scientific reports

Check for updates

OPEN

An improved YOLOv8 safety helmet wearing detection network

Xudong Song, Tiankai Zhang & Weiguo Yi✉

In the field of industrial safety, wearing helmets plays a vital role in ensuring workers' health. Aiming at addressing the complex background in the industrial environment, caused by differences in distance, the helmet small target wearing detection methods for misdetection and omission detection problems are needed. An improved YOLOv8 safety helmet wearing detection network is proposed to enhance the capture of details, improve multiscale feature processing and improve the accuracy of small target detection by introducing Dilation-wise residual attention module, atrous spatial pyramid pooling and normalized Wasserstein distance loss function. Experiments were conducted on the SHWD dataset, and the results showed that the mAP of the improved network improved to 92.0%, which exceeded that of the traditional target detection network in terms of accuracy, recall, and other key metrics. These findings further improved the detection of helmet wearing in complex environments and greatly enhanced the accuracy of detection.

Keywords YOLOv8, Attention mechanism, Pooled pyramid, Loss function, Safety helmet wearing detection

With the acceleration of industrialization, the safety supervision of construction sites has gradually received increased amounts of attention. Production safety accidents in housing and municipal engineering usually occur because site construction workers do not wear safety helmets. Although it has been clear that employees must correctly use safety equipment at work, there are still cases in which safety regulations are ignored during actual operation; therefore, it is especially important to detect the wearing of safety helmets.[1]However, traditional methods for detecting safety helmet, such as manual inspection and video surveillance, have problems such as incomplete coverage and slow response, making it difficult to meet the real-time detecting requirements of modern construction sites. This background has given rise to the study of safety helmet wearing detection using target detection technology.

Traditional safety detecting mainly relies on manual supervision by on-site safety personnel, a method that faces many challenges and drawbacks.[2] Firstly, construction sites usually have a vast footprint, complex and changing work scenarios, and a high degree of worker mobility, so it is difficult for manual supervision to achieve comprehensive coverage of the helmet-wearing status of all workers. This limitation makes it difficult to effectively carry out safety detecting and capture every worker's violation in real time. Secondly, in the traditional safety management system, a large number of safety supervisors are required to carry out prolonged supervision work, which not only consumes a large number of human resources, but also makes the supervisors fatigue easily by concentrating for a long period of time, thus affecting the accuracy and continuity of supervision. This problem is further exacerbated by the decentralized nature of workers' operations, the difficulty of safety management, and the lack of clarity of safety responsibilities. Finally, the manual supervision process inevitably involves subjective judgment, which may lead to inconsistency and inaccuracy in the supervision results. At the same time, due to the inefficiency of supervision, even if violations are detected, they may not be effective in preventing accidents due to the lack of timeliness.

With the progress of technology, the popularization of intelligent high-definition video surveillance system based on target detection technology[3] provides production and operation units with an intelligent safety detecting method that avoids human operation and has high real-time performance. Such systems are not only able to detect the working environment in real time, but also send timely warning messages to effectively prevent safety accidents. Although the application of intelligent detecting equipment has brought significant safety management improvements, detecting equipment still faces many challenges in helmet detection in the actual construction work environment. On the one hand, the detection effect of safety helmets is often affected by factors such as light intensity, changes in the background environment, equipment blockage, and changes in viewing angle and scale, and these environmental factors may lead to misdetection and omission of the detecting system, thus reducing the accuracy and reliability of detecting. On the other hand, in the complex construction environment, the safety helmet, as a relatively small target, is more difficult to detect. This not only requires the detecting system

Computer and Communication Engineering Institute, Dalian Jiaotong University, 794 Huanghe Road, Shahekou District, Dalian, Liaoning, China. ✉email: jiekexun98@163.com

to accurately recognize the helmet, but also to accurately determine its wearing status to ensure effective safety management.

In recent years, with the rapid development of deep learning technology[4,5], the target detection algorithm has also been significantly improved and perfected, especially in terms of detection accuracy and detection speed. Therefore, deep learning technology is utilized to improve the accuracy of detecting of helmet as well as small-target helmet wearing in complex backgrounds, which can not only effectively enhance the safety management of the construction site and safeguard the lives of the workers, but also reduce the potential safety accidents, and provide a strong guarantee for the safety of production in the construction industry.

In this experiment, we chose to use YOLOv8 network, and based on this, we introduced DWR[6] attention module in the backbone layer, which enhanced the feature extraction capability by multi-scale cavity convolution. Additionally, by introducing the ASPP[7,8] pooling pyramid, the information of different scales is captured, which more effectively handles the helmet detection task for helmets at different distances and background conditions. Finally, the introduction of NWD[9] loss function improves the regression accuracy and compensates the deficiency of CIoU in small target detection. The improved model proposed in this paper can effectively capture the details and improve the helmet detection accuracy in different scenarios, especially in the detection of small targets, thus greatly reducing the cases of misdetection and omission, and providing further technical support for ensuring the production safety in complex construction environments.

Related work

In recent years, a series of target detection networks based on deep learning have been proposed. The YOLO series algorithm[10–16] has gradually become one of the mainstream target detection algorithms due to its accuracy. For example, Liu et al.[17] proposed an improved YOLOv7 network (YOLOv7-AC) for underwater target detection, which improves the speed of feature extraction and network inference. Wang et al.[18] proposed a new network lightweight model, HV-YOLOv8, which effectively improves the detection accuracy of small targets and improves the adaptability to different small targets.

For helmet detection, Deng et al.[19] addressed the complexity and high resource demands of the YOLOv3 object detection algorithm by designing a new, lightweight version, ML-YOLOv3. The approach integrated the cross stage partial network (CSPNet) and GhostNet to form a more efficient residual network, CSP-Ghost-ResNet, and combines CSPNet with Darknet53 to create the new backbone network, ML-Darknet. Additionally, a lightweight multiscale feature extraction network, PAN-CSP-Network, was introduced. The resulting ML-YOLOv3 significantly reduced the floating point operations (FLOPs). Zhang et al.[20] presented a practical algorithm aimed at improving helmet detection, utilizing an enhanced version of the YOLOv5s algorithm. Firstly, the K-means method was utilized to recalibrate the size of the anchor boxes based on the dataset's label characteristics, which aims to boost the model's feature extraction accuracy. Secondly, an additional layer was integrated into the algorithm to bolster the model's capability in recognizing small-sized targets. Finally the attention mechanism was incorporated and the CIOU_Loss function was replaced with the EIOU_Loss function within the YOLOv5 framework to refine the model's precision. Li et al.[21] introduced YOLO-PL, an innovative and lightweight helmet detection algorithm derived from YOLOv4, which enhanced detection accuracy and efficiency. The development began with YOLO-P algorithms, which refined the network's ability to detect small objects and improved anchor assignment. The introduction of the Enhanced PAN (E-PAN) structure allowed for effective merging of information across different layers, improving detection accuracy. The study progressed by lightening the design through the Dilated Convolution Cross Stage Partial with X res units (DCSPX), optimizing the structure for enhanced lightness while maintaining performance, and replacing the conventional spatial pyramid pooling (SPP) module. Xia et al.[22] added a new feature output to the YOLOv5 network to detect small target helmets and used clustering methods to obtain a more appropriate prior anchor frame. Yi et al.[23] used the YOLOv5 algorithm to detect the helmet wearing situation of operators in complex scenes, which can accurately detect operators in motion, and also has better detection effect for the obscured helmets. Dai et al.[24] improved the sensitivity of the network for small target detection based on SSD using multilayer fusion to consider both shallow and deep semantic information. Tan et al.[25] improved YOLOv5 by introducing DIoU-NMS to increase the accuracy of suppressing the predicted bounding box. Fang et al.[26] established a large-scale data set and used the method of deep learning to detect the helmet. The optimization approaches for the YOLO network generally include the introduction of the attention mechanism[27–30], the improvement of the loss function[31,32], and the optimization of the pyramid pooling layer[7].

In summary, advancements have been made in the detection of safety helmets. Yet, there is a gap in targeted research within construction and engineering fields, where challenges such as identifying small-scale objects in complex backgrounds frequently result in detection failures or inaccuracies. Addressing the existing gaps in research, this paper explores a improved model for the detection of safety helmets, focusing on enhancing the accuracy of detection to better protect the safety of workers in the construction industry.

Methods

In the field of helmet detection, opting for an anchor-free YOLOv8 model circumvents the issues associated with fixed sizes and ratios inherent in traditional anchoring methods, which is particularly crucial for identifying safety helmets of various sizes and shapes. The anchor-free design streamlines the detection process, enhancing model training and generalization capabilities. Moreover, the method of directly regressing bounding boxes increases the detection accuracy for small objects and overlapping targets, a key advantage in complex construction scenarios. Additionally, YOLOv8 maintains the high-speed detection characteristics of the YOLO series and further optimizes speed and accuracy by eliminating anchors, making it more suitable for real-time safety

monitoring. Therefore, based on these considerations, choosing to improve based on the anchor-free YOLOv8 model becomes a rational choice.

YOLOv8 network

YOLOv8 is a leading end-to-end target detection network model that continues and builds on the core ideas of the YOLO series. Its structure is divided into four main parts: the input, the backbone layer, the neck hybrid feature network layer and the detect layer. On the input, YOLOv8 employs a variety of data enhancement techniques, including mosaic data enhancement and adaptive image scaling, to effectively enrich the training dataset. The backbone layer consists of an attention mechanism module, a cross-stage local network, and a spatial pyramid pooling structure, which work together to efficiently extract image features. The neck hybrid feature network layer utilizes the path aggregation network and feature pyramid network structure for multi-scale feature fusion, which enables the model to efficiently handle images of different scales. Finally, in the detect layer network, YOLOv8 employs decoupled detection headers optimized for classification and localization tasks, respectively, to further improve the accuracy and efficiency of detection.

Improved safety helmet wearing detection network

Despite YOLOv8's advancements in object detection, its performance in safety helmet detection can be compromised by distance variations and complex backgrounds typical in industrial settings, leading to potential misdetections and omissions. To address these issues, the network has been enhanced for more precise helmet detection.

As illustrated in Figure 3, the enhanced YOLOv8s model for helmet detection incorporates significant modifications: the DWR attention module is integrated into the backbone layer, and the SPPF module is substituted with the ASPP module in the network's backbone. This alteration not only broadens the receptive field of the feature map but also preserves the resolution of the original image. The model gains the ability to handle and interpret spatial information across various scales through the inclusion of receptive fields of differing sizes. Additionally, the adoption of NWD within the regression loss framework markedly diminishes localization errors while boosting accuracy. These modifications enable the model to precisely detect helmet usage across varying distances and in complex backgrounds, enhancing the detection of smaller targets.

DWR attention mechanism

The DWR module is also called Dilation-wise residual attention module, and its structure is realized in residual mode. As shown in Fig. 1, the multiscale information is efficiently extracted by the "RR–SR" two-step method (where RR is Region Residualization, and SR is Semantic Residualization), and then the generated feature maps with multiscale receptive fields are fused.

Specifically, the first step is RR, which is implemented by a conventional 3×3 convolution combined with a bulk normalization (BN) layer and a ReLU layer and is used to generate a series of concise feature maps of different sizes. These feature maps will be used as materials for the second step of morphological filtering. The second step is SR, which employs multirate null depth convolution to morphologically filter the features of regions of different sizes separately. As a result of this step, only one desired receptive field is applied to each channel feature to exclude possible redundant receptive fields.

With the above approach, the role of multirate null depth convolution changes from trying to obtain as much complex information as possible to simply morphologically filtering each succinctly expressed feature map to obtain multiscale information more efficiently.

After extracting multiscale information, multiple outputs are aggregated. All the feature maps were concatenated and then batch normalized. Then, the features are merged using point-by-point convolution to form the final residuals. Finally, the final residuals are added to the input feature maps to construct a more robust and comprehensive feature representation DWR that enhances the feature extraction capability through multiscale null convolution, which enables the network to effectively capture details and contextual information, thus improving the detection accuracy in complex scenes. In addition, the improved ability of DWR to handle targets at different scales enhances the network's generalization capability and makes it more adaptable to environmental changes, which is crucial for stability and reliability in practical applications.

ASPP pooled pyramid

In the original YOLOv8, when utilizing spatial pyramid pooling fast (SPPF) for feature extraction, since the main focus is on global feature extraction, it is not effective enough to capture fine-grained local features; this approach may become a bottleneck when accurately locating helmets occupying smaller portions of the frame and not providing sufficient contextual information.

In this regard, this paper introduces ASPP, whose structure is shown in Fig. 2. ASPP consists of two parts in parallel, the first of which performs dimensionality reduction via one 1×1 convolution and constructs the pooling pyramid. The corresponding 3×3 null-space convolutional layers are stacked to extract the features at different scales. The second part first applies global average pooling, which reduces the feature maps to a numerical value that captures the global contextual information of the image. Then, a 1×1 convolutional layer with a convolutional kernel of 256 is connected to output a single mean feature, which is subsequently upsampled back to the size of the original feature map by bilinear interpolation. Finally, the outputs of the two-part null convolutional and global pooling layers are connected, and a 1×1 convolutional layer is used to fuse the compressed multiscale features to generate the final output feature map.

ASPP is able to capture information at different scales through its null convolution with different sampling rates, and through global context fusion, it is able to handle the task of safety helmet detection at different

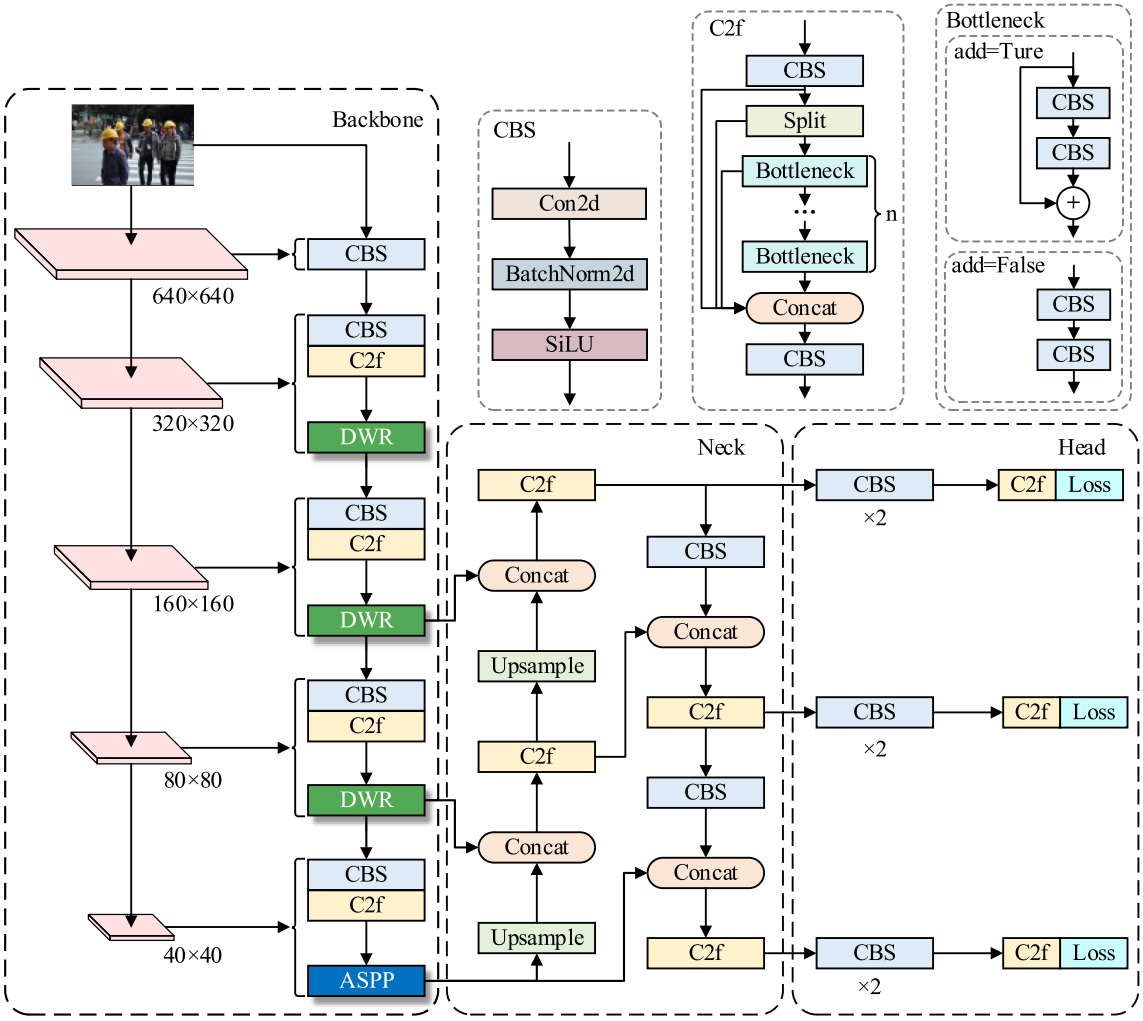

Figure 1. Improved YOLOv8 safety helmet wearing detection network(CBS modules are used to extract the initial features. The C2f. module is a residual feature learning module that enriches the gradient flow of the model through cross-layer connections, resulting in a neural network module with a stronger feature representation capability. The DWR is add to enhance the model's focus on relevant features, improving feature representation and overall detection accuracy. The ASPP module replaces the SPPF module, which uses a combination of serial and parallel maximum pooling operations to amplify different receptive fields and output feature maps with adaptive sizes. The loss function consists of Complete Intersection over Union and NWD.)

distances and under different background conditions more efficiently, especially when accurately locating small targets. This improvement enables the network to achieve greater accuracy and robustness in handling helmet detection in diverse and complex site environments.

NWD loss function

The original YOLOv8 used the CIoU as the default loss letter number, which is defined as follows:

$$R_{\mathrm{CIoU}} = \frac{\rho^2\left(b, b^{\mathrm{gt}}\right)}{c^2} + \alpha v \tag{1}$$

where b and b^{gt} represent the centers of the two bounding boxes of the prediction box and the real box, respectively; ρ is the Euclidean distance between the centers of the two bounding boxes; c is the diagonal length of the smallest closed region containing the two bounding boxes; and α is a weight parameter for balancing the effects of overlapping regions and aspect ratios, defined as follows:

$$\alpha = \frac{v}{(1 - \mathrm{IoU}) + v} \tag{2}$$

IoU denotes the intersection and concurrency ratio between the predicted bounding box and the true bounding box, and v measures the consistency of the aspect ratio, defined as follows:

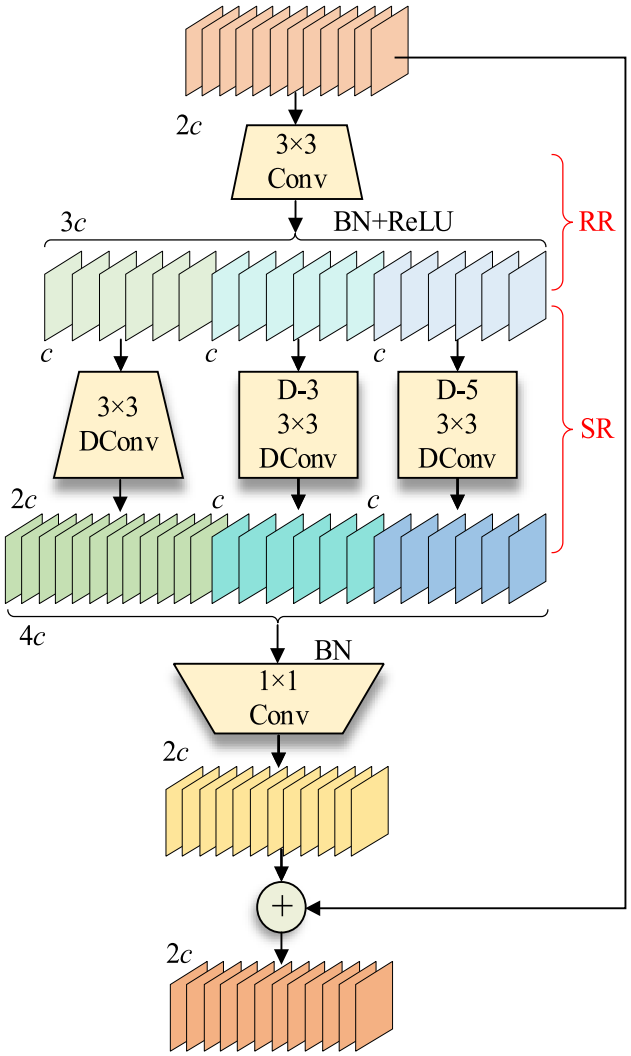

Figure 2. Flowchart of the DWR.

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \tag{3}$$

w,h and w^{gt},h^{gt} are the width and height of the predicted and real bounding boxes, respectively. Thus, the complete CIoU is defined as follows:

$$L_{CIoU} = 1 - \mathrm{IoU} + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v \tag{4}$$

The CIoU has several limitations in small target detection. The CIoU is more sensitive to slight positional deviations in small targets, especially when the helmet occupies a small area in the image, which may lead to performance degradation. To address the above limitations, this paper introduces NWD as part of the regression loss in the YOLOv8 network. The network is first modeled by representing the bounding box as a two-dimensional Gaussian distribution. Specifically, for the horizontal border $R = (cx, cy, w, h)$, where (cx, cy), w and h denote the center coordinates, width and height, respectively. Its interior elliptic equation can be expressed as:

$$\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} = 1 \tag{5}$$

(μ_x, μ_y) is the center coordinates of the ellipse and σ_x, σ_y are the lengths of the semi-axes along the x and y axes. Therefore, $\mu_x = cx$,$\mu_y = cy$,$\sigma_x = w/2$,$\sigma_y = h/2$.

The probability density function of a two-dimensional Gaussian distribution is:

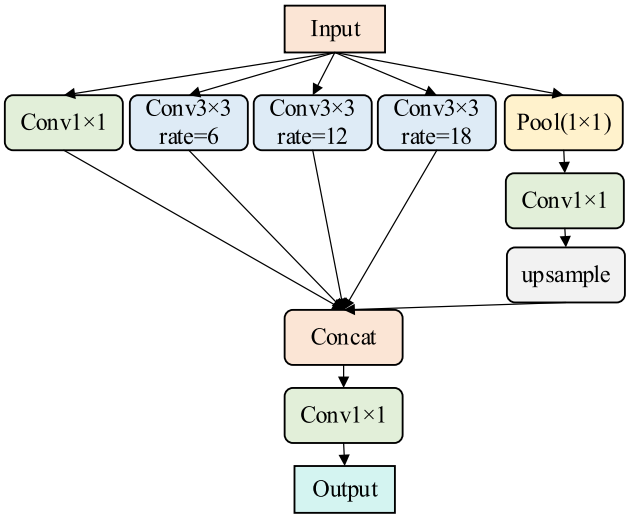

Figure 3. Schematic diagram of ASPP.

$$f(x|\mu,\Sigma)=\frac{\exp\left(-\frac{1}{2}(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)\right)}{2\pi|\Sigma|^{\frac{1}{2}}} \tag{6}$$

where x,μ,Σ are the covariance matrices of coordinate (x, y), mean vector and Gaussian distribution. When $(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)=1$, The ellipse in Eq. 5 will be the density profile of a two-dimensional Gaussian distribution. Thus the horizontal bounding box $R=(cx, cy, w, h)$ can be modeled as a 2D Gaussian distribution $N(\mu, \Sigma)$, where:

$$\mu=\begin{bmatrix}c_x\\c_y\end{bmatrix},\Sigma=\begin{bmatrix}\frac{w^2}{4}&0\\0&\frac{h^2}{4}\end{bmatrix}$$

Then, the similarity between the predicted and real targets is calculated by comparing their corresponding Gaussian distributions. The Wasserstein distance from Optimal Transport theory is used to calculate the distribution distance. For two Gaussian distributions $\mu_1=N(m_1,\Sigma_2)$ and $\mu_2=N(m_2,\Sigma_2)$, the Wasserstein distance between μ_1 and μ_2 is:

$$W_2^2(\mu_1,\mu_2)=\|m_1-m_2\|_2^2+\mathrm{Tr}\left(\Sigma_1+\Sigma_2-2\left(\Sigma_2^{1/2}\Sigma_1\Sigma_2^{1/2}\right)^{1/2}\right) \tag{7}$$

The above equation can be simplified as:

$$W_2^2(\mu_1,\mu_2)=\|m_1-m_2\|_2^2+\left\|\Sigma_1^{1/2}-\Sigma_2^{1/2}\right\|_F^2 \tag{8}$$

where $||\cdot||_F$ is Frobenius norm. And for Gaussian distributions N_a and N_b modeled by $A=\left(cx_a, cy_a, w_a, h_a\right)$ and $B=\left(cx_b, cy_b, w_b, h_b\right)$, the above equation can be further simplified as:

$$W_2^2(\mathcal{N}_a,\mathcal{N}_b)=\left\|\left(\left[cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2}\right]^{\mathrm{T}}, \left[cx_b, cy_b, \frac{w_b}{2}\cdot\frac{h_b}{2}\right]^{\mathrm{T}}\right)\right\|_2^2 \tag{9}$$

But $W_2^2(\mathcal{N}_a,\mathcal{N}_b)$ is a distance metric and cannot be used directly as a similarity metric (i.e., values between 0 and 1 as IoU). Therefore it is normalized using its exponential form to obtain a new metric, called Normalized Wasserstein distance (NWD). It is defined as follows:

$$NWD(\mathcal{N}_a,\mathcal{N}_b)=\exp\left(-\frac{\sqrt{W_2^2(\mathcal{N}_a,\mathcal{N}_b)}}{C}\right) \tag{10}$$

C is a constant closely related to the dataset. $\mathcal{N}_a$ and $\mathcal{N}_b$ represent the predicted bounding box and the true bounding box, respectively, which are modeled as two-dimensional Gaussian distributions.

The NWD is independent of the target scale and is suitable for measuring the similarity between small targets. In helmet detection, the prediction accuracy is effectively measured even when the size of the helmet in the image is low. So the NWD indicator is chosen and designed as a loss function:

$$\mathcal{L}_{NWD} = 1 - NWD(\mathcal{N}_p, \mathcal{N}_g) \tag{11}$$

$\mathcal{N}_p$ is the Gaussian distribution model for the prediction frame p and $\mathcal{N}_g$ is the Gaussian distribution model for Ground Truth box g.

With the above method, the introduction of NWD can effectively measure the similarity between the predicted frame and the real frame, reduce positioning errors, compensate for the shortcomings of the CIoU in small target detection, and improve the accuracy and robustness of detection.

Experiments and analysis of results

Datasets

In this paper, we use the publicly available SHWD safety helmet wearing detection dataset, which contains 7581 images covering 9044 helmet-wearing heads (positive samples) and 111,514 ordinary heads without helmets (negative samples). The dataset also contains images of workers taken at different distances in different scenes, which can effectively demonstrate the effects of multiple scales. To prevent overfitting due to insufficient data volume, the original dataset was augmented with data in this paper. The augmented dataset contains 10,581 images, and a Python script was used to randomly assign the training set (train) and test set (val) with 8:2 weighting (Fig. 3).

Experimental environment

In this study, the experimental environment utilizes the Windows10 operating system, with programming carried out in Python. Model training, and result testing are all conducted within the PyTorch , leveraging the CUDA (compute unified device architecture). The configuration details are outlined in Table 1.

Network training

During the training process of the YOLOv8 model, in order to optimize the model performance, this study defines specific hyper-parameters during the training process, as shown in Table 2.

Experimental evaluation indicators

In this paper, the precision (P), recall (R), average precision (AP) and mean average precision (mAP) are used as the main evaluation metrics of performance. The commonly used mAP0.5 and mAP0.5:0.95 are chosen as evaluation indices in terms of accuracy. Its specific formula is as follows:

$$P = \frac{TP}{FP + TP} \tag{12}$$

$$R = \frac{TP}{TP + FN} \tag{13}$$

Experimental configuration	Specification
Operating system	Windows10
CPU	Intel Xeon W-2265
GPU	NVIDIA GeForce RTX 3090
IDE	PyCharm
Development language	Python3.8.15
Deep learning frame	Pytorch1.13.1
CUDA	CUDA11.6

Table 1. Experimental configuration.

Name of the training parameter	Parameter value
Epochs	100
Batch	16
Initial learning rate	0.01
Momentum	0.937
Weight_decay	0.0005

Table 2. Network training hyperparameters.

$$AP = \int_{P}^{1} (R)dR \tag{14}$$

$$mAP = \frac{1}{n}\sum AP \tag{15}$$

TP is the number of correctly identified positive samples,*FP* is the number of incorrectly identified positive samples,*FN* is the number of positive samples not correctly identified, and *n* is the number of categories.

Analysis of the experimental results of the improved network

Performance measures

The developed network's efficacy is shown in graphs, which show different metrics of the performance of training and validation sets. Two separate types of loss are depicted in Fig. 4 where box_loss indicates the error between the prediction box and the calibration box and cls_boss indicates the accuracy of calculating the anchor box and the corresponding calibration classification. By observing the changes in each loss, it can be seen that the loss value of the improved network decreases and gradually converges, which proves that the model training has achieved good results.

Figure 5. shows the change in each evaluation index. Observing the changes in the four evaluation indices, it can be seen that the detection accuracy increases and stabilizes at the highest value, which shows that the improved network performs very well in terms of both the convergence effect and learning effect.

We better represent the performance improvement of the model proposed in this study by comparing the loss curve as well as the precision-recall curve with other models. As shown in Fig. 6, the commonly used evaluation metric for precision-recall curve is the area under the PR curve, i.e., PR AUC (Area Under the PR Curve), which ranges from 0 to 1, and the larger the value, the better the performance. The maximum large area under the PR curve of the model proposed in this study indicates the best performance of the classification model with the best trade-off between precision and recall. As shown in Fig. 7, compared with other networks, our network maintains a high level of convergence speed while the loss value is lower than other networks. This shows that the improved network proposed in this paper reflects a great advantage in performance.

The detection results of the improved network are shown in Fig. 8. Due to the default use of mosaic data augmentation, each picture was stitched together by a random number and size of pictures, in which 0 represents the person wearing a helmet and 1 represents the person not wearing a helmet. The figure shows that even a target that occupies a small part of the picture can effectively detect whether the helmet is worn, which shows that the improved network in this paper greatly improves the quality of helmet detection, which is highly effective.

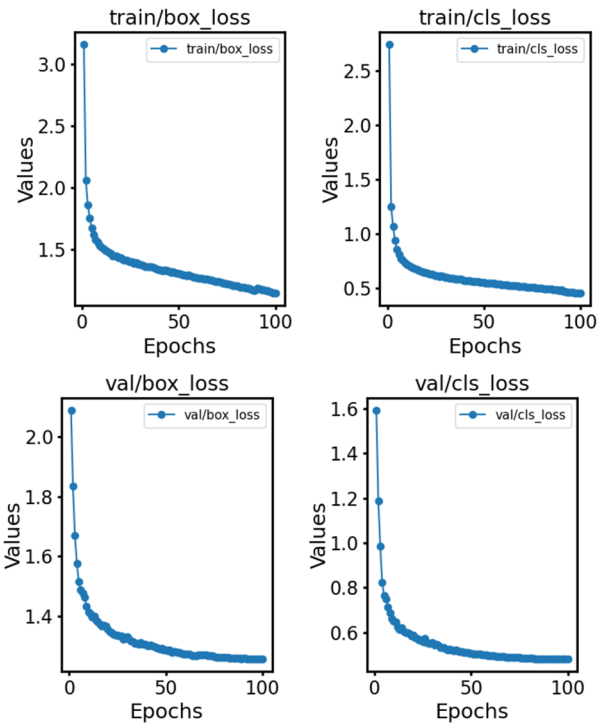

Figure 4. Loss profile of the improved network on the helmet dataset.

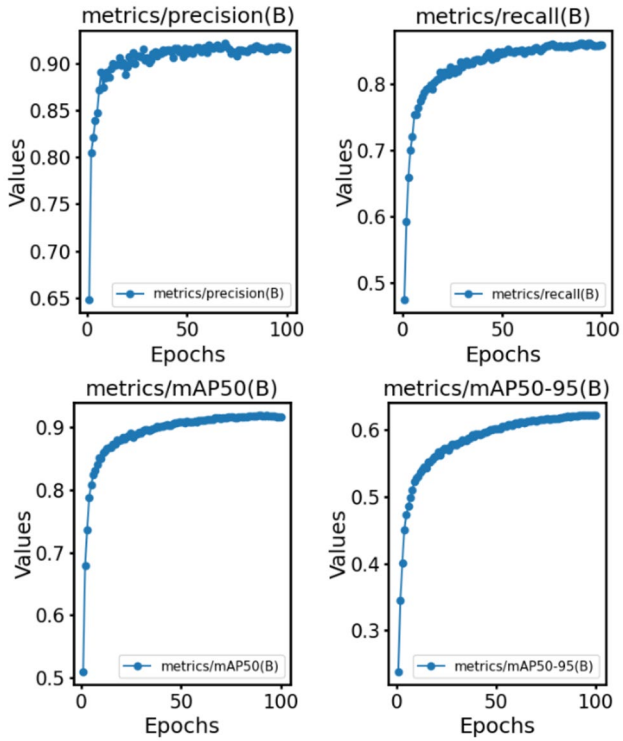

Figure 5. Evaluation metrics of the improved network on the helmet dataset.

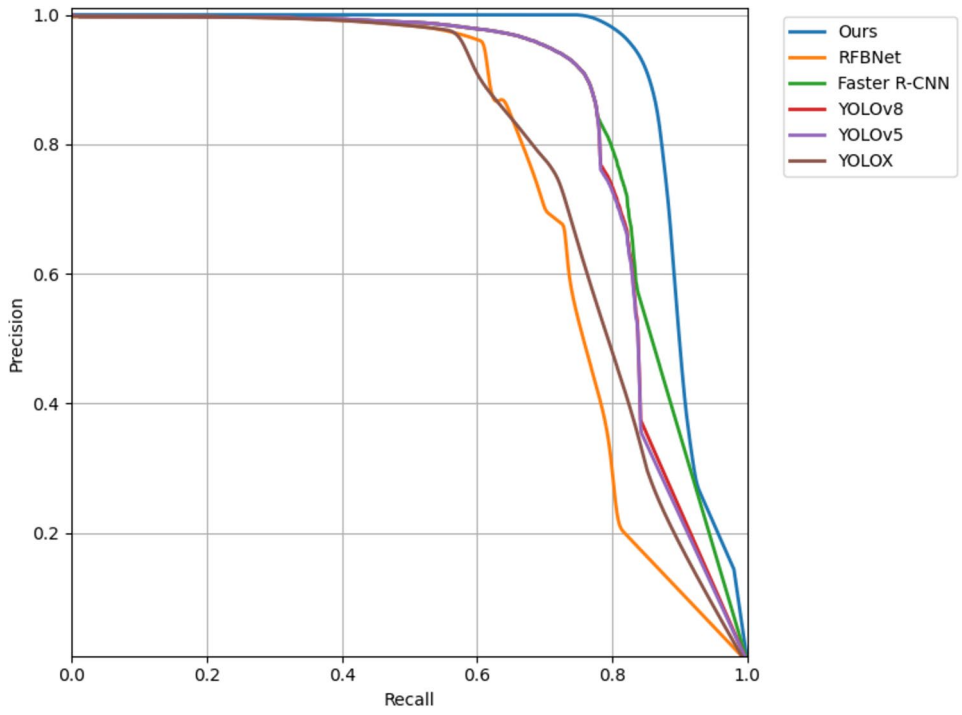

Figure 6. The precision-recall curve of other networks versus our network.

Ablation experiments

DWR module ablation experiments

Due to the complex background of helmet detection and different distances from the detection equipment, these factors can lead to misdiagnosis and omission of detection results, so the DWR module is introduced to enhance

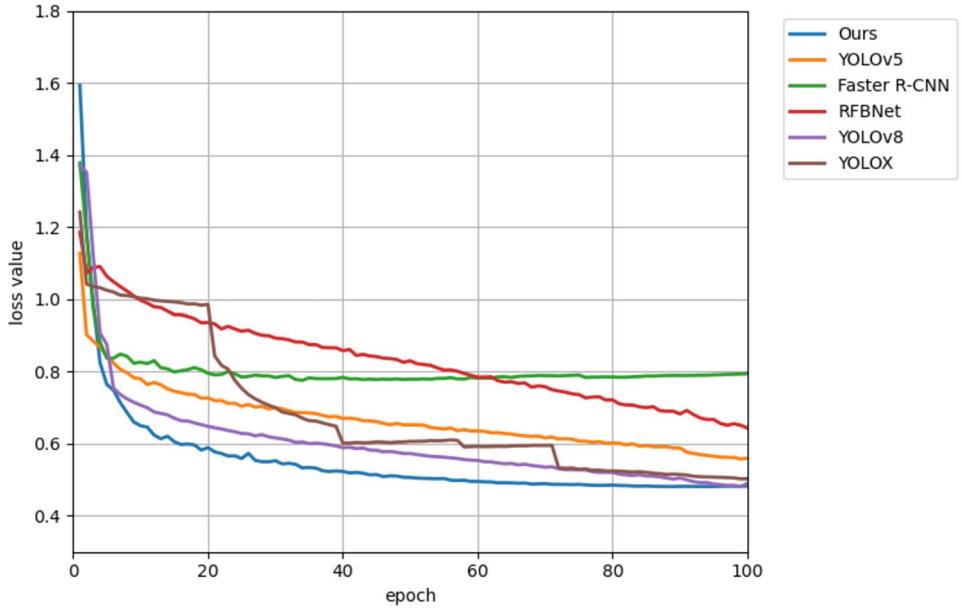

Figure 7. The loss curve of other networks versus our network.

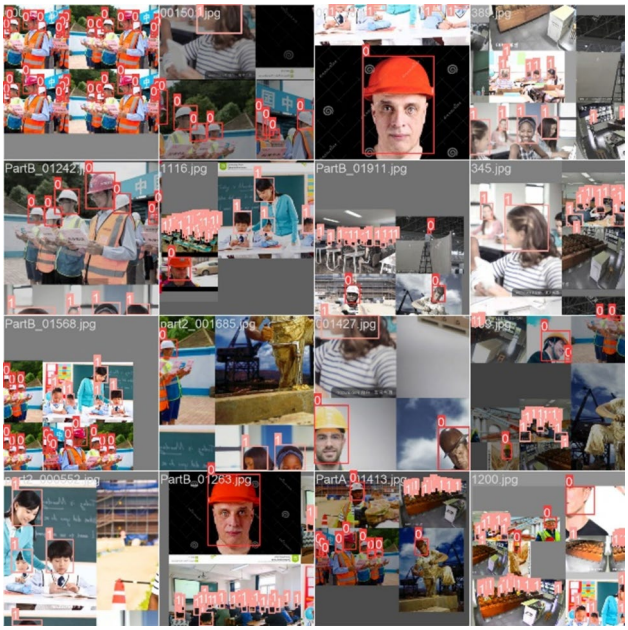

Figure 8. Safety helmet wearing detection results after network improvement.

the feature extraction ability of the network. The DWR module is added on the basis of YOLOv8s network and compared with other different mainstream attention for experiments to verify its effectiveness.

Table 3 illustrates the comparative analysis of the network's performance with the introduction of the DWR module. It shows an enhancement of 2.58% in Precision, 6.54% in Recall, and 4.97% in mAP over the conventional network. Conversely, the addition of the ECA attention module marginally increases the mAP by 0.35% but does not achieve improvements in other performance metrics. Although the CBMA attention module slightly surpasses DWR in Precision by 0.11%, it demonstrates lower performance in Recall and mAP by 2.36% and 1.32%, respectively. Therefore, the overall effectiveness in detection is superior with the DWR module compared to the other attention module.

From the analysis in Table 3, it can be seen that the introduction of the DWR module enhances the feature extraction capability, the multi-scale processing effect, and the model generalization performance of YOLOv8 in helmet target detection.

Network	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95
v8_s	89.2	79.5	86.5	0.542
v8_s_CBAM	91.6	82.7	89.6	0.582
v8_s_ECA	86.9	79.4	86.8	0.505
v8_s_DWR	91.5	84.7	90.8	0.604

Table 3. Attention module ablation experiment results.

ASPP module ablation experiments

To capture multi-scale helmet details, the original network's SPPF module was replaced with the ASPP module. This change was compared against SimSPPF and RFB modules. As shown in Table 4, the SimSPPF led to a modest increase in Precision by 0.22% and Recall by 0.51%, albeit with a slight reduction in Precision by 0.12%, compared to the original setup. The RFB improved all performance metrics, yet it fell short of the ASPP's enhancements, with Precision, Recall, and mAP lower by 0.44%, 2.31%, and 1.34% respectively. In stark contrast, the integration of the ASPP module significantly boosted the model's performance: precision increased by 1.12%, Recall by 3.40%, and mAP by 2.77%, all compared to the original network configuration.

From this analysis, it can be seen that ASPP is able to capture details and contextual information more effectively through its multi-scale null convolution, especially when dealing with different scale targets, which improves the accuracy of model detection.

NWD loss function ablation experiments

Table 5 presents a comparative analysis of the effects of various loss functions. Incorporating NWD into CIoU significantly enhances the model's performance, yielding a 3.12% higher mAP and reducing the Box loss by 29.6% compared to the original network. In contrast, the WIoU loss function does not match the effectiveness of the original network configuration. The introduction of the SIoU loss function improves mAP and decreases Box loss, however, when compared to the NWD, there's a deficit of 2.01% in mAP and a 40.03% increase in Box loss. Similarly, the performance improvements with DIoU do not rival those achieved by NWD. This evidence strongly supports that the integration of NWD with the original network not only boosts the detection accuracy of small targets but also significantly diminishes localization errors.

Integral network ablation experiments

To validate the improvement effects of the YOLOv8 network, a series of ablation experiments were designed to explore how these improvements affect the network's performance on the helmet detection task by introducing different combinations of modules, such as ASPP, NWD, and DWR, into YOLOv8. As shown in Table 6, the performances of the various combinations are evaluated by key metrics to reveal the specific contributions of the different modules to the detection effect.

According to the data presented in the Table 6, the addition of DWR, ASPP, and NWD individually contributes to notable improvements in the original network's metrics. However, the collective integration of these modules optimizes the model, resulting in significant enhancements: precision is augmented by 2.91%,

Network	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95
v8_s	89.2	79.5	86.5	0.542
v8_s_SimSPPF	89.4	79.9	86.4	0.505
v8_s_RFB	89.8	80.3	87.7	0.554
v8_s_ASPP	90.2	82.2	88.9	0.556

Table 4. Pooled pyramid module ablation experiment results.

network	mAP@0.5/%	mAP@0.5:0.95	Box_loss
v8_s(CIoU)	86.5	0.542	1.395
v8_s_WIoU	86.5	0.533	1.411
v8_s_SIoU	87.4	0.545	1.385
v8_s_DIoU	88.6	0.522	1.366
v8_s_NWD	89.2	0.561	0.982

Table 5. Loss function ablation experiment results.

Network	P/%	R/%	mAP@ 0.5/%	mAP@ 0.5:0.95
v8_s	89.2	79.5	86.5	0.542
v8_s_DWR	91.5	84.7	90.8	0.604
v8_s_ASPP	90.2	82.2	88.9	0.556
v8_s_NWD	91.2	82.1	89.2	0.561
v8_s_DWR_ASPP	91.5	84.9	90.9	0.606
v8_s_DWR_NWD	92.5	85.6	91.6	0.617
v8_s_ASPP_NWD	90.7	84.2	90.5	0.601
Ours	91.8	86.6	92.0	0.622

Table 6. Ablation experiment results.

Network	P/%	R/%	mAP@ 0.5/%	mAP@ 0.5:0.95
Faster R-CNN	89.5	80.5	87.4	0.541
RFBNet	82.6	73.4	75.6	0.416
YOLOv5	88.4	78.6	85.7	0.554
YOLOX	91.0	81.7	88.3	0.549
YOLOv8	89.2	79.5	86.5	0.542
SSD[24]	88.3	76.7	84.2	0.516
DAAM-YOLOv5[25]	87.7	78.6	85.3	0.526
Ours	91.8	86.6	92.0	0.622

Table 7. Comparative experiment results.

Recall by 8.93%, and mAP by 6.36%. These results highlight the complementary nature of the modules: DWR improves detail recognition, ASPP enhances multi-scale feature processing, and NWD reduces localization errors while ensuring the accuracy of small target detection. Their unified application not only boosts the model's overall accuracy but also its robustness and adaptability across varied scenarios, culminating in optimal helmet detection performance.

Comparative experiments

To comprehensively evaluate the performance of the improved YOLOv8 network on the helmet detection task, comparative experiments were conducted with several classical target detection networks, including traditional networks such as Faster R-CNN, RFBNet, YOLOv5 and YOLOX, and networks in the literature referenced in papers 24 and 25. As shown in Table 7, the network proposed in this paper outperforms Faster R-CNN by 5.26% in mAP, RFBNet by 21.69%, YOLOv5 by 7.35%, and YOLOX by 4.19%, and also outperforms other methods in all other metrics. It can be seen that the method proposed in this thesis reflects obvious advantages in terms of detection accuracy compared with other methods.

Conclusion

The improved YOLOv8 network proposed in this paper provides significant improvement in safety helmet wearing detection. In response to the problem of a large number of small-target helmets being detected due to the small frame occupied by helmets caused by distance and other reasons, the detection capability for small targets has been significantly improved by the introduction of the DWR, ASPP and NWD modules. The DWR module enhances the feature extraction capability of the network, the ASPP module improves the network's multiscale processing effect, and the NWD module optimizes the network's detection accuracy for small targets. The experimental results show that these improvements and enhancements cause the network to outperform the traditional target detection network in key metrics such as accuracy, recall and mAP. Therefore, the improved method proposed in this paper effectively improves the accuracy of safety helmet wearing detection, especially in small target helmet detection.

Data availability

The data used in this paper is public and has been deposited on GitHub at https://github.com/njvisionpower/Safety-Helmet-Wearing-Dataset. The data is from a third party and the ownership of the data is Njvisionpower. The above website provides a usage License. Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files.

Received: 3 February 2024; Accepted: 23 July 2024
Published online: 30 July 2024

References

1. Zhang, J., Qu, P., Sun, C. & Luo, M. Safety helmet wearing detection method based on improved YOLOv5. *J. Comput. Appl.* **42**, 1292–1300 (2022).
2. Geng, J. & Ren, B. N. Application of fuzzy comprehensive evaluation in the bid evaluation of municipal engineering construction projects. *Appl. Mech. Mater.* **584**, 2159–2164 (2014).
3. Li, W., Feng, X. S., Zha, K., Li, S. & Zhu, H. S. In *Journal of Physics: Conference Series*. 012003 (IOP Publishing).
4. Qi, S. *et al.* Two-dimensional electromagnetic solver based on deep learning technique. *IEEE J. Multiscale Multiphys. Compu. Tech.* **5**, 83–88 (2020).
5. Sadad, T. *et al.* Brain tumor detection and multi-classification using advanced deep learning techniques. *Microsc. Res. Tech.* **84**, 1296–1308 (2021).
6. Wei, H. *et al.* DWRSeg: Dilation-wise residual network for real-time semantic segmentation. arXiv preprint arXiv:2212.01173 (2022).
7. Lian, X., Pang, Y., Han, J. & Pan, J. Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation. *Pattern Recognit.* **110**, 107622 (2021).
8. He, H., Yang, D., Wang, S., Wang, S. & Li, Y. Road extraction by using atrous spatial pyramid pooling integrated encoder-decoder network and structural similarity loss. *Remote Sens.* **11**, 1015 (2019).
9. Yu, Z. *et al.* Yolo-facev2: A scale and occlusion aware face detector. arXiv preprint arXiv:2208.02019 (2022).
10. Chen, W., Huang, H., Peng, S., Zhou, C. & Zhang, C. YOLO-face: a real-time face detector. *Visual Comput.* **37**, 805–813 (2021).
11. Adibhatla, V. A. *et al.* Applying deep learning to defect detection in printed circuit boards via a newest model of you-only-look-once (2021).
12. Jocher, G. *et al.* ultralytics/yolov5: v6. 0-YOLOv5n'Nano'models, Roboflow integration, TensorFlow export, OpenCV DNN support. *Zenodo* (2021).
13. Guo, Z., Wang, C., Yang, G., Huang, Z. & Li, G. Msft-yolo: Improved yolov5 based on transformer for detecting defects of steel surface. *Sensors* **22**, 3467 (2022).
14. Kim, J.-H., Kim, N., Park, Y. W. & Won, C. S. Object detection and classification based on YOLO-V5 with improved maritime dataset. *J. Mar. Sci. Eng.* **10**, 377 (2022).
15. Wang, G. *et al.* UAV-YOLOv8: a small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors* **23**, 7190 (2023).
16. Zhang, Y. *et al.* Complete and accurate holly fruits counting using YOLOX object detection. *Comput. Electron. Agric.* **198**, 107062 (2022).
17. Liu, K. *et al.* Underwater target detection based on improved YOLOv7. *J. Mar. Sci. Eng.* **11**, 677 (2023).
18. Wang, W., Meng, Y., Li, S. & Zhang, C. *Hv-Yolov8 by Hdpconv: Better Lightweight Detectors for Small Object Detection*. Available at SSRN 4632283
19. Deng, L., Li, H., Liu, H. & Gu, J. A lightweight YOLOv3 algorithm used for safety helmet detection. *Sci. Rep.* **12**, 10981 (2022).
20. Zhang, Y.-J., Xiao, F.-S. & Lu, Z.-M. Helmet wearing state detection based on improved YOLOv5s. *Sensors* **22**, 9843 (2022).
21. Li, H., Wu, D., Zhang, W. & Xiao, C. YOLO-PL: Helmet wearing detection algorithm based on improved YOLOv4. *Digit. Signal Process.*, 104283 (2023).
22. Xia, Z. & Xiao, H. A study of campus environment security cap detection system based on YOLO v4. *Network Security Technology and Applications*, 40–41 (2021).
23. Yi, Z., Wu, G., Pan, X. & Tao, J. in *2021 33rd Chinese Control and Decision Conference (CCDC)*. 769–773 (IEEE).
24. Dai, B., Nie, Y., Cui, W., Liu, R. & Zheng, Z. In *Proceedings of the 2nd International Conference on Artificial Intelligence and Advanced Manufacture*. 95–99.
25. Tan, S., Lu, G., Jiang, Z. & Huang, L. In *2021 IEEE International Conference on Intelligence and Safety for Robotics (ISR)*. 330–333 (IEEE).
26. Fang, Q. *et al.* Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Autom. Constr.* **85**, 1–9 (2018).
27. Huang, H., Liang, Q., Luo, D. & Lee, D. H. Attention-enhanced one-stage algorithm for traffic sign detection and recognition. *J. Sens.* **2022** (2022).
28. Guo, M.-H., Liu, Z.-N., Mu, T.-J. & Hu, S.-M. Beyond self-attention: External attention using two linear layers for visual tasks. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 5436–5447 (2022).
29. Huang, H., Chen, Z., Zou, Y., Lu, M. & Chen, C. Channel prior convolutional attention for medical image segmentation. arXiv preprint arXiv:2306.05196 (2023).
30. Yu, Y., Zhang, Y., Cheng, Z., Song, Z. & Tang, C. MCA: Multidimensional collaborative attention in deep convolutional neural networks for image recognition. *Eng. Appl. Artif. Intell.* **126**, 107079 (2023).
31. Gevorgyan, Z. SIoU loss: More powerful learning for bounding box regression. arXiv preprint arXiv:2205.12740 (2022).
32. Zhang, S. *et al.* Diag-IoU Loss for Object Detection. *IEEE Transactions on Circuits and Systems for Video Technology* (2023).

Acknowledgements

This work was supported by the National Natural Science Foundation of China [No. 52175379] and the Liaoning Provincial Science and Technology Department [No. 2022JH2/101300268].

Author contributions

X.D.S. and T.K.Z. designed the concept and the experimental approach. T.K.Z. developed the model and performed the experiments. T.K.Z. wrote the first draft of the manuscript. X.D.S. and W.G.Y. reviewed the manuscript and corrected the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to W.Y.