Original Research Article

Reflective Safety Clothes Wearing Detection in Hydraulic Engineering Using YOLOv3-CCD

ABSTRACT

The construction site of the hydraulic engineering has a high danger factor and the correct wearing of reflective safety clothes ensures the safety of workers. Therefore, the detection and testing of reflective safety clothes wearing is an important task in the construction site of a hydraulic engineering. However, the traditional manual supervision strategy has the problems of low efficiency, narrow scope, and poor real-time performance in complex working conditions. Based on YOLOv3 that is a classical target detection model, this paper proposes a reflective safety clothes detection algorithm (YOLOv3-CCD) based on an attention mechanism and an improved loss function. By adding the CA (Coordinate Attention) mechanism module to the backbone network, the characterization ability of the target feature is enhanced, so as to solve the Long-Term dependencies problem in the detection process; The loss function is changed from IOU-Loss (Intersection Over Union Loss) to CIOU-Loss (Complete-IOU Loss), so that the network takes the aspect ratio into consideration when selecting the prediction box, which improves the accuracy of target positioning; In the post-processing of the algorithm, we improved NMS (Non-Maximum Suppression) to solve the problem of dense target detection being missed. Experimental results show that compared with the original YOLOv3 network model, the algorithm has stronger robustness and the overall detection accuracy is 1.8% higher than that of the original network. Moreover, the detection speed is 32 frames per second, which is faster than the original network.

Keywords: YOLOv3; target detection; attention mechanism; loss function

1. INTRODUCTION

Reflective safety clothing is commonly used in the construction site of the hydraulic engineering and other production workplaces, and it plays a significant part in safety warnings. However, safety accidents frequently occur due to the negligence or weak safety awareness of the workers who do not wear reflective safety clothes or do not wear them in a standard way [1]. Intelligent monitoring of complex environments such as construction sites is the guarantee of an operator's safety. Therefore, it is of great significance to apply target detection technology to detect whether workers are wearing reflective safety clothes correctly.

In traditional detection, manual annotation features such as SIFT (Scale-invariant Feature Transform), HOG (Histogram of Oriented Gradient), and SURF (Speeded-up Robust Feature) are commonly used to carry out feature matching in big data feature databases [2]. However, due to the difference in image background and the diversity of illumination conditions, manual annotation features cannot achieve good robustness. With the fast development of convolutional neural networks, the two-stage detection algorithms represented by the CNN (Convolutional Neural Network) algorithm, as well as the one-stage detection algorithms represented by the SSD (Single Shot MultiBox Detector) and the YOLO (You Only Look Once) algorithms, have been widely implemented in actual applications in the industrial fields [3-5]. With the continuous improvement of these algorithms, R-CNN (Regions with CNN features), Fast R-CNN, and Faster R-CNN algorithms were

born successively, promoting the development of target detection [6]. The YOLO algorithm was proposed after the R-CNN algorithm and Darket Net was used as the backbone network [7]. In the algorithm, the process of candidate area extraction was omitted, and the whole image was divided into multiple units. Each unit was only responsible for predicting the target of the center point in the unit, and the detection speed was raised to a new height while ensuring the detection accuracy

The detection of industrial equipment such as safety helmets and reflective safety clothing is a classic problem of target detection applied to engineering security. In order to enhance the feature information of safety helmets and reflective safety clothing, scholars at home and abroad have proposed data enhancement and feature fusion schemes, respectively, and achieved certain improvement. For example, Ahmad et al. proposed using several built-in sub-networks to achieve adaptive detection of cross-scale targets [8]. Lin et al. proposed a multi-scale fusion structure of convolutional neural networks, which fused the shallow layer with rich semantic information and the deep layer with high resolution by using the method of horizontal connection [9]. Such a structure with top-down and horizontal connections is called a "Feature Pyramid Network" (FPN). The fusion information contains both deep and shallow features, which improves the expression of features without increasing the amount of computation. The YOLOv3 network adds the FPN network to the three-scale feature map behind the Darknet53 backbone network to improve the detection of target objects.

Combining previous experience and based on the YOLOv3 algorithm, a YoLov3-CCD reflective safety clothing detection algorithm based on attention mechanism and improved loss function is proposed in this paper. The purpose is to raise the weight of reflective safety clothing's attention, reinforce its characteristic information, adapt to the complex environment of a construction site, and improve the detection accuracy. The main work of this pa-per is as follows: (1) Adding a CA mechanism to the YOLOv3 backbone network to improve the reflective safety clothing's characteristic information. (2) CIOU-Loss (Complete-IOU Loss) is used as the Loss function of the boundary box, and the influence of the length-width ratio of the prediction frame on the detection effect is taken into account, so as to solve the problem of low accuracy in small sample detection. (3) DIOU-NMS (Distance-IOU Non-Maximum Suppression) is used in the post-processing of the algorithm to solve the location problem of the prediction box. (4) Training the network on the data set, and conducting comparative tests with other networks to verify the accuracy of the model for reflective safety clothing detection.

2. PRINCIPLES OF YOLOV3 ALGORITHM

The YOLOv3 network is the most classic algorithm for target detection in the YOLO series. The network integrates multi-scale prediction (SSD), Full Convolution Net (FCN), Feature Pyramid Net (FPN), Dense Net and other excellent structures [10]. With excellent detection accuracy and amazing detection speed, it is renowned through-out the target detection field. Figure 1 shows the network structure of YOLOv3.

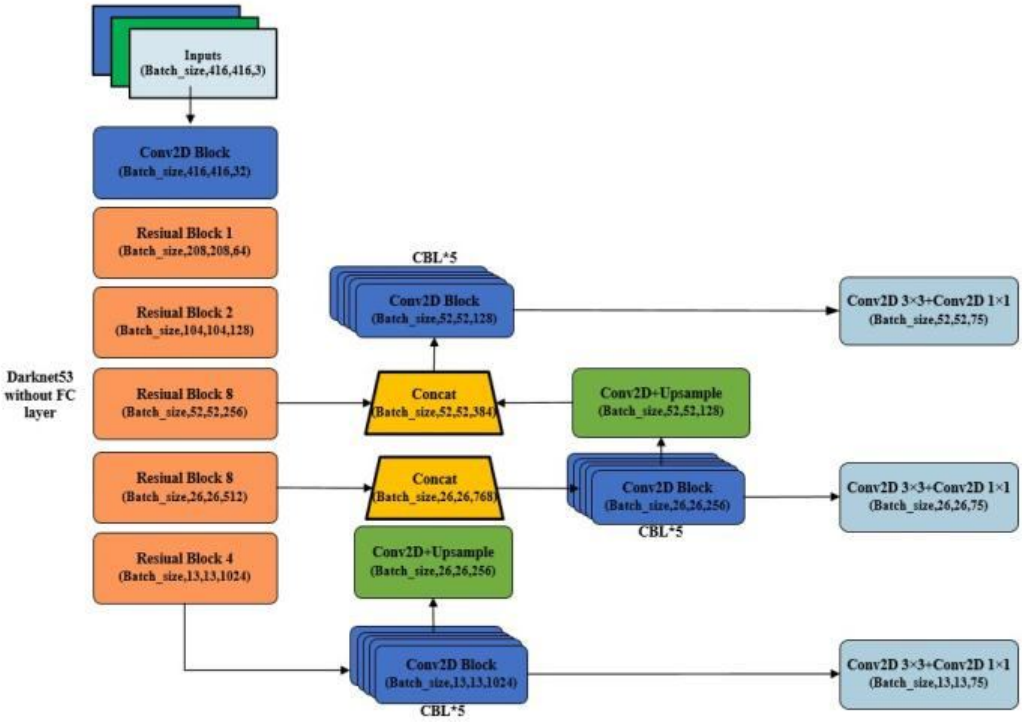


Fig. 1. YOLOv3 network structure diagram

YOLOv3 has two basic components, namely the convolution component (CBL) and the residual component (Res Unit). Figure 2 shows the schematic structure of the YOLOv3 model. The convolution component consists of the convolution layer, normalization and activation function (Leaky relu), which can reduce the size of the feature map [11]. The residual component is borrowed from the Resnet network, which can be used to build deeper net-works and prevent the loss of valid information. In addition, the convolution with a step size of 2 is used to replace the pooling process in YOLOv3, which can prevent the loss of feature information of small objects in the pooling process. Therefore, YOLOv3 provides better detection performance than other networks.

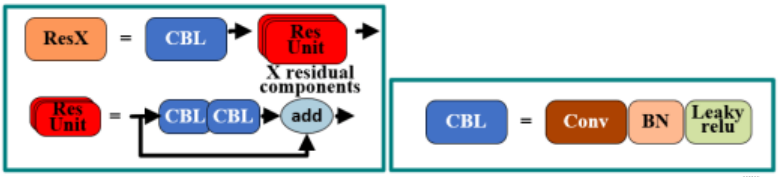


Fig. 2. Residual component and convolution component

In the backbone of the YOLOv3 Backbone network, a total of 52 convolutional operations have been carried out [12], and three scale feature layers have been respectively extracted to preserve the feature information of the deep, shallow, and middle layers. Therefore, YOLOv3 can achieve multi-scale prediction through the fusion of scales. Then, each scale corresponds to a predictive decoding process, the purpose of which is to calculate the position information (b_x、b_y、b_w、b_h) of the target box in picture. Figure 3 is the target box location information diagram. The calculation process is as formula (1)-(4):

$$b_x = \sigma(t_x) + c_x \#(1)$$
$$b_y = \sigma(t_y) + c_y \#(2)$$
$$b_w = p_w e^{t_w} \#(3)$$
$$b_h = p_h e^{t_h} \#(4)$$

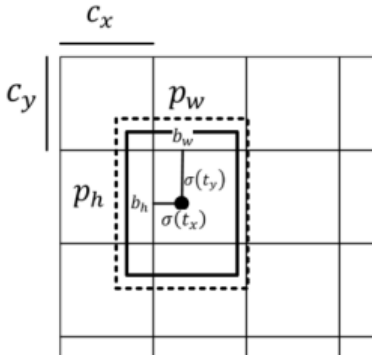


Fig. 3. Location information diagram of the target box

The prediction principle of YOLOv3 is to divide the whole map into 13*13, 26*26, 52*52 grids respectively. Each grid is responsible for the detection of one region. c_x and c_y are the number of grids from the upper left corner of the grid where the target box centroid is located to the upper leftmost corner. p_w and p_h represent the edge lengths of the prior frame. t_x and t_y are the offsets of the target box centroid relative to the upper left corner of the grid where it is located. t_w and t_h are the edge lengths of this box. σ denotes the Sigmoid activation function, which not only solves the problem of category mutual exclusion, but also increases the flexibility of the network with replacing the softmax function [13].

3. IMPROVEMENTS BASED ON YOLOV3

3.1 Improvement of the backbone network

When seeing an image, the human brain will assign the majority of its attention to prominent places in the picture and less attention to areas such as the background. Therefrom, to assign different weights to the feature graph, an attention module is added to the target detection network. The general attention mechanism is separated into two parts: channel attention and spatial attention. SENET and CBAM are two popular attentional mechanisms.

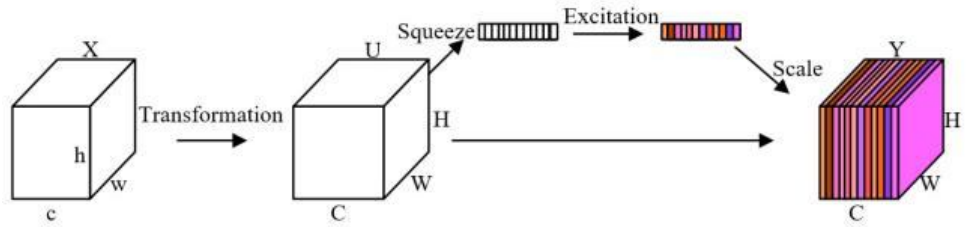


Fig. 4. SENET structure

SENET is a typical channel attention mechanism, and its structure is shown in Figure 4. The key step is to obtain a weight matrix and multiply the weight value by the spatial information (H*W) of the original channel to reconstruct the feature map [14].

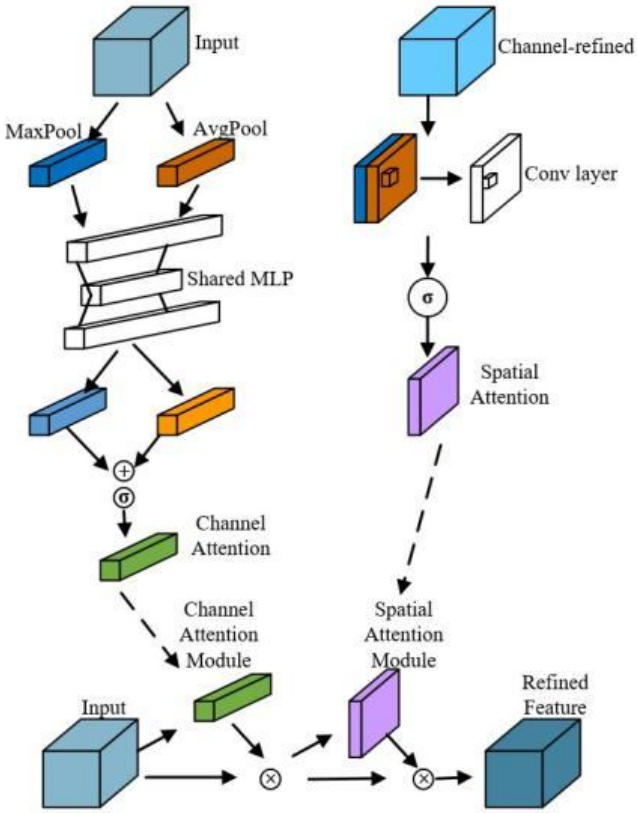


Fig. 5. Schematic diagram of the CBAM structure

CBAM is a combination of a channel attention mechanism and a spatial attention mechanism [15], and its structure is shown in Figure 5. Compared with SE, CBAM added spatial location information and achieved better detection effects, but its long calculation process caused the problem of long-range dependency which still had a negative impact on detection performance.

In order to alleviate the feature information loss and long-range dependency of SENET and CBAM in two-dimensional global pooling [16], this paper proposes a lightweight attention mechanism, Coordinate Attention (CA), which considers channel and space in parallel.

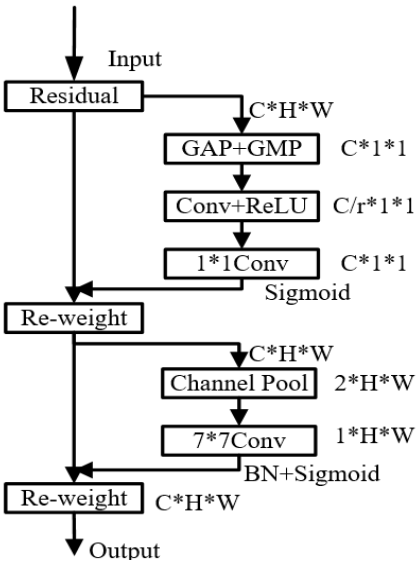
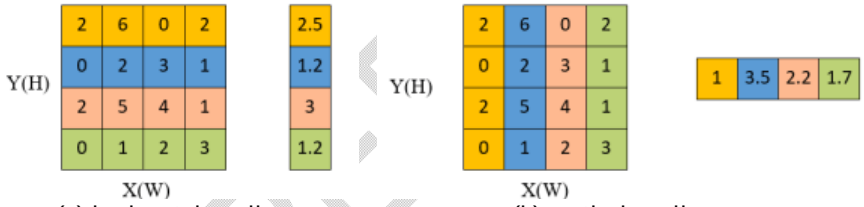


Fig. 6. Structure diagram of the CA mechanism

As shown in Figure 6, in the mixed-domain attention mechanism CA, the given input (input with dimension C*W*H) is globally pooled twice; That is, the two pooling cores with the scales (H, 1) and (1, W) are pooled along the horizontal direction (X) and vertical direction (Y) of the input feature map respectively. Finally, we can get two feature graphs (C*1*H) and (C*W*1) embedded with location information. The calculation principle is shown in Figure 7.



(a) horizontal pooling (b) vertical pooling

Fig. 7. Schematic diagram of horizontal and vertical pooling

The detailed calculation process of horizontal direction and vertical direction are shown in the formulas (5) and (6) below:

$$Z_c^h(h) = \frac{1}{W}\sum_{0\le i\le W} x_c(h,i)\ Z_c^h \in R^{C*1*H} \#(5)$$

$$Z_c^w(w) = \frac{1}{H}\sum_{0\le j\le H} x_c(j,w)\ Z_c^w \in R^{C*W*1} \#(6)$$

The two embedded feature graphs Z_c^h and Z_c^w are aggregated along the spatial dimension to obtain the global information, which overcomes the problem of long-range dependency and helps the network locate the target. The number of parameters is controlled by changing the number of characteristic channels through 1*1 convolution (F_1) and then activated by a nonlinear activation function (δ). The calculation process is shown in the formula (7):

$$f = \delta\left(F_1([Z^w, Z^h])\right) \#(7)$$

Then a split operation is performed along the spatial dimension to obtain the two separated feature graphs (f_w、f_h), and then 1*1 convolution (F_w、F_h) and a sigmoid function (σ) are performed, respectively, to obtain the two attention vectors (g^h、g^w) with the same number of channels as the input. The lower sampling ratio, r = 16, is introduced to balance the performance and computation of the network. The calculation formulas are as follows:

$$f^h \in R^{\frac{C}{r}*1*H}\ g^h = \sigma\left(F_h(f^h)\right) \#(8)$$

$$f^w \in R^{\frac{C}{r}*W*1} g^w = \sigma\left(F_w(f^w)\right)\#(9)$$

Finally, g^h and g^w are multiplied by the original input as weights to obtain the final output of the CA module. The calculation process is expressed by the formula (10):

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j)\#(10)$$

The CA module is a novel attentional mechanism that not only alleviates the problem of long-range dependency but also improves the accuracy of the network without increasing the amount of computation. The module has been loaded into the MobileNet network and achieved good detection results. Based on this experience, in this paper we apply the CA module to the backbone network of the YOLOv3 model to adapt to the complex environment such as uneven site illumination, and thus improve the detection accuracy of reflective safety clothing. The detection effect is shown in Figure 8. The reflective safety clothing that was missed in the original YOLOv3 model but detected in the YOLOv3-CCD model. It can be seen that the application of the CA module can improve the ac-curacy and performance of the reflective safety clothing detection network.

(**a**) (**b**) (**c**)

Fig. 8. Comparison of detection effects: (a) The original image; (b) Detection results based on YOLOv3; (c) Detection results of Yolov3 with CA mechanism

The heat map uses colors to distinguish the weight values of different positions of the image. The darker the color, the higher the weight value of its corresponding position. We compare the heat map before and after the improved algorithm and find that the network with the added attention mechanism shows a better heat map. Figure 9 illustrates that the algorithm assigns more weights to the reflective safety clothes in the images after adding the CA mechanism.

(a) (b) (c)

Fig. 9. Comparison of heat map effects: (a) The original image; (b) Heat map based on YOLOv3; (c) Heat map of Yolov3 with CA mechanism

3.2 Improvement of loss function

The detection errors of YOLOv3 include the center coordinate error, width-height coordinate error, confidence error ($Loss_{conf}$) and classification error ($Loss_{class}$). All three items are calculated using the binary cross-entropy loss function, except for the width-height coordinate error, which is calculated using the mean variance loss function. The loss of center and width-height coordinates is called the positioning Loss ($Loss_{loc}$), so the original YOLOv3 loss function can be expressed by the formula (11) as follows:

$$Loss_{object} = Loss_{loc} + Loss_{conf} + Loss_{class} \#(11)$$

In this paper, we mainly improve the calculation of $Loss_{loc}$. Previously, the calculation methods of IOU-Loss (Intersection Over Union), GIOU-Loss (Generalized IOU) and DIOU-Loss (Distance IOU) [17] have been proposed. GIOU solves the problem when the bounding boxes do not overlap on the basis of IOU. DIOU considers the information of the center distance of the bounding boxes based on IOU. But they are not available for all cases.

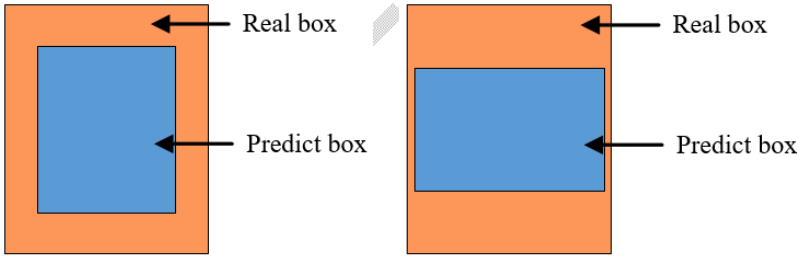


Fig. 10. Comparison of the positions of the prediction frame and the real frame

As shown in Figure 10, the IOU, GIOU and DIOU values of the prediction box and the real box are the same, but obviously the prediction box in the left picture (a) has higher accuracy. In order to obtain the optimal prediction box [18], this paper introduces a new calculation method to solve this problem—CIOU-Loss, as shown in formulas (12)-(14):

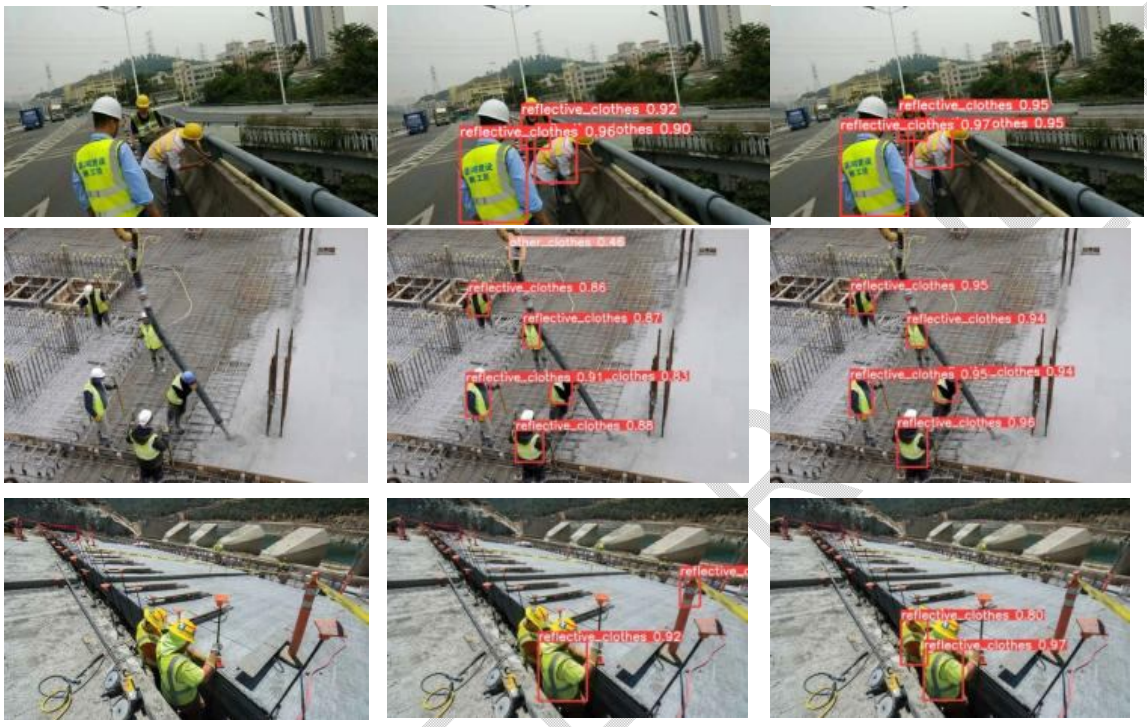
$$L_{CIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \#(12)$$

$$\alpha = \frac{v}{(1 - IOU) + v} \#(13)$$

$$v = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2 \#(14)$$

v is a parameter that is used to measure the similarity of the ratio of the width and height of the two frames. The closer the ratio is, the smaller the v value is and the lesser the loss is. w^{gt} and h^{gt} represent the width and height of the predicted

frame, respectively. w and h represent the width and height of the real frame. α is the weight parameter, when the v value is constant, the larger the IOU the larger the α is and the larger the loss is. It means that the algorithm pays more attention to the aspect ratio between the two frames when the IOU is large. CIOU integrates the center distance, overlap ratio and aspect ratio between the two frames. We apply this method to the YOLOv3 network and find that the detection accuracy is improved to some extent. As shown in Figure 11 below, the confidence of the detection results with the improved loss function is significantly increased.



(**a**) (**b**) (**c**)

Fig. 11. Comparison of optimization effects of loss functions: (a) The original image; (b) Detection results based on YOLOv3; (c) Detection results of Yolov3 with CIOU-Loss

3.3 Improvement of post-processing of target detection

The final output of the neural network in the YOLO algorithm is several candidate frames with confidence. Usually, a fixed threshold is used to remove some frames with low confidence, and then an NMS (Non-Maximum Suppression) algorithm is used to remove detection frames with high overlap.

The operation principle of NMS is to select the box M with the highest score from the generated box set B and put it into the final result set D [19]. Then, the detection boxes that overlap with frame M and whose IOU is greater than the fixed threshold Nt are screened and deleted from the set B. This method not only effectively removes a large number of redundant detection frames, but also reduces the computational burden of iterative operation as well as improves the detection efficiency. However, because the method of fixed threshold is used to forcibly delete adjacent detection boxes, the algorithm may mistakenly delete dense objects, thus reducing the detection performance of the algorithm. If the fixed threshold value is simply raised, some redundant boxes will not be deleted, and the detection accuracy of the algorithm will decrease [20].

The post-processing improvement method of the algorithm proposed in this paper for reflective safety clothing detection is to change the NMS with a fixed threshold value to the DIOU-NMS that considers the center distance between the real box and the detection box [21]. Using DIOU as the standard for NMS can effectively solve the problem of missed detection caused by the forced deletion of adjacent detection frames. The calculation process of this method is expressed by the formula (15):

$$S_i = \begin{cases} S_i, & IOU - R_{DIOU}(M, B_i) < \varepsilon \\ 0, & IOU - R_{DIOU}(M, B_i) \geq \varepsilon, \end{cases} R_{DIOU} = \frac{\rho^2(b, b^{gt})}{c^2} \#(15)$$

Where M represents the high confidence detection frame, B_i represents the generated detection frame set, ε is the NMS threshold, and S_i is the classification confidence. Figure 12 shows that the improved non-maximum suppression improves the detection accuracy of reflective safety clothing in dense scenes.

(**a**) (**b**) (**c**)

Fig. 12. Comparison of optimization effects of loss functions: (a) The original image; (b) Detection results based on YOLOv3; (c) Detection results of Yolov3 with DIOU-NMS

4. EXPERIMENT AND RESULT ANALYSIS

4.1 Experimental environment and data set

Through self-collection, real shooting, reorganization of open source datasets and web crawlers, this paper obtained a total of 7083 still images as the dataset for model training. It contains nearly 10,000 manual labeling boxes of clothes, including reflective clothes ('reflective _clothes') and other clothes ('other _clothes'). The XML file of the training set was converted into a labeled file in TXT format in code form, and the ratio of the training set to the test set was 4:1. The experimental environment is shown in Table 1.

Table 1. Experimental environment

Hardware and software platforms	
The operating system	Win10
CPU	Intel 10400f CPU
GPU	GeForce RTX3080
Memory	16GB
Development platform	PyTorch1.7.0
Development of language	Python3.8.5

4.2 Metrics for model evaluation

In order to quantitatively compare the detection performance of the model, this paper assumes that TP is the number of detection pairs and FP is the number of error detections among detected objects. Among undetected objects, FN represents the number of positive samples, whereas TN is the number of negative samples. The four values of TP\FP\TN\FN can be further interpreted using a binary confusion matrix.

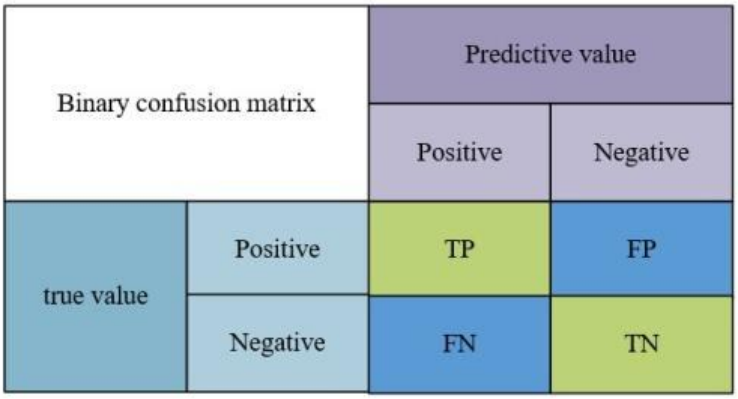


Fig. 13. Binary confusion matrix

The following indicators are used to evaluate the model:

Recall (R): the ratio of the number of detection pairs to the total number of positive samples. The formula is as follows:

$$R = TP/(TP + FN)\#(16)$$

Precision (P) is the ratio of the number of detected pairs to the total number of detected pairs. The following is the formula:

$$P = TP/(TP + FP)\#(17)$$

The weighted average harmonic of Precision and Recall is F_1, and its formula is as follows:

$$F_1 = \frac{2 \times P \times R}{P + R}\#(18)$$

The most widely used evaluation metric in object detection is mAP@0.5, which is determined by averaging the mean accuracy across all categories at IOU = 0.5.

$$mAP = \sum_{i=1}^{N} \int_0^1 P_i(r)dr/N\ \#(19)$$

The parameters in this experiment were set as follows: the image size was 640*640 as input, the epoch of the training was set as 200 times, and the batch size was set as 8.

4.3 Experimental results and analysis

In this paper, the improvements are compared and experimented under different combinations. The evaluation indexes are F1 and mAP@0.5, where mAP@0.5 (%) is determined by averaging the mean accuracy across all categories at IOU = 0.5. ΔmAP@0.5 % is the difference between a mAP value and the number above it in the table.

The CA mechanism module has the advantage of being plug and play. In the experiment, after the module is added to the backbone of the YOLOv3 network, the improved YOLOv3 network structure is shown as Figure 14.

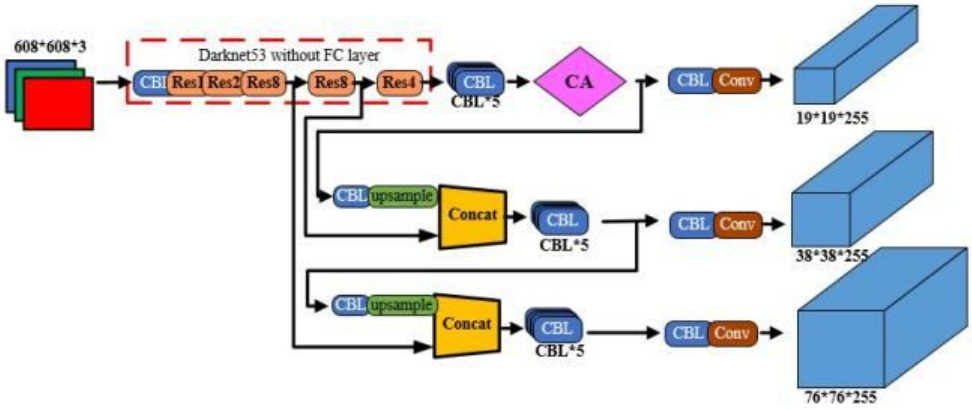


Fig. 14. Improved YOLOv3 network

As shown in Table 2 below, the addition of the CA attention mechanism module improves the detection accuracy of the original YOLOv3 network and improves the detection performance of the target. More specifically, the accuracy of detection results is improved by 0.8% compared with the original model when only the CA mechanism is added, and the accuracy of detection with improved loss function and NMS is improved by 1.3% compared with the original network. When improving the backbone network, loss function, and post-processing together, the accuracy value obtained is 1.8% better than the original network. The overall improved detection frame rate of YOLOv3 is 32fps, which is 8 higher than the 24fps of the original network.

Table 2. Comparison of ablation results

CA	CIOU-Loss+DIOU-NMS	F_1(%)	mAP@0.5 (%)	ΔmAP@0.5 (%)	Recognition speed /(frames/s)
		93	87.3		24
√		94	88.1	0.8	26
	√	95	88.6	0.5	31
√	√	96	89.1	0.5	32

Through the above analysis, we can know that the CA mechanism can enhance the attention weight of the target features, improve the ability of the network to extract the target features, and improve the detection performance of the network. Figure 15 shows the comparisons of different improved detection results and further il-lustrates the ablation experiment.

(a) (b) (c)

(d)

(e)

Fig. 15. Comparison of detection effects: (a) The original image; (b) Detection results based on YOLOv3; (c) Detection results of Yolov3 with CA mechanism; (d) Detection results of Yolov3 with DIOU-NMS; (e) Detection results of Yolov3 with CA mechanism and DIOU-NMS

The detection results are also compared with those of different algorithm models, as shown in the Table 3. The mAP value of the improved network is 1.8% higher than that of the original YOLOv3 model and 0.5% higher than that of the YOLOv4 model. Moreover, the detection frame rate of the YOLOv3-CCD is 32 fps, which is 7 higher than the 25 fps of YOLOv4 and 8 higher than the 24 of the original YOLOv3. Compared with YOLOv5, the mAP value of YOLOV3-CCD is lower, but the detection frame rate is higher. Therefore, YOLOv3-CCD has the best detection performance, and the application value of YOLOv3-CCD is higher.

Table 3. Performance comparison of different algorithms

Model	F_1(%)	mAP@0.5 (%)	ΔmAP@0.5 (%)	Recognition speed /(frames/s)
YOLOv3	95	87.3		24
YOLOv4	95	88.6	1.3	25
YOLOv3-CCD(ours)	96	89.1	0.5	32
YOLOv5	96	91	0.9	27

The plots of mAP for different models were obtained by multiple sets of experiments, with the number of training as the horizontal coordinate and the average accuracy as the vertical coordinate, as shown in Figure 16. The result shows that the YOLOv3-CCD model not only has a high detection accuracy but also has a fast convergence speed.

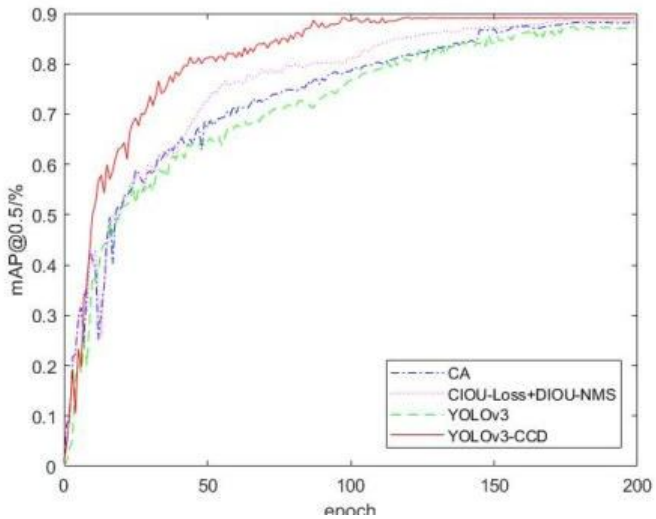


Fig. 16. The plots of mAP for various models

5. CONCLUSION

In this paper, the YOLOv3-CCD network model is proposed to solve the problem of missing and mischecking of reflective safety clothes in complex construction environment of hydraulic engineering. The optimization and improvement of the YOLOv3 network structure are realized to improve the detection performance and accuracy regarding whether an operator is wearing reflective safety clothing. Firstly, the CA mechanism module is added into the backbone network of YOLOv3, which improves the attention weight of target feature and is beneficial to feature extraction. Secondly, based on the YOLOv3 network, loss function and non-maximum suppression post-processing are improved. The CIOU-Loss function and DIOU-NMS are used to improve coordinate loss and prediction box position, respectively. Finally, compared with other target detection networks, the results show that the improved network model not only improves the detection accuracy by 1.8% compared with the original net-work while ensuring the detection efficiency, but also has a significant improvement in detection speed. The improvement of this model is not only of great significance to the detection of reflective safety clothes in the construction environment, but can also be extended to the detection of other important objects in hydraulic engineering.

REFERENCES

1. Black, A.; Bui, V.; Henry, E.; Ho, K.; Pham, D.; Tran, T.; Wood, J. Using retro-reflective cloth to enhance drivers' judgment of pedestrian walking direction at night-time. Journal of Safety Research 2021, 77, 196-201.

2. Xiao, Y.; Tian, Z.; Yu, J.; Zhang, Y.; Lan, X. A review of object detection based on deep learning. Multimedia Tools and Applications 2020, 79, 23729-23791.

3. Biswas, D.; Su, H.; Wang, C.; Stevanovic, A.; Wang, W. An automatic traffic density estimation using Single Shot Detection (SSD) and MobileNet-SSD. Physics and Chemistry of The Earth 2019, 110, 176-184.

4. Xu, D.; Wu, Y. Improved YOLO-V3 with DenseNet for Multi-Scale Remote Sensing Target Detection. SENSORS 2020, 20.

5. Guo, Z.; Wang, C.; Yang, G.; Huang, Z.; Li, G. MSFT-YOLO: Improved YOLOv5 Based on Transformer for Detecting Defects of Steel Surface. SENSORS 2022, 22.

6. Wei, B.; Hao, K.; Gao, L.; Tang, X. Detecting textile micro-defects: A novel and efficient method based on visual gain mechanism. Information Sciences 2020, 541, 60-74.

7. Liu, C.; Wu, Y.; Liu, J.; Han. J. MTI-YOLO: A Light-Weight and Real-Time Deep Neural Network for Insulator Detection in Complex Aerial Images. Energies 2021, 14.

8. Ahmad, T.; Ma, Y.; Yahya, M.; Ahmad, B.; Nazir, S.; ul Haq, A. Object Detection through Modified YOLO Neural Network. Scientific Programming 2020, 2020.

9. Yang, W.; Zhang, J.; Chen, Z.; Xu, Z. An efficient semantic segmentation method based on transfer learning from object detection. IET Image Processing 2021, 15, 57-64.

10. Huang, J.; Zhang, H.; Wang, L.; Zhang, Z.; Zhao, C. Improved YOLOv3 Model for miniature camera detection. Optics and Laser Technology 2021, 142.

11. Hong, Z.; Yang, T.; Tong, X.; Zhang, Y.; Jiang, S.; Zhou, R.; Han, Y.; Wang, J.; Yang, S.; Liu, S. Multi-Scale Ship Detection from SAR and Optical Imagery via A More Accurate YOLOv3. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2021, 14, 6083-6101.

12. Duan, L.; Yang, K.; Ruan, L. Research on Automatic Recognition of Casting Defects Based on Deep Learning. IEEE Access 2020, 9, 12209-12216.

13. Deng, Z.; Yang, R.; Lan, R.; Liu, Z.; Luo, X. SE-IYOLOV3: An Accurate Small Scale Face Detector for Outdoor Security. MATHEMATICS 2020, 8.

14. Xue, M.; Chen, M.; Peng, D.; Guo, Y.; Chen, H. One Spatio-Temporal Sharpening Attention Mechanism for Light-Weight YOLO Models Based on Sharpening Spatial Attention. Sensors 2021, 21.

15. Guo, N.; Gu, K.; Qiao, J.; Bi, J. Improved deep CNNs based on Nonlinear Hybrid Attention Module for image classification. Neural Networks 2021, 140, 158-166.

16. Xie, H.; Wu, Z. A Robust Fabric Defect Detection Method Based on Improved RefineDet. Sensors 2020, 20.

17. Zhou, Q.; Qin, J.; Xiang, X.; Tan, Y.; Xiong, N. Algorithm of Helmet Wearing Detection Based on AT-YOLO Deep Mode. Computers, Materials and Continua 2021, 69, 159-174.

18. Guo, C.; Cai, M.; Ying, N.; Chen, H.; Zhang, J.; Zhou, D. ANMS: attention-based non-maximum suppression. Multimedia Tools and Applications 2022, 81, 11205-11219.

19. Jiang, S.; Xu, T.; Li, J.; Huang, B.; Guo, J.; Bian, Z. IdentifyNet for Non-maximum Suppression. IEEE Access 2019, 7, 148245-148253.

20. Lawal, O. YOLOMuskmelon: Quest for Fruit Detection Speed and Accuracy Using Deep Learning. IEEE Access 2021, 9, 15221-15227.

21. Yap, M.; Hachiuma, R.; Alavi, A.; Brungel, R.; Cassidy, B.; Goyal, M.; Zhu, H.; Ruckert, J.; Olshansky, M.; Huang, X.; Saito, H.; Hassanpour, S.; Friedrich, C.; Ascher, D.; Song, A.; Kajita, H.; Gillespie, D.; Reeves, N.; Pappachan, J.; O'Shea, C.; Frank, E. Deep Learning in Diabetic Foot Ulcers Detection: A Comprehensive Evaluation. Computers in Biology and Medicine 2021, 135.