An Improved YOLOv8 Algorithm Model for Detection of Personal Protective Products in Chemical Plants

Hekang Cheng[1,a], Suqun Cao[1,*], Daheng Wang[1,b], Xinze Shen[1,c], Kun Zhao[2,d], Jianhui Wu[1,e], Tao Jiang[1,f], Haitao Zhou[1,g], Di Zhang[1,h], Jianxue Zhao[1,i]

[1.]Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian, JS 223003, China
[2.]School of Art and Design, Guilin University of Electronic Science and Technology, Guilin, JS 541004, China

[a]chenghekang01@163.com), *caosuqun@126.com, [b]daheng_wang@163.com, [c]shenxinze817@gmail.com, [d]15717325283@163.com, [e]wjh_516328659@163.com, [f]yjgm0412@163.com, [g]zhouhaitao881@gmail.com, [h]zhangdi202209@163.com, [i]15725310789@139.com

Abstract. With the rapid development of society and the acceleration of industrialization, the chemical industry, with its irreplaceable position, has become an important pillar of the national economy. In the key field of chemical plants, the issue of production safety is particularly prominent. Operators are in such a special environment, whether the personal protection measures are in place is directly related to their personal safety, this paper is based on the YOLOv8 algorithmic model, respectively, in the YOLOv8 backbone network and the Neck part of the optimization, which is mainly to optimize the Bottleneck module in the backbone network to the MishNextBlock module, and in the YOLOv8 Neck's pyramid structure, two modules are added: Spatial and Channel Attention (SCA) and Global Attention Upsample (GAU), which are fused into a new Path Aggregation Network (PAN) structure. PAN structure, thus forming the improved YOLOv8 algorithm model. And it is shown experimentally that the improved YOLOv8 algorithm realizes a significant improvement in the detection accuracy of small-scale targets, near and far distance multi-scale and fuzzy small targets in chemical plant protective equipment.

Keywords: YOLOv8; GUA; SCA; Chemical plant protective products

1 INTRODUCTION

With the rapid development of science and technology, the chemical industry is facing more and more challenges in the production process. Chemical plant protection products are directly related to production safety and the economic benefits of enterprises. Traditional manual monitoring methods face problems such as inefficiency and inabil

A. Haldorai et al. (eds.), *Proceedings of the 2024 3rd International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID 2024)*, Atlantis Highlights in Intelligent Systems 11,
https://doi.org/10.2991/978-94-6463-490-7_53

equipment (PPE), referred to as chemical protective equipment, the correct use of which is essential to protect workers from harmful chemicals and hazardous environments. Therefore, it is of great significance to explore an efficient and accurate detection method for chemical protective products to improve the safety level of chemical industry.

Currently, the main detection techniques are categorized into sensor-based and vision-based approaches. Sensor-based approaches such as Bukhari et al. [1]. integrate optical imaging, passive infrared, and fingerprint sensors to achieve multilevel hazard assessment and multilayer alarm routing to achieve security item alerts. In addition, researchers such as Uddin and Nyeem [2] used ultrasonic array sensors hidden inside walls to achieve security item detection.

Vision-based approaches, on the other hand, utilize surveillance cameras and image processing techniques to identify whether an individual is using safety guards correctly or not. Lee et al. [3] and other researchers have effectively improved safety protection in the workplace by using the YOLO-EfficientNet model for the identification of helmets, heads, and hats in three categories. Han et al. [4] and others have proposed SSD target detection algorithms through a cross-layer attentional mechanism refinement of features so as to achieve improved accuracy in detecting helmets. The computer vision algorithm developed by Siebert et al. [5] automates the detection of bicycle helmet use in traffic videos with high accuracy. The YOLO-PL algorithm proposed by Li et al. [6] is based on YOLOv4, which optimizes the accuracy and efficiency of helmet detection. These vision methods have low cost but are susceptible to illumination and occlusion, and the algorithm complexity is high and less adaptable to the external environment.

To address the above problems, this study proposes an improved YOLOv8 algorithmic model for personal protection detection in chemical plants to improve the accuracy of chemical protection detection for small-scale targets, multi-distance multi-scale targets, and fuzzy small targets.

2 THE IMPROVED YOLOV8 ALGORITHM

In this paper, we optimize the backbone and neck part of the YOLOv8 algorithm model respectively, in which Bottleneck is optimized as MishNextBlock module in the backbone network, and two modules SCA and GAU are added to the PAN structure in PAN-FPN in the neck part to form the Multi-scale feature fusion and attention mechanism (MFAE-PAN) structure. scale feature fusion and attention mechanism (MFAE-PAN) structure. The improved YOLOv8 structure is shown in Fig. 1.

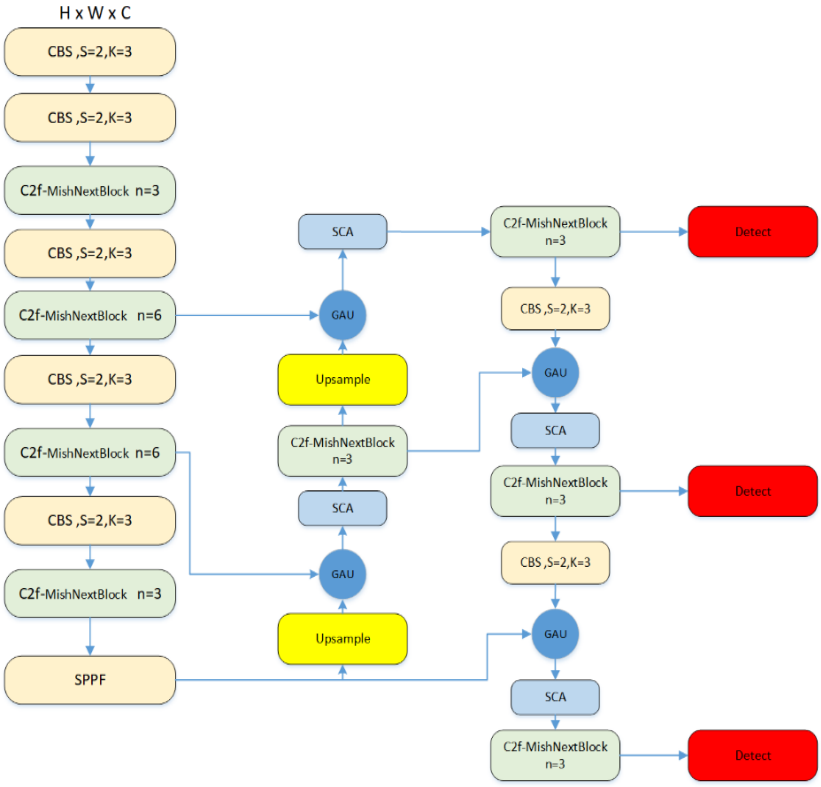


Fig. 1. Improved YOLOv8 algorithm model

2.1 Improvements to the backbone network of YOLOv8

In chemical plants, the scenario of a dense chemical personnel operating environment is very common, and therefore, the safety monitoring of densely populated areas is crucial. In order to improve the accuracy of personnel detection in such chemical plant environments, this paper introduces the MishNextBlock module with reference to the cascaded fusion network (CFNet) proposed by Zhang et al. [7]. The Bottleneck module in the original YOLOv8 algorithm model is replaced by the MishNextBlock module, which fuses both local and global features for interaction. In this case, the structure of the MishNextBlock module is shown in Fig. 2. The “d” in the figure indicates that the convolution used is deep convolution, and 5x5 is the corresponding convolution kernel size. The “a” in the figure indicates that it is the abbreviation of windowed attention mechanism, and the window size used in the MishNextBlock module is 5x5. The “e” in the figure indicates that it is the expansion factor, i.e., the multiplication rate from the number of channels of the input feature maps to the output feature maps, the n in the figure indicates that it is the number of channels of the feature maps, and the Mish in the figure indicates that the Mish activation function is used. denotes the Mish activation function used.

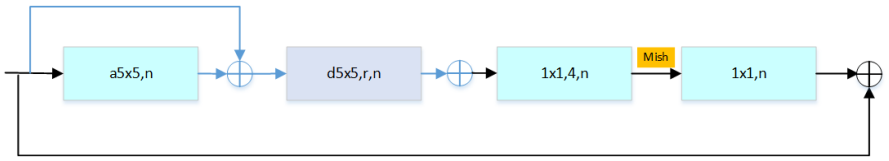


Fig. 2. MishNextBlock module structure diagram

Mish as an activation function is formulated as follows.

$$Mish(x) = x \otimes tanh(ln(1 + e^x)) \quad (1)$$

Where tanh(x) is the hyperbolic tangent function, which compresses the input values to compress them into the range of (-1,1). It is an S-shaped nonlinear function that provides the neural network with the necessary nonlinear properties to help the network learn complex data patterns; x is the neuron input information.

2.2 Improvements in the Neck section of YOLOv8

Aiming at the problems of poor image clarity, fuzzy edges, and difficult multi-target detection in chemical plant video surveillance, this paper optimizes the PAN module in YOLOv8, and proposes a PAN structure that combines multi-scale feature fusion and attention mechanism: multi-scale feature fusion and attention mechanism (Multi scale feature fusion and attention mechanism, MFAE-PAN). Among them, the module utilizes the concept of depth-separable convolution to effectively reduce the number of parameters while increasing the network depth. The multi-scale sensory field is formed by organizing different scale convolutional layers to capture image features. In addition, two key modules are incorporated into the module: the GAU and the SCA. where the GAU module fuses and optimizes the detection effect of different feature layers by up-sampling and fusing neighboring levels of feature maps to form a new integrated feature layer. The SCA module, on the other hand, is dedicated to highlighting the main features in these fused feature maps and focusing on important high-frequency feature information through the top-down transfer mechanism of the feature pyramid, which is especially critical for recognizing blurred edges and minute details in chemical plant images. Among them, the structure of MFAE-PAN is shown in Fig. 3.

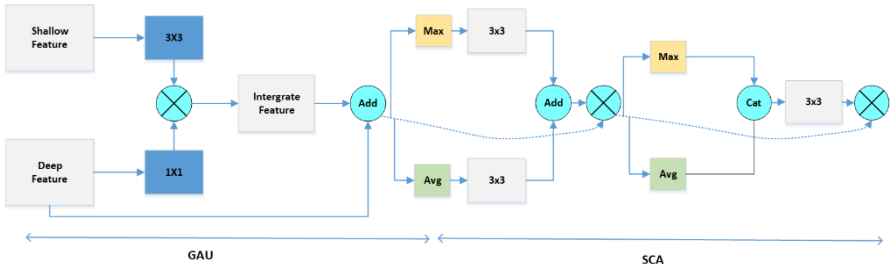


Fig. 3. MFAE-PAN module composition

As shown in Fig. 3, small convolutional kernels are used in the depth feature map of the GAU module due to the fact that small convolutional kernels improve the local sensitivity of the features in this feature map, reduce the number of parameters and the amount of computation, and also improve the nonlinear capability of the module. The shallow feature map in the GAU module is responsible for capturing the basic visual elements of the image such as edges, texture and color and other basic information, which helps the subsequent layers to process the high level features in a more refined way, so the main 3x3 convolution kernel is used. After the convolutional operation, the MFAE-PAN module performs the feature map fusion, after which the module connects the deeper features in order to preserve them through residual linking, which is mainly between the feature maps of the deeper parts of the network as described above and the fused feature maps as described above.

Where the above process can be expressed using equation (2), where F_{aug} is the final feature map of the GAU module, F_d is the feature map produced by the deeper convolutional layer, F_S is the feature map produced by the shallower convolutional layer, and the convolution in the module is expressed using $Cov_{k,s,p}$,where "k" is the size of the convolutional kernel, "s" is the step size, and "p" is the padding, the ⊗ denotes matrix multiplication. The computational formula is shown below:

$$F_{aug} = F_d + (Cov_{1,1,0}(F_d) \otimes Cov_{3,1,1}(F_S)) \tag{2}$$

In the overall operation of the module, the shallow feature map is convolved using the convolution block $Cov_{3,1,1}$, where "3,1,1" denotes the convolution kernel size, step size, and padding from front to back, respectively, and similarly, the deep convolution part uses the convolution block $Cov_{1,1,0}$. The two feature maps are multiplied using ⊗ and the result of the multiplication is residually concatenated with F_d to obtain the final output F_{aug}.Where the spatial attention module calculation process is shown in equation (3) below:

$$F_{SA} = F_{aug} \otimes (Cov_{1,1,0}\big(Avg(F_{aug}\big) + Cov_{3,1,1}(Max(F_{aug})) \tag{3}$$

As shown in Equation 3, where $Avg(\cdot)$ in the formula denotes average pooling and $Max(\cdot)$is the maximum pooling layer, the spatial attention output result F_{SA} is mainly obtained by convolving the maximum pooling result and the average pooling result of the input F_{aug} by the convolution blocks $Cov_{3,1,1}$ and $Cov_{1,1,0}$, respectively, and adding them together, and then after that, residual connecting of the added results with the original input feature maps F_{aug}, so as to obtain the final output result F_{SA}.

$$F_{CA} = F_{SA} \otimes Cov_{3,1,1}\big(Cat(Avg(F_{SA}), Max(F_{SA})\big) \tag{4}$$

Similarly, as shown in Equation 4, F_{CA}in Equation is the output of the channel attention mechanism , where $Cat(\cdot,\cdot)$ in Equation denotes the feature fusion, in which the main fusion F_{SA} after average pooling and maximum pooling, respectively, and the fused result is convolved by 3x3, and this step is mainly for the purpose of intercepting the deep information, and finally, the convolved result is concatenated with the previous spatial attention result F_{SA}, in order to obtain the output feature map F_{CA}.

3 EXPERIMENTAL RESULTS AND ANALYSIS

3.1 Datasets

The construction part of the chemical care products dataset mainly focuses on fusing the data pictures collected in the field of chemical plants with the actual working pictures of the Internet chemical plants, and screening out the positive and negative samples with different situations as much as possible. The dataset collects more than 10,000 pictures in total, which includes categories such as workers, safety helmets, and safety suits, and is divided into a training set, a testing set, and a validation set according to the ratio of 8:1:1.

3.2 Experimental environment and evaluation criteria

In designing the experiments in this subsection, the experimental environment involved is pytorch1.7.1, python3.9, and the GPU is NVDIA GeForce RTX 3080.The main parameters of the experimental data include Batchesize of 32, initial learning rate of 0.01, thread setting of 2, Momentum of 2, and a total of 300 training epoch. The evaluation metrics of the experimental results in this paper are Recall (R), Precision (P), Average Precision (AP) and mean average precision (mAP).

3.3 Experimental results and comparative analysis

The P-R plots of precision and recall for the three categories of chemical care products detection based on the improved YOLOv8 algorithm are shown in Fig.4. as follows:

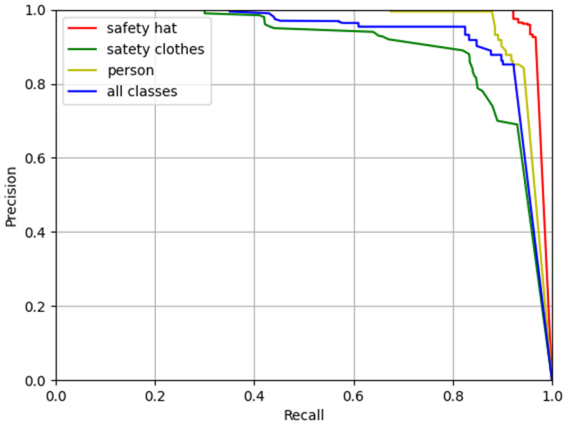
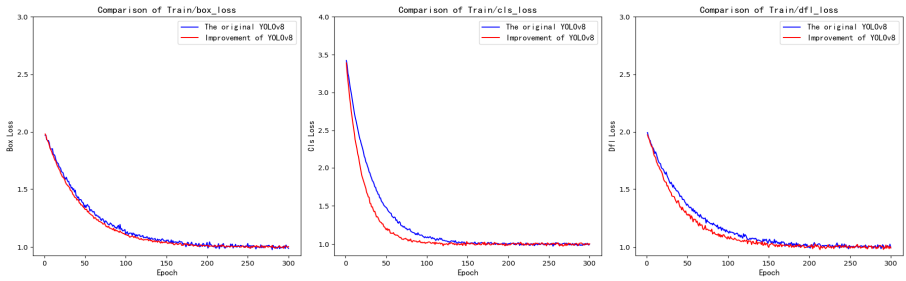


Fig. 4. Comparison chart of experimental results

The above figure shows that in the high recall interval, helmet recognition exhibits high precision due to the larger dataset, reflecting the model's ability to accurately detect helmets in most cases; whereas the precision for safety suits is relatively low.

Overall, the improved YOLOv8 algorithm demonstrates a balanced and efficient performance in chemical care products detection.

This section is mainly based on the optimized YOLOv8 algorithm model in the chapter and the actual performance of the unimproved YOLOv8 in the detection of chemical care products in the same environment, through the training and validation of the model with 300 Epoch, compared with the box_loss, cls_loss, dfl_loss indexes, and the results of the comparison are as follows in Fig.5.



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

Fig. 5. Comparison chart of experimental results. (a) Box Loss, (b) Distribution Focal Loss, (c) Classification Loss.

As shown by the three loss functions in Fig. 5, the optimized and adjusted model has a significant improvement in the decreasing speed of each type of loss, which confirms the effectiveness of the improved algorithm in improving the accuracy and robustness of target detection.

3.4 Ablation experiments

In this study, ablation experiments are also conducted, in which the main ablation targets include: the MishNextBlock module of the backbone network and the MFAE-PAN module of the Neck part in the improved YOLOv8. The results are shown in Table 1 below. where "✖" indicates that the model does not have this module and "✔" indicates that it contains this module.

Table 1. Comparison Table of Ablation Experiments

comparison group	MishNext-Block module	MFAE-PAN module	mAP50 /%	mAP0.95 /%	FPS /ms
1	✖	✖	95.75	96.85	20.3
2	✔	✖	97.23	97.21	21.8
3	✖	✔	97.56	98.22	21.4
4	✔	✔	98.19	98.32	22.2

As shown in Table 1 of the ablation experiments, in the improved YOLOv8 algorithm model composed of the original YOLOv8 algorithm with the addition of the

MishNextBlock module and the MFAE-PAN module respectively, the mAP0.5 accuracy is improved by about 2.44% and the mAP0.95 accuracy is improved by about 1.47%, although the target detection time is improved by a small margin. The overall detection accuracy is improved.

3.5 Visualization of experimental results

Comparing the performance of the original YOLOv8 algorithm with the improved YOLOv8 checking algorithm in the detection of chemical guards, the experimental results are extracted from the graphs, as shown in Fig. 6. below.

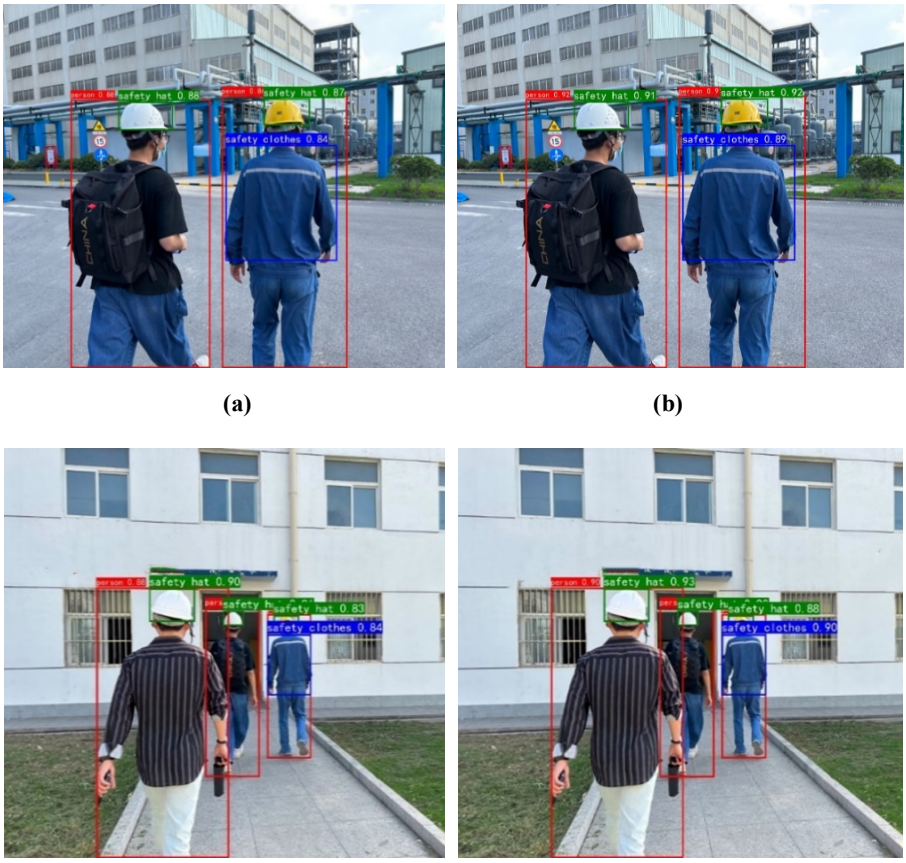


Fig. 6. Proximity and Long-Range Detection Results. (a) Proximity YOLOv8 detection, (b) Proximity Improvement YOLOv8 Detection, (c) Long-range YOLOv8 detection, (d) Long-range improvement of YOLOv8 detection

4 CONCLUSIONS

In this paper, for the problem of poor accuracy in multi-scale and small target detection in the detection of PPE in chemical plants, based on the latest YOLOv8 algorithm model, optimization is proposed in the backbone network as well as the Neck part of the model, respectively, in which the Bottleneck module is modified to MishNextBlock, and a multi-scale feature fusion and attention mechanism is proposed to be Combined PAN structure: MFAE-PAN, MFAE-PAN structure is mainly realized by adding two modules SCA (Spatial and Channel Attention) and GAU (Global Attention Up-sampling) module, the experimental analysis shows that the improved YOLOv8 algorithm can reach an average accuracy of 98.32%, and an average detection speed of 22.2ms, and both the The improved algorithm model in this paper shows high accuracy and high real-time performance for multi-scale multi-target detection at both near and far distances, which is of great practical application value in the field of chemical product detection.

ACKNOWLEDGMENT

This work was supported by the New Generation Information Technology Innovation Project [grant number: 2022IT123]; the Postgraduate Research & Practice Innovation Program of Huaiyin Institute of Technology [grant number: HGYK202319]; the Postgraduate Research & Practice Innovation Program of Jiangsu Province [grant number SJCX23_1858]; Joint Research Project of Jingshen Salt&Chemical Industry Co.,Ltd and Huaiyin Institute of Technology.

REFERENCES

1. Bukhari S M S, Zafar M H, Abou Houran M, et al. Secure and privacy-preserving intrusion detection in wireless sensor networks: Federated learning with SCNN-Bi-LSTM for enhanced reliability. Ad Hoc Networks, 2024, 155: 103407. https://doi.org/10.1016/j.adhoc.2024.103407
2. Uddin M N, Nyeem H. Engineering a multi-sensor surveillance system with secure alerting for next- generation threat detection and response. Results in Engineering, 2024, 22: 101984.https://doi.org/10.1016/j.rineng.2024.101984
3. Lee J Y, Choi W S, Choi S H. Verification and performance comparison of CNN-based algorithms for two-step helmet-wearing detection. Expert Systems with Applications, 2023, 225: 120096. https://doi.org/10.1016/j.eswa.2023.120096
4. Han G, Zhu M, Zhao X, et al. Method based on the cross-layer attention mechanism and multiscale perception for safety helmet-wearing detection. Computers and Electrical Engineering, 2021, 95: 107458.https://doi.org/10.1016/j.compeleceng.2021.107458
5. Siebert F W, Riis C, Janstrup K H, et al. Automated detection of bicycle helmets using deep learning[J]. Journal of Cycling and Micromobility Research, 2024, 2: 100013. https://doi.org/10.1016/j.jcmr.2024.100013

6. Li H, Wu D, Zhang W, et al. YOLO-PL: Helmet wearing detection algorithm based on improved YOLOv4. Digital Signal Processing, 2023: 104283. https://doi.org/10.1016/j.dsp.2023.104283
7. Zhang G, Li Z, Li J, et al. Cfnet: Cascade fusion network for dense prediction. arXiv preprint arXiv:2302.06052, 2023. https://doi.org/10.48550/arXiv.2302.06052