

# Improvement of Helmet Detection Algorithm Based on YOLOv8

Danyang Li

School of Computer Science and Engineering  
Xi'an Technological University  
Xi'an, China  
E-mail: xnldy169988@163.com

Jianguo Wang\*

Research Institute of Artificial Intelligence and  
Data Science  
Xi'an Technological University  
Xi'an, China  
E-mail: wjg\_xit@126.com  
\*Corresponding author

**Abstract**—In order to solve the problems of safety helmets in complex factory environments due to the complex background, dense targets, etc., which cause the YOLOv8s algorithm to be prone to leakage and misdetection, and low recognition accuracy, a safety casque detection algorithm based on the YOLOv8s improved YOLOv8s-improved is proposed. By incorporating a deformable convolutional module into the backbone network of YOLOv8s, the occurrences of false negatives and false positives are effectively reduced, and detection accuracy is enhanced. To tackle the issue of small target detection being easily disturbed by image backgrounds and noise, the CBAM attention mechanism is embedded to sift out the relatively important information from a large amount of information, and enhances the ability of helmet information extraction; for the problem that the loss of small target classification and localization is not easy to calculate, a new For the problem of small target classification and localization loss is not easy to calculate, a new IoU loss function is introduced to improve the training effect of the model. The experiment shows that the detection accuracy mAP of the improved YOLOv8s algorithm in this paper is 1.3% higher than that of the original YOLOv8s algorithm. Experimental results have shown that the improved algorithm proposed in this paper not only reduces false positives and false negatives in helmet wearing detection, but also enhances the detection capability for small targets, thus improving the performance of helmet wearing detection to a certain extent.

**Keywords**-Yolov8 Algorithm; Helmet Detection; Deformable Convolution; Attention Mechanism

## I. INTRODUCTION

It is essential to ensure the safety of people in factories, and helmets play a vital role in this

regard. Therefore, it is particularly important to monitor whether workers in factories are wearing helmets correctly. However, many complex construction sites still use manual inspection methods to carry out inspections. Undoubtedly, this method suffers from poor timeliness, low detection accuracy, and consumes a lot of labor costs. Therefore, it is of great significance for the research of realizing automatic detection of whether workers wear helmets correctly. In recent years, object detection methods based on deep neural networks have been widely applied in various industries in daily life. Compared with manual detection, deep learning-based detection methods greatly improve the detection efficiency of helmet wearing status, achieve real-time automatic detection around the clock, and effectively reduce labor costs.

## II. RELATED WORK

The object detection algorithm based on deep learning is divided into two level detection algorithm and one level detection algorithm.

One-stage target detection methods mainly convert the target discovery problem into a regression problem and solve it, and the typical algorithms are represented by YOLO series (e.g. YOLOv4, YOLOv5, YOLOv8, etc. algorithms, SSD algorithms). Two-level target detection algorithms, the detection task is carried out in two steps, first the generation of candidate regions, and then classification and localization, typical

algorithms are represented by R-CNN algorithm, Faster R-CNN algorithm.

In 2020, Hui Wang [1] proposed a new detection method based on the Faster R-CNN algorithm. Xie [2] et al proposed Drone YOLO model. But the model has more parameters, large computation and low real-time performance. Based on YOLOv3, Kai Xu [3] et al. mitigated the positive and negative sample imbalance problem by increasing the feature map and using K-means clustering algorithm. Jin Yufang [4] et al. proposed an improved YOLOv4 algorithm that combines the Head classifier with multilevel features by narrowing down the target features and improving the feature fusion module. Song Xiaofeng [5] and others, on the other hand, proposed a helmet wearing detection method that fuses feature environments and improves YOLOv5, which can better detect the small target of helmets.

YOLOv8 is highly scalable compared to other YOLO algorithms. In 2023 Ultralytics released the YOLOv8 algorithm [6][7]. Chen Yifang [8] and others suggested reconfiguring the feature extraction network and feature fusion network to achieve the goal of reducing computational load of the model, while introducing a deformable convolutional network (DCN) into the backbone network to strengthen feature selection capabilities; a global attention mechanism (GAM) was led into the neck network, thereby improving detection accuracy. Geng Huan [9] and others proposed a target detection algorithm in view of an upgrade YOLOv8 model, which can be deployed on the edge computing device, additionally improving test precision.

Although, the above research has improved the effect of helmet detection through different improvement methods but still have the following shortcomings: in target detection, small target labeling frame low resolution, distribution of dense and easy to overlap the problem; Small object detection is easily affected by image background and noise interference, information extraction ability is weak, recognition and localization accuracy is low; small target classification and localization loss is not easy to calculate. The YOLOv8s algorithm is used as the

baseline model, the variable convolution module added to the backbone network through the feeling wildness of the detection points on the feature map, the attention mechanism added to the neck side to strengthen the features of the small targets, and the IoU criterion designed for the network to be able to adaptively adjust the proportions of the various parts of the loss function at different stages to increase the detection ability of small goals. In the end, the optimized model is tested and verified on the casque data set. The method can be adapted to the complex scene of construction sites and obtain better helmet detection.

### III. PROPOSED METHODOLOGY

#### A. Deformable Convolution Improvement

The Deformable Convolution Network (DCN) family of algorithms is designed and proposed to enhance the model invariance to complex targets. In traditional convolution operation, the convolution kernel is a fixed rectangular structure, which is obsolete in DCN. DCN introduces deformable convolution, the convolution kernel can adopt the optimal convolution kernel structure based on different phases, feature maps or even pixel points. In DCN, each point on the convolution kernel learns an offset, which allows the convolution kernel to learn different structures in view of different parts of the data. This means that at each pixel point of the input feature map, we can learn a pair of offsets (x and y coordinates) that provide different convolutional effects for each location. These offsets are shared among different channels within the same feature map, thereby forming a deformable convolution module. The module structure is illustrated in Figure 1.

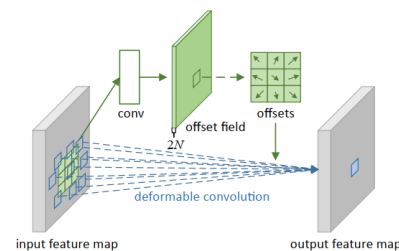


Figure 1. DCN module structure

Ordinary convolution is to sample a set of pixels from the enter feature map, and use

convolution operation to compute the sampling result to obtain the convolved result.

$$y(p_o) = \sum_{p_n \in R} w(p_n) \cdot x(p_o + p_n) \quad (1)$$

As for a deformable convolution, it is achieved indirectly by modifying the result of sampling to change the shape of the convolution kernel. Here we use  $\Delta p_n$  to represent  $p_n$  for expansion, where  $\{\Delta p_n \mid n = 1, 2, \dots, N\}$ , in which case the deformable convolution is computed as:

$$y(p_o) = \sum_{p_n \in R} w(p_n) \cdot x(p_o + p_n + \Delta p_n) \quad (2)$$

The offset obtained by the convolution operation as above is a small number and cannot be sampled directly for use. So DCN uses the sampling method of bilinear interpolation to achieve the effect of using the offset, which is described by the above equation.

The proposal of transformable folding aims to improve the flexibility of convolutional neural networks to target objects with irregular shapes and expand its acceptance range. The traditional convolution operation is only suitable for feature extraction on regular rectangular receptive fields, while in true conditions, lots of target objects possess inordinance fashion, which requires the use of deformable convolution to recognize these irregularly shaped target substances. flexible convolution can adaptively adjust the form and extent of the receptive field according to the irregular shape of the target object, and this feature improves the robustness of the CNN in dealing with complex scenes. Therefore, we choose to add deformable convolution module to the backbone network of YOLOv8s. By adding the deformable convolution module, more and more detailed image features can be extracted, thus laying the foundation for feature fusion and prediction on the detection head later. It is used to improve the model's focus on small and medium-sized targets. After adding the deformable convolution module, because the sensory field of the location of the small target is adaptively changed, the model can better adjust the regression parameters of the prediction frame when the prediction frame is regressed, to increase

the attention to the small target, and then improve the overall performance of the model.



Figure 2. YOLOv8s+DCN

### B. Attention mechanism CBAM improvement

In recent years, attention mechanisms have performed well in various deep learning tasks. Research has shown that attention mechanisms play a positive role in human visual perception, helping people efficiently and adaptively process visual information and focus on prominent areas of the image, thus enabling them to make the most accurate judgments. Among them, Convolutional Block Attention Module (CBAM) is a simple and effective attention module used to transmit convolutional neural networks. Our module leverages intermediate function cards to sequentially extract attention along two dimensions—channel and spatial. These attention cards are then multiplied with the input function cards to boost adaptability. As a lightweight and multifunctional module, CBAM can be easily integrated into any CNN architecture with minimal overhead and can be trained end-to-end alongside the CNN.

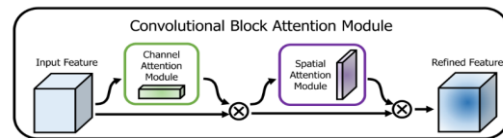


Figure 3. CBAM Network Architecture

There are several reasons to incorporate the CBAM attention module into the neck of YOLOv8s: (1) Enhanced feature representation: CBAM effectively adjusts channel and spatial weights in the feature map, allowing it to better

capture and represent key image features. (2) Improved computational efficiency: Compared to other attention modules, CBAM is less computationally demanding and more efficient (3).

**Better task adaptability:** CBAM is versatile and works well for a wide variety of visual tasks. The neck of the YOLOv8 network, situated between the backbone and prediction layers, plays a critical role in feature integration. Its structure allows for effective fusion of multi-scale features, which is crucial for accurate predictions. Thus, the design of the neck significantly impacts the algorithm's performance. In the modified network, as shown in Fig4, the CBAM module is placed after the un-sampling structure during the up-sampling phase of PAN-FPN and after each C2f module in the down-sampling phase, right before the CBS module convolution. This allows feature enhancement before fusion, enabling the model to focus more on small target details and improve the accuracy of both recognition and localization for small objects.

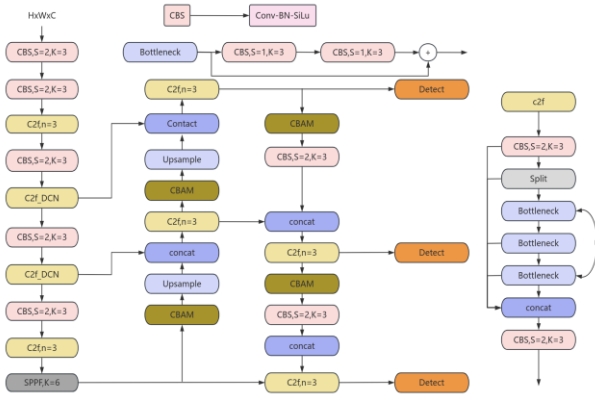


Figure 4. YOLOv8s+DCN+CBAM Network Architecture

### C. Improvement of IOU

IoU Loss function is a loss function commonly used in target detection to measure the Intersection over Union (IU) ratio between the true and predicted frames. This loss function evaluates the accuracy of prediction by calculating the ratio of the intersection over union of two bounding boxes. Specifically, the size of the two boxes is first compared, the area of their overlapping parts is calculated, and then the area

of the overlapping parts is divided by the total area of the two boxes to get the IoU Value. IoU The range of the value is [0,1], and a larger value indicates a higher similarity between the model prediction result and the real labeling. IoU The definition is as follows.

$$IOU = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

However, IoU has a big drawback. Firstly, if the two frames are not intersected, then  $IoU = 0$ , cannot reflect the size of the distance between the two. At the same time, because of the  $loss=0$ , there is no gradient back propagation, so we can't learn and train. Secondly, when the IoU ratio between the predicted and ground truth boxes is identical, the locations of the predicted boxes may still differ. Since the loss remains the same, this alone cannot determine which predicted box is more accurate.

Among them. Wise – IoU A dynamic approach is used to compute the category prediction loss in theIoU loss, defined as follows.

$$L_{WIOU} = R_{WIOU} L_{IOU} \quad (4)$$

$$R_{WIOU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (5)$$

Among them.  $W_g, H_g$ , is the size of the smallest closed box. To prevent  $R_{WIOU}$  generating gradients that prevent convergence,  $W, H$ , is separated from the computational graph (superscript \* indicates this operation), effectively eliminating the impediment to convergence.

In view of this, a new IoU criterion for loss function calculation, named GIOU is as follows:

$$L_{GIOU} = R_{WIOU} L_{IOU} + \frac{1}{2} \left( \frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*} \right) + \alpha v \quad (6)$$

The GIOU ability to dynamically adjust the bounding box regression loss is like that of theWise – IoU . At the beginning of training, when the prediction box and the true box are IoU is small, the model should focus on boosting



IoU. Currently, the  $R_{WIoU}L_{IoU}$  in the  $R_{WIoU}$  will be able to effectively increase the IoU smaller penalty strength. At the late stage of training, the prediction frame and the real frame are IoU is higher and stabilized.  $R_{WIoU}L_{IoU}$  The value in  $L_{IoU}$  then decays to a smaller value, and the model automatically shifts the focus to the regression of the center point and aspect ratio to further accurately predict the box position, thus improving the performance. GIoU Improvements are made to the traditional IoU intersection and concatenation operations in the model. Simultaneously, dynamically adjust the bounding box regression loss and mitigate the penalty on geometric metrics such as distance and aspect ratio. This enables more comprehensive consideration of the differences between predicted boxes and actual boxes in terms of IoU, position, size, and shape, thereby improving the accuracy of object detection. By halving the second and third terms of the loss function, the GIoU intervene in the punishment of geometric metrics at a lower level, avoiding too much intervention in model training and enhancing the generalization ability of the model.

In a nutshell. GIoU In each scenario compared to CIoU and Wise – IoU exhibits better adaptability and robustness to evaluate the target detection performance and classification tasks more effectively.

#### IV. EXPERIMENTATION

##### A. Introduction to the data set

The dataset is Safety-Helmet-Wearing, which contains 7581 images, including 6064 images in the training set and 1517 images in the validation set, and the dataset is labeled with Labelling tool. During the labeling process, the heads wearing helmets are labeled as "Helmet" and the heads not wearing helmets are labeled as "No Helmet".

Due to occlusion issues, the feature information in the image is weakened or disappeared. Therefore, the dataset annotation method is further optimized by reducing the detection contour of the safety helmet. The results are as follows:



Figure 5. Before



Figure 6. After

This improved annotation method will more accurately mark the key parts of the safety helmet, avoiding unnecessary background interference and greatly reducing the number of pixels required for annotation. This improved annotation method can mend the quality and stability of data.

##### B. Evaluation indicators

To evaluate the performance of the improved target detection model, the target detection model before and after the improvement is compared mainly in terms of detection accuracy and test speed. Firstly, taking the two types of categorized targets in this study as an example, when focusing only on all the images in the test set containing targets wearing helmets, the targets wearing helmets are counted as Positive cases (Positive), and the targets not wearing helmets are counted as Negative cases (Negative). The prediction results of the targets can be categorized into four cases: (1) True Positives (TP), which indicates lots of cases that were correctly classified as positive cases during the model detection process, in this paper, it is the number of instances of the target that actually wore a helmet and was classified as wearing a helmet by the YOLO model; (2) False Positives (FP), which indicates instances that were incorrectly classified as positive cases by the detected model. classified as positive instances, in this paper it is the number of instances where the target was not wearing a helmet but was classified by the YOLO model as wearing a helmet category; (3) True Negatives (TN), which indicates instances that were correctly classified by the detected model as negative instances, in this paper it is the number of instances where the target was

no-wearing a helmet and was classified by the YOLO model as wearing a helmet in the category of not wearing a helmet; and (4) False Negatives (FN), denoting the number of negative instances incorrectly classified by the detected model, in this paper the number of instances where the target was wearing a helmet but classified by the YOLO model in the category of not wearing a helmet.

Precision indicates the ratio of the number of correct samples detected to the total number of samples detected. In this paper, Precision refers to the proportion of images taken out as wearing helmet categories that are wearing helmets. It is calculated as  $Precision = TP / (TP + FP)$

Recall, also known as the check rate, indicates the ratio of the number of correct samples detected to the total number of samples in the test set. In this paper, it refers to how many images in the test set that contain the target wearing a helmet are correctly detected. A higher recall rate indicates better checking ability of the model. The specific formula for recall is as follows:  $Recall = TP / (TP + FN)$

Accuracy refers to the percentage of correctly detected samples in the total number of samples.  $Accuracy = TP + TN / (TP + TN + FP + FN)$

The curve formed by recall rate and accuracy, with recall rate as the horizontal axis and accuracy rate as the vertical axis, is called for the P-R curve. The area enclosed by the curve and coordinate axis is called the average accuracy, which is also an evaluation indicator for measuring the performance of the model on the dataset. *mAP* represents the mean of all categories of AP in the entire dataset in addition, under different threshold conditions, *mAP* will take different forms, among which *mAP@0.5* It refers to the average accuracy of all categories in object detection tasks when IoU reaches 0.5. In common object detection evaluations, *mAP@0.5* Usually used to evaluate the performance of algorithms.

### C. Experimental results

During the training phase, we used tensor board to record the loss function of the model on the training and validation sets. As can be seen

from the figure below, the training loss and validation loss of the model are gradually decreasing as the number of training times increases, indicating that the model keeps learning more accurate features. At the end of training, the model is evaluated on the dataset using the model and the following results are obtained.as shown in the figure7:

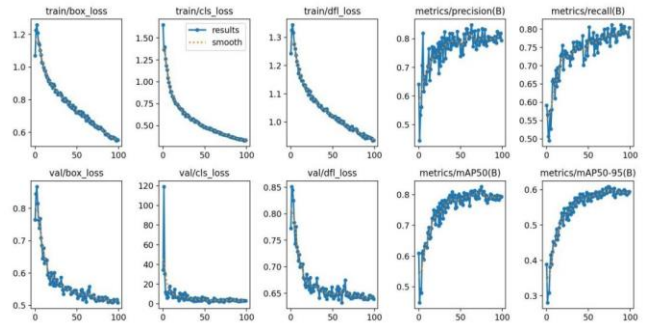


Figure 7. Loss Curve

Fig 8 illustrates the P-R curves of YOLOv8s, YOLOv8s-DCN, YOLOv8s-CBAM, YOLOv8s-DCN-CBAM, and YOLOv8s-improved models for each category on the Safety-Helmet-Wearing dataset.

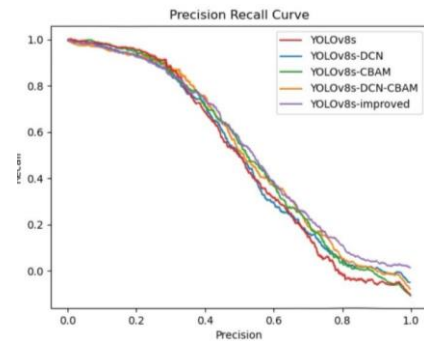


Figure 8. P-R curve

Table 1 shows the results of the ablation experiments on the dataset for different models:

TABLE I. EXPERIMENTAL RESULTS

Method				P (%)	R(%)	mAP@.5(%)
YOLO v8s	DC N	CBA M	GIOW			
√				67.3	58.9	64.6
√	√			72.6	56.3	64.3
√		√		71.4	60.1	65.7
√	√	√		70.1	58.4	65.7
√	√	√	√	72.5	59.3	65.9

As analyzed in figure4 P-R curve and Table 1, the YOLOv8s-improved algorithm is significantly more effective than other YOLOv8s-improved algorithms in the detection of safety helmets in terms of precision and recall. Adding CBAM attention module or DCN module separately to YOLOv8s can improve precision and recall respectively, but the overall detection performance is basically the same as that of YOLOv8s, which is not a great improvement to the overall performance of the model. From the last two rows of the table, adding the DCN module and the CBAM attention module to YOLOv8s, respectively, improves the accuracy rate while basically maintaining the recall rate, indicating that the added DCN module and CBAM module help to improve the network performance. Finally, the new IoU loss function calculation criterion is applied to the yolov8-DCN-CBAM model, and compared with YOLOv8s, YOLOv8s-improved can effectively improve the target miss detection while guaranteeing the recall rate.

## V. CONCLUSIONS

Using YOLOv8 as the basic model to realize the worker's helmet in complex scenes detection can timely detect and correct unsafe behaviors and effectively reduce number of incidents, minimizing casualties and economic losses. The purpose of this paper is to construct a helmet detection algorithm, which innovatively incorporates the improved YOLOv8s model and proposes the YOLOv8s-improved algorithm. The key to this improvement is to add realizable convolutional modules in the backbone network to improve the perceptual field of the points detected on the feature map; such a structural improvement injects the whole algorithm with a deeper level of information comprehension and improves the perceptual power of the model, which in turn improves the detection of workers wearing helmets. Increasing the attention mechanism in the neck to focus more on local details and edge information retention, which enables the model to better capture the target's textures, boundaries, and small changes, which significantly improves the detection and localization of small targets and

more effectively reduces the occurrence of missed detection. IoU Guidelines allow for Designing Networks makes it possible to adaptively adjust the proportions of each part of the loss function at different stages to better address the detection challenges in different scenarios. The experiments show that the improved YOLOv8s-improved model has improved precision in detection, increased recall, reduced missed detections, and the detection speed meets the requirements of real-time detection, which provides a certain means for the subsequent detection of factory workers' safety helmets in complex scenarios.

## REFERENCES

- [1] Hui Wang. Helmet Detection and Identity Recognition Based on Improved Faster R-CNN. Xi'an University of Science and Technology, 2020.
- [2] Chun hui Xie, JinMing Wu, Haiyu Xu. Improved small target detection algorithm for UAV imagery with YOLOv5. *Computer Engineering and Applications*, 2023, 59(9): 198-206. XIE C H, WU J M, XU H Y. Small object detection algorithm based on improved YOLOv5 in UAV image. *Computer Engineering and Applications*, 2023, 59(9): 198-206.
- [3] Xu K, Deng C. Helmet wearing recognition algorithm based on improved yolov3. *Progress in laser and Optoelectronics*, 2021, 58(06):300-307.
- [4] Jin F Y, Wu X, Dong H, et al. Helmet detection algorithm based on improved yolov4. *Computer science*, 2021, 48(11): 268-275.
- [5] XiaoFeng SONG, YunJun WU, BingBing LIU, et al. Improved YOLOv5s algorithm for helmet wearing detection. *Computer Engineering and Applications*, 2023, 59(2). 194-201. Song X F, Wu Y J, Liu B B, et al. Improved YOLOv5s algorithm for helmet wearing detection. *Computer Engineering and Applications*, 2023, 59(2): 194-201.
- [6] Redmon J, Divvala K S, Girshick B R, et al. You Only Look Once: Unified, Real-Time Object Detection. *CoRR*, 2015, abs/1506.02640.
- [7] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger. *CoRR*, 2016, abs/1612.08242.
- [8] Yifang Chen, Shang Zhang, Xiukang Ran, et al. Improved YOLOv8-based aircraft target detection algorithm for SAR images. *Telecommunication Technology*, 2023-08-04.
- [9] Huantong Geng, Zhenyu Liu, Jun Jiang, et al. An embedded road crack detection algorithm based on improved YOLOv8. *Computer Application*, 2023-08-28.
- [10] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks. *Proceedings of the IEEE international conference on computer vision*. 2017:7