Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 11, No. 2, 2025

Optimization Research Method for Personal Protective Equipment Wearing Detection on Offshore Platforms Based on YOLOv7

Feng Sun *

College of Mechanical Engineering, Tianjin University of Technology and Education, Tianjin 300222, China
* **Corresponding author Email**: 2723521868@qq.com

Abstract: In response to the monitoring requirements of offshore operations, this paper addresses the issue of subjective oversight in the wearing of personal protective equipment (PPE) by inspection personnel on offshore platforms during their work process. By integrating object detection algorithms from the field of computer vision, a monitoring method for PPE usage by workers on offshore platforms is proposed. This method builds upon the YOLOv7 neural network and introduces the CBAM attention mechanism and the Focaler-IOU bounding box loss function to optimize the recognition accuracy of detection targets, resulting in an increase in the mean average precision (mAP) by 5.9% compared to the original model. The feasibility and effectiveness of the integrated detection method were validated through testing on a field dataset.

Keywords: Process Monitoring; Yolov7; Target Recognition; CBAM Mechanism; Labor Insurance Supplies Testing.

1. Introduction

The operating environment of offshore platforms is known for its complexity and variability, which poses a high safety risk to workers. With increasing concern about safety issues on industrial sites, it is critical to ensure that workers are properly wearing protective equipment such as hard hats, goggles, ear muffs, protective clothing, gloves and protective shoes in such environments. These devices are the first line of defense to reduce potential unsafe factors and health risks in the work environment. Therefore, how to effectively supervise the wearing of labor insurance supplies has become an important issue to be solved.

The traditional inspection method of labor insurance supplies mainly relies on manual work, which is not only inefficient, but also prone to negligence and omission. However, with the rapid development of computer vision technology and the field of deep learning, especially the progress of object detection technology, new solutions have been brought to this challenge. Deep learning-based methods have demonstrated superior ability to identify and classify objects in images, making them ideal for improving operational safety on offshore platforms.

In this paper, a technical scheme of wearing inspection of labor insurance products using YOLOv7 is proposed, aiming to improve the safety level of offshore platforms through intelligent video surveillance system. The system will be integrated into inspection monitoring robots, which are responsible for performing safety inspection tasks at various key locations. They capture pictures, video or live footage of the scene and transmit this data to the detection server. The server uses deep learning algorithms to analyze the collected data and automatically determine whether workers are complying with the safety dress code, such as whether they are wearing the necessary labor protection products. Once any unsafe behavior is detected, the system will immediately issue an alert so that corrective action can be taken in time. This intelligent monitoring method can not only greatly improve work efficiency, but also effectively reduce the risk of human error, thus providing a safer working environment for offshore platforms.

The Convolutional Neural Network (CNN) object recognition method based on deep learning can automatically extract features through neural networks, and has the characteristics of strong robustness, high precision and high efficiency. Target detection algorithms mainly include R-CNN[1], SPPNet [2], Faster R-CNN[3], YOLO[4], etc. Based on YOLOv4, Li H added an enhanced PAN (E-PAN) structure [5] to improve the accuracy of the helmet. Following YOLOv4, Chien-YaoWang et al proposed YOLOv7 object detection model [6] in 2022. In the architecture, extended efficient aggregation network (E-ELAN) was adopted to control the shortest and longest gradient path, so that the model could learn and converge more effectively. By improving the attention mechanism of YOLOv7 [7], Jakubec M[7] reduces the computational requirements of the model and improves the detection accuracy of the helmet. Combined with the above analysis, relevant scholars at home and abroad have introduced deep learning methods into helmet wearing detection and achieved good detection results.

This paper will test and analyze the wearing of multi-objective labor protection products (safety helmet, labor protection clothing, goggles, ear muffs, gloves and labor protection shoes, etc.) for offshore platform workers, and optimize and improve the model network framework and function module on the basis of the original model.

2. Target Detection Analysis based on YOLOv7 Model

2.1. YOLOv7 Model Introduction

YOLOv7 model is one of the models with the best comprehensive performance among current target detection models. Its network structure is mainly composed of Input, Backbone and Head. Input input data images with a resolution of 640×640 are input. Backbone is used to double channels and extract features, while Head is used to predict results. Its structure is shown in Figure 1.

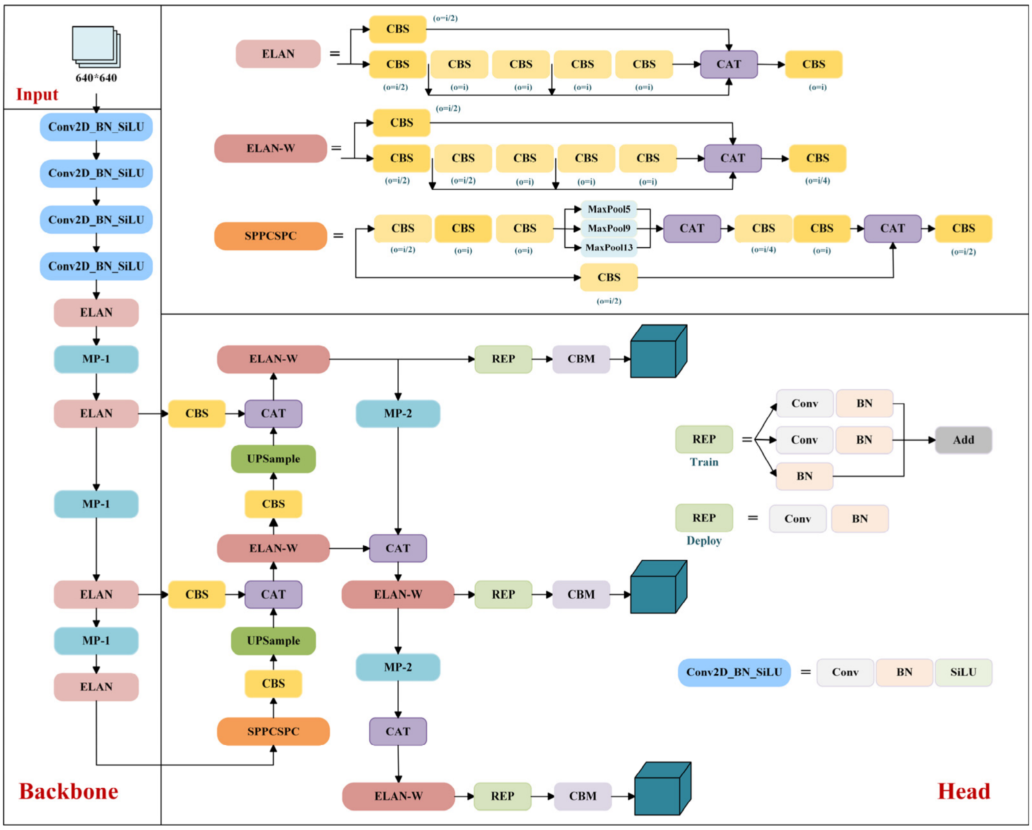


Figure 1. Structure diagram of Yolov7

As shown in Figure 1, ELAN (Efficient Layer Aggregation Networks) module and MP (Max Pooling) structure are used in the Backbone part. ELAN module is an efficient network structure with two branches, as shown in Figure 2. The first branch is changed through a 1×1 convolutional channel, and the second branch is extracted by four 3×3 convolutional modules on the basis of the first branch, and finally the four features are superimposed to obtain the final feature extraction result. ELAN can effectively aggregate the feature information of different layers by controlling the shortest and longest gradient paths, which improves the accuracy and robustness of the target detection algorithm. The function of MP structure is to use downsampling to reduce the feature loss of the model.

In the Head part, the SPPCSPC module deals with four branches of MaxPool(5,9,13,1). These four different branches represent that the SPPCSPC module is capable of processing different objects. The maximum pooling of the module has four kinds of receptive fields, and the maximum pooling of the SPPCSPC module is used to obtain receptive fields. The algorithm ADAPTS to different resolution images and can be used to distinguish large targets from small ones. The UPSample module is an upsample module that uses the method of nearest neighbor interpolation. The REP module is divided into training module and reasoning module. The training module has three branches. The first branch is a 3×3 convolution for feature extraction. The second branch is a 1×1 convolution for smoothing features. The third branch is a direct move that finally adds them together. The inference module contains a 3×3 convolution of weight addition from the reparameterized transformation of the training module.

2.2. Figures Optimization and Improvement based on YOLOv7 Model

2.2.1. Convolutional Module Optimization

PConv convolution module is a part of encoding module and decoding module in U-Net structure. It is a convolution module based on void convolution, which is used to process images with occluded objects and helps to improve the model's performance of small target detection. The PConv convolution module consists of two convolution layers, a mask layer and a normalization layer, wherein the mask layer is used to generate the mask and the normalization layer is used to normalize the mask and the feature map. In the coding module, the PConv convolution module is used to extract image features, while in the decoding module, the PConv convolution module is used to restore the feature map to the original image.

2.2.2. Insert CBAM Attention Mechanism

The attention mechanism is a bionic mechanism that mimics human vision and is applied to machines. In neural networks, the attention module is usually an additional neural network, and the attention mechanism helps the model to focus on the information target data and ignore irrelevant target information when processing sequence data. The attention mechanism can be applied to any type of input, and it is a very effective resource allocation scheme to solve information overload in the case of limited computing power.

As shown in Figure 2, CBAM is composed of Channel Attention Module and Spatial Attention Module. Channel attention helps to enhance the feature representation of different channels. The channel attention module uses global maximum pooling and global average pooling to obtain the information of each channel in the feature graph, and uses two convolution layers to fuse the pooled feature graph, and then uses sigmoid activation function to scale the resulting feature graph to the range of [0,1], and serves as the weight coefficient of the channel attention. The spatial attention module takes the feature map output by the channel attention module as the input feature map of the module, and then performs global maximum pooling and global mean pooling based on the channel, and fuses the two pooling results, and finally generates the feature map through sigmoid.

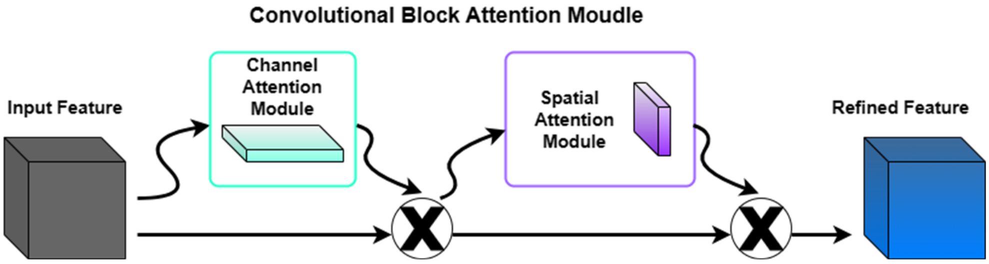


Figure 2. CBAM attention module

2.2.3. Improved Loss Function

Loss function is a non-negative real function used to quantify the difference between model prediction and real label. The calculation of loss function in YOLOv7 object detection algorithm is related to category loss, confidence loss and coordinate regression loss.

At the moment, Loss functions commonly used to improve YOLOv7 include CIOU (Complete Intersection over Union), GIOU (Generalized Intersection over Union), and WIOU (Wise) Intersection over Union) and MPDIOU (Minimum Points Intersection over Union). Different loss functions have different effects on the detection speed, detection accuracy and system memory. Therefore, for the valve switch data set, the method of experimental comparison is used to select the optimal loss function.

3. Experimental Analysis

3.1. Model Training

All experiments in this paper were run under the same server conditions: the server operating system was Ubuntu 20.04 and Intel® Core™ i7-13700K processor was used. The server graphics card is configured as NVIDIA GeForce RTX 3090 24G. The deep learning environment is configured as Pytorch learning framework, CUDA version is 11.6 and uses GPU-accelerated mode for training. Transfer learning is adopted in the training process, and the model is trained in this paper

3.2. Based on YOLOv7 Labor Insurance Products Wear Detection Analysis

The data set used in the labor insurance products testing experiment in this paper comes from the on-site shooting of the offshore platform, and the data set after data enhancement is about 5,500 pieces. The data set mainly includes multiple types of pictures under complex scenes such as single-person multi-target and multi-person multi-target (as shown in Figure 3). Among them, 5000 training sets and 500 verification sets were divided.

(a) (b) (c) (d)

Figure 3. Training data set

3.3. Marine Field Target Recognition based on YOLOv7

3.3.1. Model Optimization Ablation Experiment

The results of target detection ablation experiment are shown in the table below.

Table 1. Target detection ablation results

Model	CBAM	SE	GIOU	Focaler-iou	mAP@0.5
Yolov7-1					77.2%
Yolov7-2	√				**81.3%**
Yolov7-3		√			78.3%
Yolov7-4			√		80.2%
Yolov7-5				√	**82.6%**
Improved model	√			√	**83.1%**

In order to verify the improvement effect of CBAM attention mechanism on the model, SE attention mechanism was added as the control group for comparison. According to the data in Table 1, the improvement effect of SE attention mechanism was smaller than that of the original model, while the average accuracy of CBAM attention mechanism was increased by 4.1%. The adaptive feature extraction capability of the original model is enhanced.

Compare the GIOU, CIOU and Focaler-iou loss functions. The model using Focaler-iou loss function has the best improvement effect, and the average accuracy reaches 82.6%, which is 5.4% higher than that of the original model. Focaler-iou is a bounding box regression loss function that focuses on different regression samples and improves the detection performance of heavy detection tasks by focusing on different regression samples. The CBAM attention mechanism and Focaler-iou loss function are combined to improve the comprehensive detection performance of the optimized model. The average accuracy reaches 83.1%, which is 5.9% higher than that of the original model, which verifies the superiority of the improved model.

3.3.2. Algorithm Comparative Analysis

There are 9 types of experimental targets in this paper. Since the detection accuracy of labor protection clothing, safety helmet, labor protection shoes and workers are high and the accuracy gap is small, only the 3 types of targets with obvious improvement effect in Table 2 are compared and analyzed here. The following models all use the same data set for training reasoning, and the model parameters are set the same. The model performance pairs are shown in Table 2.

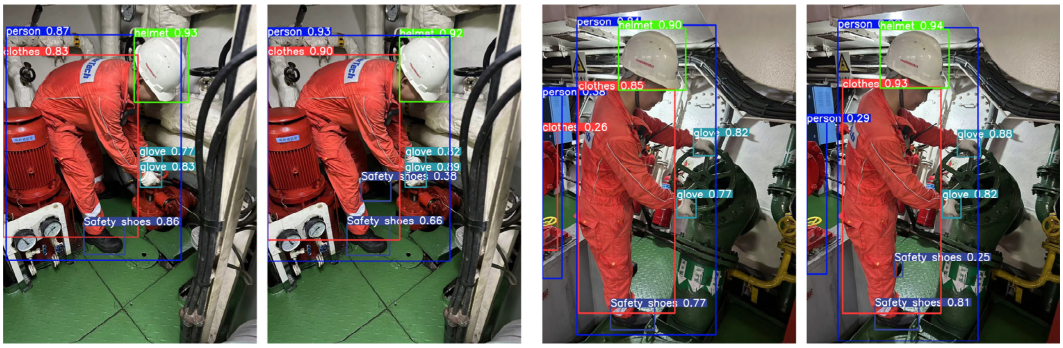
Table 2. Model performance comparison

Model	Earmuffs	Earplug	Goggles	mAP@0.5
SSD	85.8%	31.3%	70.0%	62.37%
Yolov5	82.6%	37.4%	64.6%	61.53%
Yolov7	86.7%	43.3%	73.1%	67.70%
Improved model	90.8%	56.1%	79.9%	**75.60%**

Based on the YOLOv7 model, the ablation experiment and the comparison experiment were used to analyze the data of 3 types of targets with obvious improvement among the 9 types of targets. The mAP@0.5 value is 75.60%, which is 4.9% higher than that of the original model.

3.3.3. Analysis of Model Testing Results

Compare the detection results of the improved model with the original model, and the detection results are as follows:



(a) (b) (c) (d)

Figure 4. Comparison of Yolov7 detection results

In Figure 4, Figure (a) and Figure (c) are the detection results of the original model, and Figure (b) and Figure (d) are the detection results of the optimized model. Through the direct comparison of the detection effects of the two models, the detection effect of the optimized model is better and the detection accuracy is higher, which verifies the effectiveness of the improved model.

4. Conclusion

Based on Yolov7 model, this paper optimizes the original model by optimizing the Pconv convolutional module, adding CBAM attention mechanism and optimizing loss function. The optimized model has significantly improved the detection accuracy, and its mAP@0.5 value reaches 83.1%, which is 5.9% higher than that of the original model. The effectiveness of the improved model is verified by model comparison experiments.

References

[1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[2] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.

[3] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.

[4] Jiang P, Ergu D, Liu F, et al. A Review of Yolo algorithm developments[J]. Procedia Computer Science, 2022, 199: 1066-1073.

[5] Li H, Wu D, Zhang W, et al. YOLO-PL: Helmet wearing detection algorithm based on improved YOLOv4[J]. Digital Signal Processing, 2023: 104283.

[6] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 7464-7475.

[7] Jakubec M, Lieskovska E, Brezani A, et al. Deep learning-based automatic helmet recognition for two-wheeled road safety [J]. Transportation research procedia, 2023, 74: 1171-1178.