DeepSORT can track the same target in multiple frames and assign a unique identifier to each target.[12]. The detection effect after combining YOLOv7 with DeepSORT is shown in Fig. 12. Each oxygen cylinder has a unique tracking ID, and the tracking ID of the same oxygen cylinder remains unchanged until the target disappears from the surveillance camera. The target tracking algorithm generates the same ID for the same object in different frames, which is referred to as the tracking ID in this paper. The judgment of the tracking ID makes the same oxygen cylinder that needs to be alarmed in different frames to be alarmed only once.

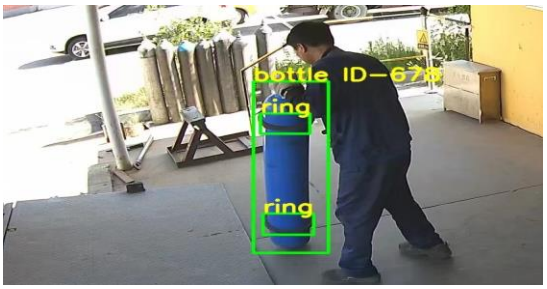

Fig. 12. Detection results after adding target tracking algorithm

Method for determining whether an oxygen cylinder is equipped with sufficient rubber rings and whether an alarm is required.

The preparation works to be done before judgment is as follows: Obtain the predicted information of the oxygen cylinder and rubber ring in the image through the YOLOv7 model, where the predicted box information includes the coordinates of the upper left corner and the lower right corner of the predicted box. Divide predicted information into two parts, one is the predicted information of the oxygen cylinder and the other is the predicted information of the rubber ring. Convert the predicted box information of the rubber ring from the form of upper left corner coordinates and lower right corner coordinates to the form of center point coordinates and predicted box width and height. Put the predicted information of the oxygen cylinder into the DeepSORT model and then obtain the predicted information of the oxygen cylinder and its corresponding tracking ID. Set up a container called All-Check to store the tracking ID of the oxygen cylinder that has already been alerted.

The determination method for whether an oxygen cylinder is equipped with sufficient rubber rings and whether an alarm is required is shown in Fig. 12. Obtain the predicted information and tracking ID of an oxygen cylinder in the image, as well as the predicted information of all rubber rings in the image, and determine whether the number of rubber ring center points contained in the predicted box of the oxygen cylinder is greater than or equal to two. If it is less than two, it indicates that the oxygen cylinder is not equipped with enough rubber rings. To prevent duplicate alarms, it is necessary to determine whether the tracking ID of the oxygen cylinder that needs to be alarmed exists in the All-Check

container. If it does, no alarm will be issued. If it does not exist, the tracking ID of the oxygen cylinder will be placed in All-Check and an alarm will be issued.

DeepSORT uses the Duke University reid dataset for pre-training, and uses exponential moving average to update the appearance state of the i trajectory at frame t. The appearance state of the i trajectory at frame t is updated using an exponential moving average. In equation 6, f_i^t is the appearance embedding of the current matching detection, and α is the momentum term.

$$e_i^t = \alpha e_i^{t-1} + (1-\alpha) f_i^t \quad (6)$$

The NSA Kalman algorithm is used, which presents formulas for adaptively calculating the noise covariance. In equation 7 is a preset constant measurement noise covariance, and is the detection confidence score for state k.

$$\tilde{R}_k = (1-c_k) R_k \quad (7)$$

Matching is performed using appearance feature distances and motion information. The cost matrix in equation 8 is the weighting of the appearance model A_a and motion model cost A_m. λ is the weighting factor.

$$C = \lambda A_a + (1-\lambda) A_m \quad (8)$$

The detection effect after combining Yolov7 with DeepSORT is shown in Fig.12, where each oxygen cylinder has a unique tracking ID and the tracking ID of the same oxygen cylinder does not change until the target disappears from the surveillance camera.

The formula for calculating the false alarm rate is shown in Equation 1. From the above data, it can be believed that the determination method for whether sufficient rubber rings are installed on the oxygen cylinder, and whether an alarm is needed can effectively determine whether sufficient rubber rings are installed on the oxygen cylinder [13]. In addition, the true value of the number of alarms refers to the actual number of alarms that need to be triggered, and the detection value of the number of alarms refers to the number of alarms generated according to the judgment rules. It can be believed that the detection value of the number of alarms is the same as the detection value of the number of oxygen cylinders with insufficient rubber rings, proving that the target tracking algorithm has played a role in preventing duplicate alarms.

$$fa = \frac{t_n}{t_a} \quad (9)$$

In equation 9, fa is the false alarm rate, t_n is the number of false alarms, that is, the detection value of the number of alarms minus the true value of the number of alarms, t_a is the true value of the total number of oxygen cylinders.

This section uses multiple videos for testing. As shown in TABLE IV, the true value of the total number of oxygen cylinders in the table represents the total number of oxygen cylinders, they are not densely arranged and displayed as complete in the video. The true value of the total number of rubber rings on the oxygen cylinder is the total number of rubber rings on all non-densely arranged oxygen cylinders displayed in the video. From TABLE III, it can be believed that one oxygen cylinder was missed and 89 rubber rings were missed. This resulted in 56 more oxygen cylinders with insufficient rubber rings being detected, while 57 oxygen cylinders with sufficient rubber rings were missed, resulting in a false alarm rate of 1.9%.

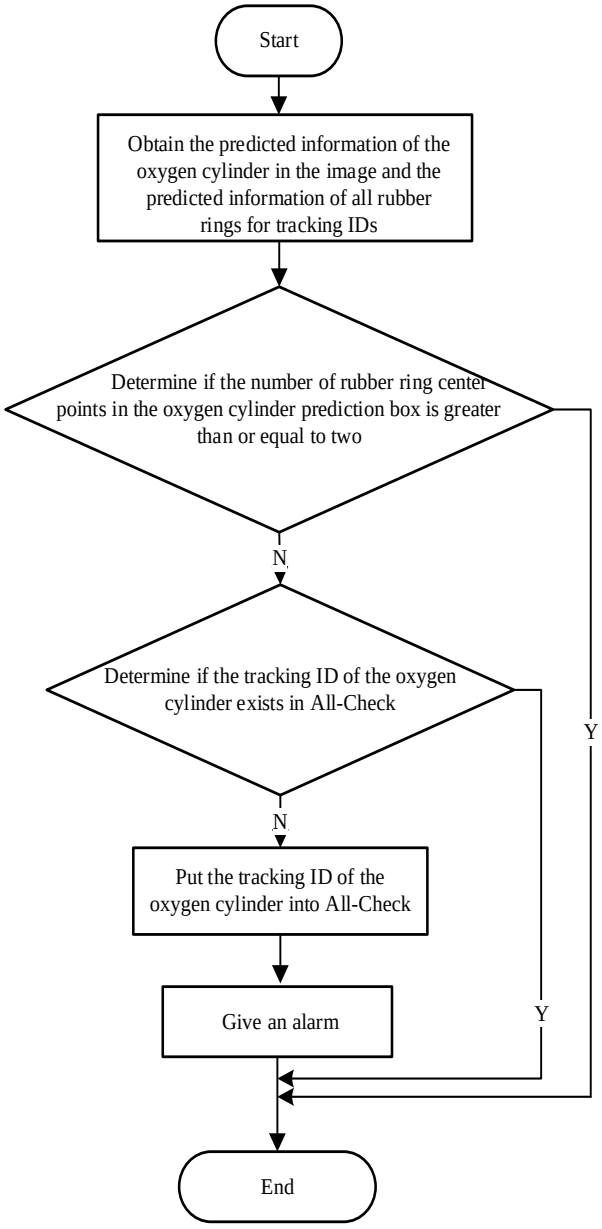

Fig. 12. Determination method for whether sufficient rubber rings are installed on oxygen cylinders and whether an alarm is required

TABLE IV
TEST RESULTS OF JUDGMENT METHODS

	Total number of oxygen cylinders	The total number of rubber rings	The number of oxygen cylinders with insufficient rubber rings	Sufficient rubber rings for oxygen cylinders	Alarms
True value	3016	4301	1298	1958	1298
Detection value	3015	4212	1354	1901	1354
False alarm rate			1.9%		

V. System Overview

This system uses a camera for image acquisition, uses OpenCV library functions to establish a connection with the camera through an RTSP address and then performs video decoding to obtain real-time image information. For the detection system of oxygen cylinders and rubber rings, Python is used to preprocess the images and then the processed images are placed in a deep learning model to obtain the predicted box information of the target on the image. As only the oxygen cylinder needs to be tracked, only the predicted box information of the oxygen cylinder is placed in the DeepSORT detector for target tracking, obtaining the anchor box information of the target and the tracking ID corresponding to the oxygen cylinder anchor box, determine whether an alert is issued by judging and analyzing the obtained tracking ID. This system uses the SE5 box as the hardware platform for algorithm deployment and the BM1684 chip, supporting up to 17.6T of computing power.

A. Video Decoding

Decoder parses the input video. Secondly, frame decoding is performed, and the decoder can directly decode the I frame to obtain the complete image [15]. For frames P and B, the decoder needs to perform motion compensation and reconstruction to obtain the complete image. Then, color space conversion is performed to ensure that the decoded image matches the display format of the device. Finally, perform some filtering, denoising, and other operations on the image. In this article, RTSP addresses are used to obtain surveillance video streams and then perform decoding operations.

B. Model Quantification

Quantification reduces model size and computational power requirements by reducing the number of weights and activation values. Quantization operations generally involve converting a model from a high-precision floating-point number to a high-precision integer. In this article, the model is converted from a 32-bit floating-point number to an 8-bit integer to reduce computational complexity, improve inference speed, and reduce device energy consumption, allowing the model to be deployed in embedded devices.

C. Introduction to System Modules

The functions that each module can achieve are as follows:

(1) Image acquisition module

Obtaining the video stream through the RTSP address [16], decode it and connect the camera, SE5 box, and PC to the same network segment to ensure a connection is established between the three [17]. Finally, you can get the video stream by putting the RTSP address into the specified location in the programs and then decode it.

(2) Object detection and tracking module

This system uses a combination of YOLOv7 object detection model and DeepSORT object tracking model for the detection of oxygen cylinders and rubber rings. This module is used to receive decoded video images. After processing the images, the tracking ID of the oxygen cylinder and the predicted information of the oxygen cylinder and rubber ring are obtained. The tracking ID and predicted box of the oxygen cylinder and rubber ring are drawn on the image. Then, it determines whether the oxygen cylinder is installed with sufficient rubber rings and whether the oxygen cylinder is rolling on the ground. Finally, based on the tracking ID, it determines whether an alarm is needed, Notify the automatic alarm module to issue an alarm when an alarm is needed. The overall process is shown in Fig. 13. The image of the drawn prediction box generated throughout the entire process will be displayed on the human-computer interaction interface. In addition, this module will restart regularly, and after the restart, the container storing the alarm oxygen cylinder tracking ID will be emptied, and the tracking ID will start to grow from 1.

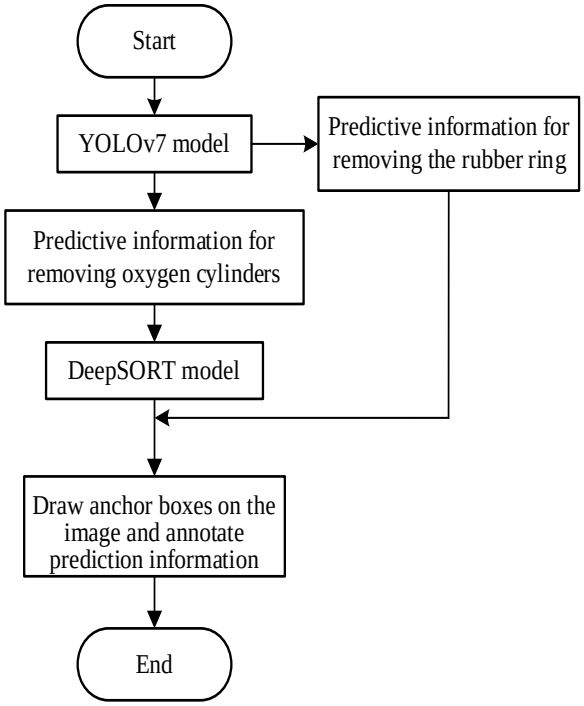

Fig. 13. Flow Chart of Object Detection and Object Tracking Module

(3) Automatic alarm module and human-machine interaction interface

This module sends an alarm upon receiving confirmation signals from the target detection and tracking modules and when the alarm is triggered, the indicator light on the PyQt interface will light up. In order to facilitate the use of this system by users, PyQt is used to create a human-machine interaction interface, allowing staff to have a more intuitive understanding of the on-site situation and quickly detect alarm information. When an alarm occurs, the alarm image will be saved to ensure that users can trace the source of abnormal conditions on site.

This article designs login interface, oxygen cylinder and rubber ring recognition interface, oxygen cylinder tracking interface, alarm log interface and programs based on the functions to be implemented in each interface to make the detection system more complete. The interface design for identifying and tracking oxygen cylinders for industrial transportation is shown in the Fig. 14.

Fig. 14. Safety production warning system

VI. Conclusion

This article mainly focuses on detecting abnormal conditions in oxygen factory. Based on the YOLOv7 algorithm, detect whether sufficient rubber rings are installed on the oxygen cylinder and whether the oxygen cylinder is rolling on the ground. By adding a small object detection layer on the basis of YOLOv7, the detection problem for small objects such as rubber rings has been solved; The problem of repeated alarms is solved by using the target-tracking algorithm; Determine whether there is sufficient rings on the oxygen cylinder by writing the post-processing procedure; Design a detection system using trained models and select hardware and software platforms for the system.

References

[1] W. Wang, S. Xu, Y. Teng, "Design of Belt Sprinkler Monitoring System Based on Image Processing Technology," *IAENG Internaitonal Journal of Computer Science*, vol. 49, no. 1, pp. 94-100, 2022

[2] C. Y. Wang, A. Bochkovskiy, H. Y. M. Liao, "YOLOv7: Trainable Bag-of-freebies Sets New State-of-the-art for Real-time Object Detectors," *Institute of Information Science*, pp. 7464-7475, 2023.

[3] K. He, X. Zhang, S. Ren, J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904-1916, 2015

[4] J. Redmon, A. Farhadi, "YOLO9000: Better, Faster, Stronger," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6517-6525, 2017.

[5] X. K. Zhu, S. C. Lyu, X. Wang, Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," *IEEE/CVF International Conference on Computer Vision Workshops*, pp. 2778-2788, 2021.

[6] H. M. Hossein, M. Hadis, "Fine-tuned YOLOv5 for Real-time Vehicle Detection in UAV Imagery: Architectural Improvements and Performance Boost," *Expert Systems with Applications*, vol. 231, no. 30, pp. 120845, 2023.

[7] T. Y. Lin, P. Dollár, R. Girshick, K. M. He, B. Hariharan, S. Belongie, "Feature Pyramid Networks for Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 936-944, 2017.

[8] J. Dang, X. F. Tang, S. Li, "HA-FPN: Hierarchical Attention Feature Pyramid Network for Object Detection," *Sensors (Basel, Switzerland)*, vol. 23, no. 9, pp. 4508, 2023.

[9] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You Only Look Once: Unified, Real-time Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788, 2016.

[10] Y. P. Liu, B. C. Jiang, H. He, Z. J. Chen, Z. F. Xu, "Helmet Wearing Detection Algorithm Based on Improved YOLOv5," *Scientific Reports*, vol. 14, pp. 8768, 2024.

[11] A. Krizhevsky, I. Sutskever, G. Hinto, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.

[12] Y. H. Du, Z. C. Zhao, Y. Song, Y. Y. Zhao, F. Su, T. Gong, H. Y. Meng, "Strongsort: Make Deepsort Great Again," *IEEE Transactions on Multimedia*, vol. 25, pp. 8725-8737, 2023.

[13] C. Szegedy, W. Liu, Y. Q. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going Deeper with Convolutions," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1-9, 2015.

[14] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *CoRR*, 2014.

[15] B. Bross, Y. K. Wang, Y. Ye, S. Liu, J. L. Chen, G. J. Sullivan, J. R. Ohm, "Overview of the Versatile Video Coding (vvc) Standard and Its Applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3736-3764, 2021.

[16] G. J. Sullivan, J. R. Ohm, W. J. Han, T. Wiegand, "Overview of the High Efficiency Video Coding (hevc) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649-1668, 2012.

[17] T. Wiegand, G. J. Sullivan, G. Bjontegaard, A. Luthra, "Overview of the h. 264/avc Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560-576, 2003.

MINGHAO MA was born in Liaoning Province, P. R. China, received the B.S. degree in Communication Engineering from University of Science and Technology Liaoning, Anshan, P. R. China, in 2022.

He is currently pursuing the M.S. degree in Electronic Information with University of Science and Technology Liaoning, Anshan, P. R. China. His research interest is machine vision

SHAOCHUAN XU was born in Liaoning Province, P. R. China, received the B.S. degree in automation from University of Science and Technology Liaoning, Anshan, P. R. China, received the M.S. degree in control science and engineering from University of Science and Technology Liaoning, Anshan, P. R. China, in1995, and 2004.

He is currently a professor in the School of Control Science and Engineering, University of Science and Technology Liaoning, Anshan, P. R. China. He published more than 20 academic papers, more than 20 patents and software copyrights. His research interests include research on industrial intelligent control and machine vision.

DEHUA LIU was born in Liaoning Province, graduated from Anshan Liaoning University of Science and Technology with a bachelor's degree in automation in 2010.

He is an intermediate engineer in Anshan Iron & Steel Group Mining Design and Research Institute Company Limited. His main research direction is automatic control.