scientific reports

Check for updates

OPEN

Lightweight detection model for safe wear at worksites using GPD-YOLOv8 algorithm

Jian Xing, Chenglong Zhan✉, Jiaqiang Ma, Zibo Chao & Ying Liu

To address the significantly elevated safety risks associated with construction workers' improper use of helmets and reflective clothing, we propose an enhanced YOLOv8 model tailored for safety wear detection. Firstly, this study introduces the P2 detection layer within the YOLOv8 architecture, which substantially enriches semantic feature representation. Additionally, a lightweight Ghost module is integrated to replace the original backbone of YOLOv8, thereby reducing the parameter count and computational burden. Moreover, we incorporate a Dynamic Head (Dyhead) that employs an attention mechanism to effectively extract features and spatial location information critical for site safety wear detection. This adaptation significantly enhances the model's representational power without adding computational overhead. Furthermore, we adopt an Exponential Moving Average (EMA) SlideLoss function, which not only boosts accuracy but also ensures the stability of our safety wear detection model's performance. Comparative evaluation of the experimental results indicates that our proposed model achieves a 6.2% improvement in mean Average Precision (mAP) compared to the baseline YOLOv8 model, while also increasing the detection speed by 55.88% in terms of frames per second (FPS).

Keywords Deep learning, Site safety wearable detection, YOLOv8, Ghost module

Since construction sites are mostly exposed to outdoor construction environments, there are more significant risk factors than in other industries. Despite countless efforts by government agencies to improve safety standards on construction sites, worker injuries and fatalities continue to occur [1–3]. In civil engineering construction, working-at-height scenarios are complex and varied, with many potential hazards. The leading causes of work-at-height injuries and deaths include personnel not wearing safety helmets as required and wearing irregularities, which can result in significant injuries and deaths[4]. Helmets are the most effective intervention to reduce the incidence and severity of head injuries[5].

Meanwhile, construction workers can improve the wearer's visibility in complex construction environments by wearing reflective vests. However, due to the hot weather, there are cases of construction workers not wearing them or wearing them incorrectly, so monitoring safe wearing has attracted significant attention in recent years[6]. There is a growing interest in safety wear because it plays a vital role in reducing accidental injuries and fatalities on construction sites[7]. Wearing safety helmets can reduce the fatality rate by 40% and the risk of serious injury by 70%[8]. Safety helmets play a decisive role in worker head protection. Because the traditional manual inspection management mode is inefficient and expensive[9], so deep learning, with its powerful learning ability, is widely used in image processing[10–12], intelligent construction site[13] detection technology has become the current trend.

A safety helmet, as well as reflective clothing wear detection, belongs to the category of target detection, which can be categorized into two-stage detection and single-stage detection. In two-stage detection, Regional Convolutional Neural Network (R-CNN)[14] and Fast Regional Convolutional Neural Network (Fast R-CNN)[15] were proposed, which can achieve 70% of the mAP value on the VOC2007 dataset. Still, the overall detection rate is slow, and then the Ultra-Fast Convolutional Neural Network (Faster R-CNN) was proposed successively by Ren et al.[16] detector. However, Fast R-CNN cannot share the parameters of multiple correlation regions in the second stage. The simultaneous use of fully connected layers may lead to information loss. Unlike two-stage detection, single-stage detection can directly generate detection bounding boxes and category probability values. The YOLO[17] model was proposed by Redmon et al. to convert the two-step detection process into an abstract regression problem. Liu et al. proposed a multiscale-based SSD[18] detection technique, which can efficiently discover multiple small targets. However, the SSD algorithm must be further optimized for secondary pre-processing after deep convolution. Shi et al. added feature pyramids in YOLOv3 to improve the recognition

School of Electronic Information Technology, Northeast Forestry University, Harbin 150040, People's Republic of China. ✉email: zcl9279sz@163.com

accuracy of people and helmets[19]. Wu et al. used the DenseNet network instead of the Darknet53 feature extraction network to enhance helmet detection accuracy in YOLOv3[20]. Song et al. used the parallel network RepVGG module to replace the Res8 module in the original YOLOv3 network[21], but the detection ability is poor for the occluded images. Qian et al. added Context Aggregation (CA) on top of the YOLOv5 backbone network and further parameter compression using depth separable convolution (DWConv) to improve the network detection accuracy[22]. Wang replaced the original network's Union Loss generalized intersection and replaced it with Distance Intersection over Union Loss to solve the problem of localization error when the population is dense[23]. Yung et al. trained three detection models using YOLOv5, YOLOv6, and YOLOv7 algorithms and summarised the advantages and disadvantages of each algorithm[24].

The above research is fundamental, but there are still the following problems: (1) Some algorithms still need to improve their computation of model and low detection accuracy. (2) Although some algorithms have high detection accuracy, the excessive parameters and computational redundancy significantly burden the equipment. (3) Helmet detection has received much attention but needs to be better integrated into reflective clothing.

Therefore, this paper proposes a safety wear detection model based on improved YOLOv8. To address these problems, this study adopts a lightweight network and an attention-aware module based on safety wear detection, which effectively solves the problem of too many features caused by the complex background and can effectively suppress the redundancy of information. A standard safety wear image dataset, including 5220 images in different environments, such as long-range small and densely covered targets, is established. From the perspective of the practicality of the detection results, this paper can meet the real-time detection accuracy requirements of the construction site and significantly reduce the leakage rate by using a lightweight network.

Related work

Dataset

To verify the superiority of this experimental model, the self-acquisition dataset for construction monitoring of high-pier large-span girder bridges of the Yinkun Expressway used in this paper consists of the publicly available SHWD safety helmet-wearing detection dataset[25]. The dataset has a total of 5220 images. For the dataset of safety helmets and reflective vests in this paper, to avoid the problems of overfitting caused by too little data, an unbalanced number of samples of the species, and background variations under the disturbances of complex environments at the site, data augmentation was performed on the original data to improve the robustness of the model by using nearest neighbor or bilinear interpolation, mean blurring of the images, deleting the percentage of pixels on the fly, changing the brightness of the images, and shifting the image. A Python script was used to divide the training set randomly, test set, and validation set with weights in the ratio of 8:1:1. The labels contain helmet, without_helmet, reflective_clothes, and other_clothes, respectively, to enhance the performance of the safety wear detection model by increasing the diversity of data. As in Fig. 1. On this basis, multiple performance metrics evaluate and validate the YOLOv8 model regarding precision, recall, and average accuracy. These metrics are developed to ensure the correctness of the model in real-world applications.

Experimental environment

The experiment is based on Pytorch1.10 framework, CUDA11.3 version for training, YOLOv8 is used as the base model, batch size is set to 16, the input image resolution size is 1024 × 1024, a total of 200 rounds of iterations, the learning rate is 0.01, and the Adam optimizer is used, the specific experimental configuration is shown in Table 1.

Performance evaluation metrics

To comprehensively de-validate the effectiveness of the safety wearable detection model, the mean Average Precision (mAP)[26] is chosen as the evaluation index. mAP is calculated by the precision P (Precision) and the recall R (Recall) together.

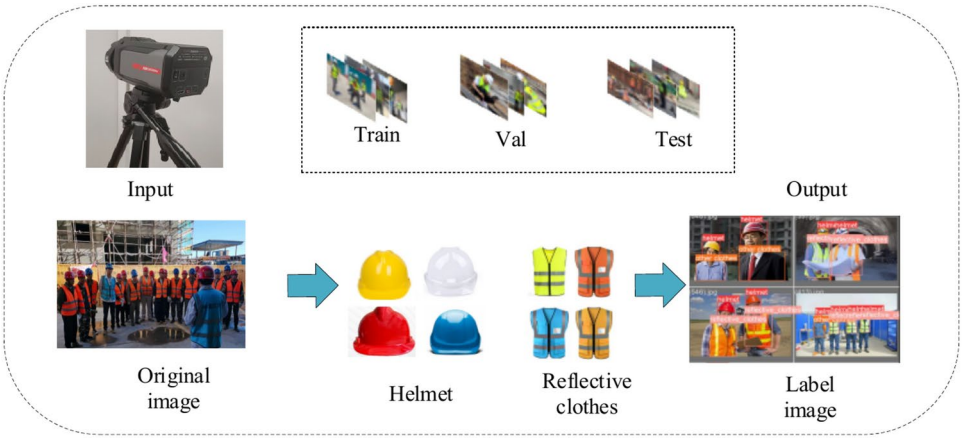

Fig. 1. Data set processing.

Project	Experimental environment
Operating system	Ubuntu
CPU	Intel(R) Xeon(R) Gold 6326 CPU @ 2.90 GHz
GPU	NVIDIA A100 80 GB PCIe
Random access memory	80 GB
Python version	3.8

Table 1. Experimental environment configuration.

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

To comprehensively assess the two indicators, P and R, the F1-score was introduced as the weighted harmonic mean of precision and recall, representing their balanced average with a range from 0 to 1.

$$F1 - score = 2\frac{P \times R}{P + R} \tag{3}$$

In the above equation, TP (true positives) denotes the number of correctly identified, FP (false positives) represents the number of backgrounds detected as targets, and FN (false negatives) indicates the number of targets detected as backgrounds. Precision P and recall R affect each other, and Average Precision AP (Average Precision) is introduced to combine both effects on model precision. In multi-class target detection, the average of multi-class AP values is usually calculated separately for the target classes to obtain the mAP of the model and evaluate its performance. The detection speed of a model is typically measured in frames per second (FPSs), indicating the number of frames detected per second.

$$AP = \int_0^1 P(R)dR \tag{4}$$

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k} \tag{5}$$

$$FPS = \frac{Framenum}{ElapsedTime} \tag{6}$$

Improved the model architecture of Yolov8

To achieve efficient detection of safety helmets and reflective clothing in complex backgrounds, for the existence of small targets with feature information challenging to extract and lack of feature expression ability and other issues, this paper introduces the P2 detection layer on the YOLOv8 model, which significantly enriches the semantic feature information, improves the detection accuracy at the same time can also effectively detect the miniature target safety helmets, using a lightweight Ghost network model instead of the YOLOv8's backbone, adopting Dyhead dynamic detection head, extracting features and spatial location information of site safety wear through the attention mechanism in its structure, significantly improving the accuracy of the expressive ability of the site safety wear detection model without increasing the computational load, and adopting EMASlideLoss loss function. Site safety wear detection accuracy is further improved while reducing the number of parameters and computational load. The structure of the overall enhanced YOLOv8 model is shown in Fig. 2.

P2 small target detection layer

Due to the complexity of construction site images, for long-distance images, safety helmets, and reflective clothing belong to small target detection, so introducing the P2 small target detection layer in the Head layer structure enables YOLOv8 to detect small target objects more effectively. Small targets usually occupy fewer pixels in an image and are, therefore, more likely to be ignored or misjudged. With the dedicated P2 layer[27], YOLOv8 can detect and localize small targets more acutely, improving the accuracy of small target detection for safety helmets and reflective clothing. Improvements are shown in Fig. 3.

Lightweight ghost network

In this paper, the network backbone network layer of YOLOv8 is replaced with the Ghost Module proposed by Han[28] to replace the original C3 structure. The number of parameters and load of the model are further reduced to improve the detection speed and accuracy.GhostNet is a lightweight neural network architecture based on the MobileNetv3 architecture by replacing the Bottleneck block with Ghost Bottlenecks and introducing the "Squeeze and Anomaly" (S&A) model. "Ghost Bottlenecks is a new residual module introduced in GhostNet, which uses the Ghost Module instead of the traditional depth-separable convolution, and the Ghost Module is a

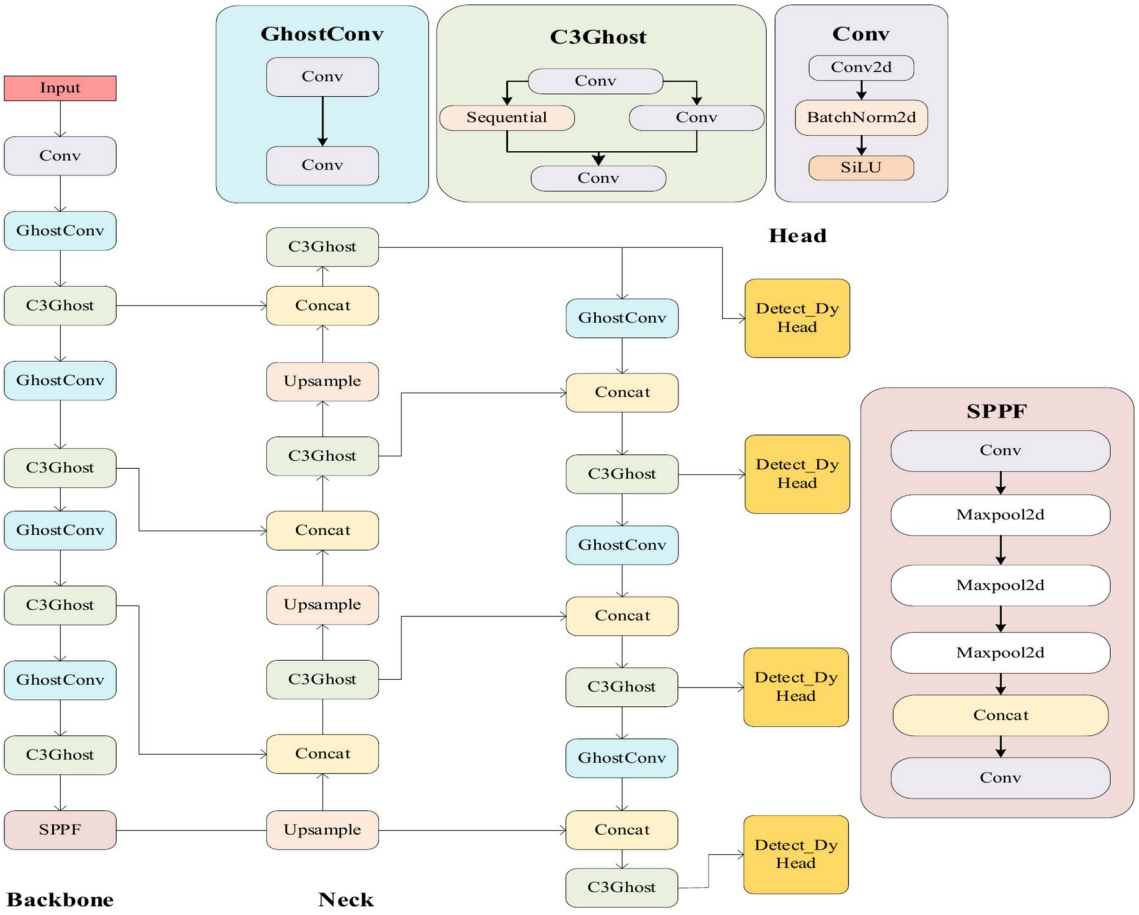

Fig. 2. Structure of the proposed YOLO-P2-Ghost-Dyhead model.

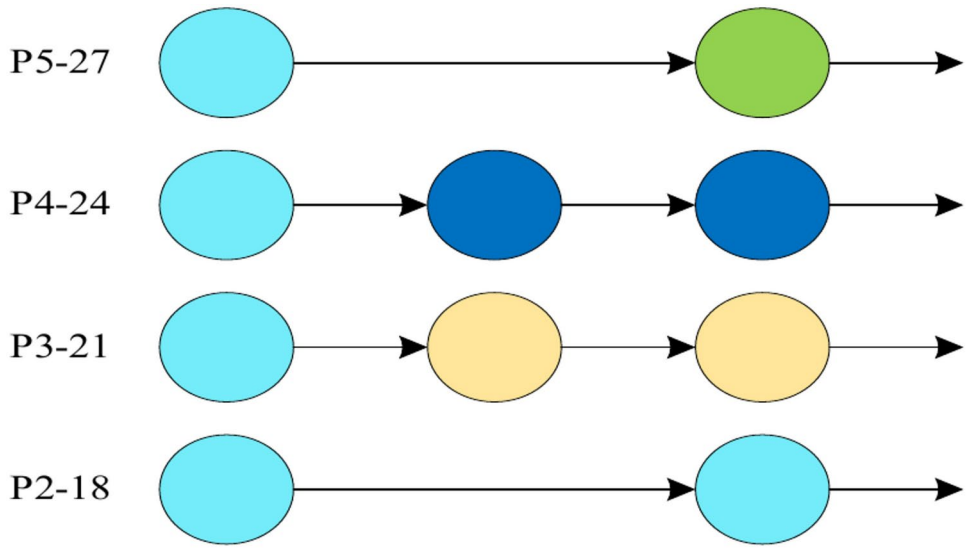

Fig. 3. Introduction of P2 small target detection layer.

new method for generating multiple low-dimensional The Ghost Module linearly combines multiple generated low-dimensional "ghost" feature maps into a final output feature result map, which improves the performance of the network by reducing the computational complexity. In addition, the SE (Squeeze-and-Excitation) structure is applied to GhostNet, which utilizes global information learning and channel attention mechanisms to help

the network better capture essential features and improve network performance. Figure 4 illustrates the specific principle.

Assume that the input feature map of size h×w×c and the output feature map of size h×w×c. The input feature maps are the heights and widths of h and w, and the output feature maps are the heights and widths of h and w. There is a constant map and m×(s − 1) = n×s×(s − 1) linear operations, each with a convolution kernel of size d×d and a regular convolution kernel of size k×k. Using linear operations of the same size in the Ghost module, after s transformations, d and k have similar sizes when s is much smaller than c. The theoretical ratio of the Ghost module to the standard convolution is:

$$\begin{aligned} r_s &= \frac{c' * h' * w' * c * k * k}{\frac{c'}{s} * h' * w' * c * k * k + (s-1) * \frac{c'}{s} * h' * w' * d * d} \\ &= \frac{c * k * k}{\frac{1}{s} * c * k * k + \frac{s-1}{s} * d * d} \\ &\approx \frac{s * c}{s + c - 1} \approx s \end{aligned} \tag{7}$$

$$\begin{aligned} r_c &= \frac{c' * c * k * k}{\frac{c'}{s} * c * k * k + (s-1) * \frac{c'}{s} * d * d} \\ &\approx \frac{s * c}{s + c - 1} \approx s \end{aligned} \tag{8}$$

In summary, by adopting the new lightweight convolution module Ghost Module, firstly, the computation amount is reduced by a small amount of convolution. Secondly, the convolution is carried out by superimposing the feature maps one by one by Φ in the figure. Finally, the number of parameters of FLOPs can be 1/s of the original one, which can further improve the detection speed while improving the detection accuracy.

Detection head optimisation

In YOLOv8, the number of YOLOv8 detection head parameters is significantly increased. It needs to be trained independently because YOLOv8 uses a decoupled head, which is divided into two branches, Cls classification, and Box regression, and removes the computational branch of Obj loss, which will result in misalignment of the feature space and a complex computational effort. In this experiment, the Dyhead target detection head[29] is used; the structure of Dyhead is illustrated in Fig. 5, which effectively unifies scale attention, spatial attention, and task attention in the same architecture and is used for coherently combining the multi-head self-attention attention mechanism, which significantly improves the effectiveness and representation capability of the target detection head, and makes the modeled network more adapted to complex safety wearable detection scenarios.

Given the feature tensor $\mathcal{F} \in R^{L \times S \times C}$, the generalised form of customisation can be described as follows:

$$W(\mathcal{F}) = \pi(\mathcal{F}) \cdot \mathcal{F} \tag{9}$$

By converting the attention function into three sequential attentions, each attention is allowed to focus on only one dimension:

$$W(\mathcal{F}) = \pi_C(\pi_S(\pi_L(\mathcal{F}) \cdot \mathcal{F}) \cdot \mathcal{F}) \cdot \mathcal{F} \tag{10}$$

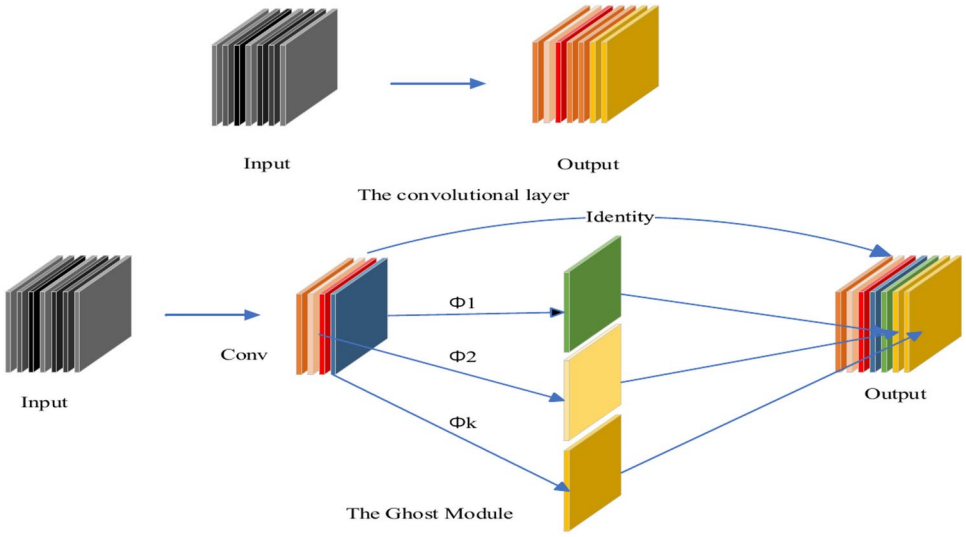

Fig. 4. Conventional convolution and Ghost module.

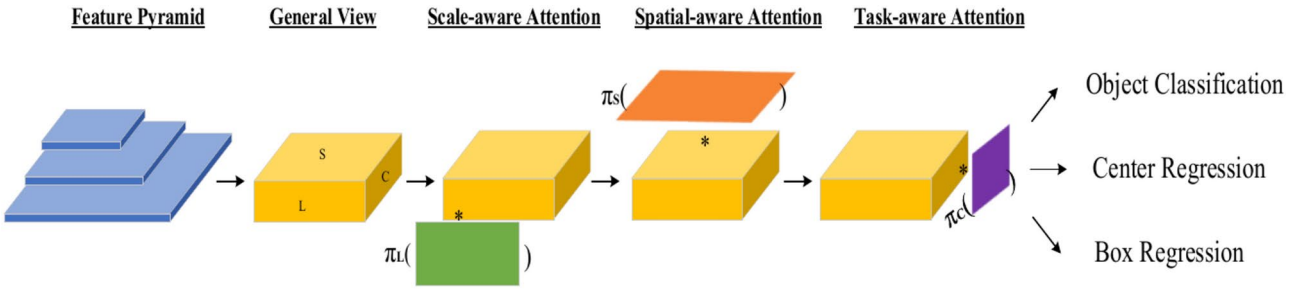

Fig. 5. Dyhead detection gead architecture.

Scale-aware Attention π_L We first introduce scale-aware attention to fuse different scale features based on their semantic importance:

$$\pi_L(\mathcal{F}) \cdot \mathcal{F} = \sigma\left(f\left(\frac{1}{SC}\sum_{S,C}\mathcal{F}\right)\right) \cdot \mathcal{F} \tag{11}$$

where $f(\cdot)$ is a linear function approximated by a 1×1 convolution and $\sigma(x)$ is a hard-sigmoid activation function.

Spatial-aware Attention π_S Next another spatial location-aware attention module is introduced to focus the discriminative power of different spatial locations. Due to the high latitude of S, it will be decoupled: the attention learning is first sparsified using deformation convolution, followed by feature cross-scale integration:

$$\pi_S(\mathcal{F}) \cdot \mathcal{F} = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K} w_{l,k} \cdot \mathcal{F}(l; p_k + \Delta_k; c) \cdot \Delta m_k \tag{12}$$

where K is the number of sparse sampling positions.

Task-aware Attention π_C To improve the generalisation of joint learning and goal expressivity, feature channels can be dynamically switched on and off to assist different tasks:

$$\pi_C(\mathcal{F}) \cdot \mathcal{F} = max(\alpha^1(\mathcal{F}) \cdot \mathcal{F}_c + \beta^1(\mathcal{F}), \alpha^2(\mathcal{F}) \cdot \mathcal{F}_C + \beta^2(\mathcal{F})) \tag{13}$$

where $[\alpha^1, \alpha^2, \beta^1, \beta^2]^T = \theta(\cdot)$ is a hyperparameter to control the activation threshold and $\theta(\bullet)$ is analogous to DyReLU. Finally, the above attention mechanisms are stacked several times in a sequential manner.

Through the above findings, the input target feature information processed by Dyhead's scale-aware attention module and spatial location-aware attention module becomes more sensitive to target detection at different scales. The features become sparser so that it can be more focussed on panoramic target detection at various locations, which further enhances the detection capability of the detection head and improves the detection accuracy.

Optimisation of the loss function

SlideLoss is an improved loss function[30], which focuses on both the prediction results and the scale information of the target object, using the sliding window mechanism to segment the image into multiple chunks and compute the loss function separately, which can more accurately obtain the classification prediction results of the target object at different scales and locations. The underlying concept is shown in Fig. 6. To reduce hyperparameters, we use the average of the IoU values of all bounding boxes as a threshold μ, with anything less than μ as a negative sample and anything greater than μ as a positive sample[31]. We hope that the model will learn to optimise these samples and make fuller use of them to train the network. Assuming the predicted bounding box is denoted as Bp and the actual detection box as Br, the formula for IoU[32,33].

$$IoU = \frac{B_P \cap B_r}{B_P \cup B_r}. \tag{14}$$

$$f(x) = \begin{cases} 1 & x \le \mu - 0.1 \\ e^{1-\mu} & \mu < x < \mu - 0.1 \\ e^{1-x} & x \ge \mu \end{cases} \tag{15}$$

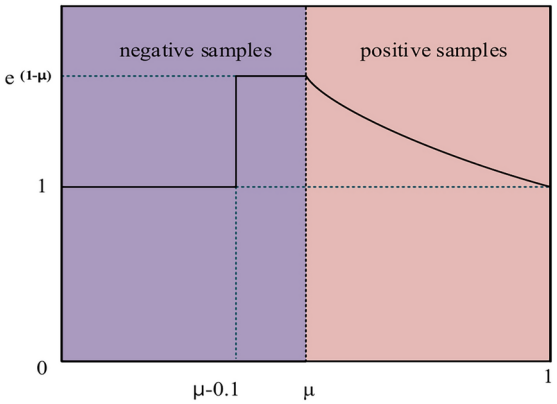

Fig. 6. SlideLoss loss function.

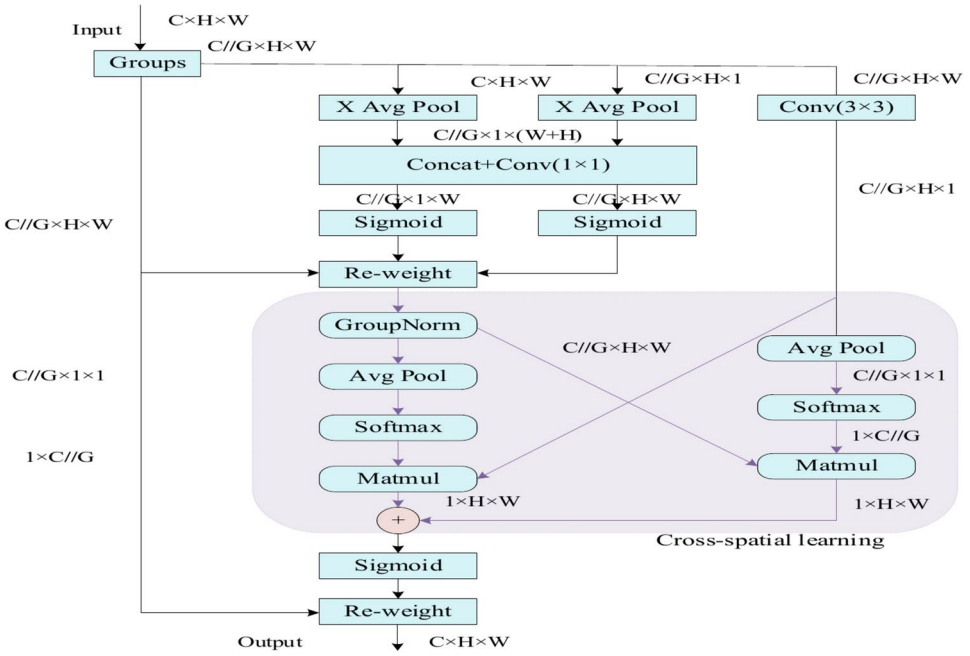

Fig. 7. EMA module structure diagram.

Due to the complexity of the site construction environment, the concept of Exponential Moving Average calculation (EMA)[34] is added to SlideLoss due to the poor detection of small targets, such as safety helmets and reflective clothing, when the distance is too far. The model's understanding of details and large-scale structures in the image is enhanced by integrating information from different scales. The specific structure of EMA is shown in the Fig. 7. The CA module is used as the design concept in EMA.

The parallel sub-network structure of the EMA module reduces greater depth and more complex processing sequences.EMASlideLoss is designed according to the scale size of the target object in the image. EMASlideLoss is intended to solve the problem of imbalance between small and large targets in traditional target detection algorithms to improve the accuracy of small target detection.EMASlideLoss EMA in EMASlideLoss refers to the Exponential Moving Average, which smoothes out the trend of loss values[35]. In EMASlideLoss, the Exponential Moving Average is used to adjust the weights of the loss function to make it more concerned about the detection of small targets, thus enhancing the detection of small targets by the safety wearable detection model.

Analysis of model improvement performance

Benchmark model comparison experiment

Safety helmet and reflective clothing wear detection belong to the scope of target detection; target detection can be divided into two-stage detection and single-stage detection. The training environment and platform remained unchanged. Faster-RCNN, YOLOv3, YOLOv5, YOLOv7, and YOLOv8 algorithms were used for training on the same data set. Test the detection effect of different models. Test results are shown in Table 2 and Fig. 8.

Models	mAP50/%	P/%	R/%	Helmet mAP	Ref-clothing mAP	Parameter size
Faster-RCNN	74.5	83.2	65.3	85.6	79.8	25,279,070
YOLOv3	82.7	87.2	72.7	87.8	83.5	25,983,713
YOLOv5	85.7	87.3	85.8	89.8	86.1	2,764,577
YOLOv7	83.2	84.5	86.2	91.3	85.5	7,402,946
YOLOv8	86.3	88.6	87.7	93.5	87.6	3,006,428
Ours	92.5	92.0	91.1	97.5	92.4	1,400,245

Table 2. Comparative experiments with different models.

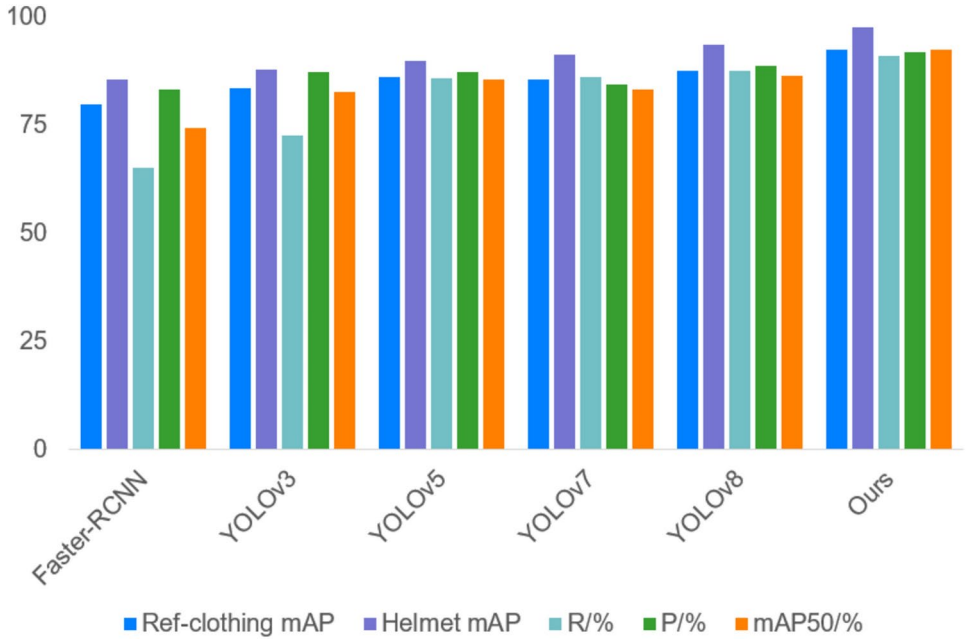

Fig. 8. Comparative experiments with different models.

By comparing with other classical models, the improved model has the best mAP value, accuracy and recall. Secondly, the model parameters as usual decreased by 53.4% compared with the original model of YOLOv8. Through the above study, it is obvious that the method proposed in this paper has obvious advantages compared with the existing typical network models.The results in the table show that the YOLO-P2-Ghost-Dyhead model in this paper achieves the highest average detection accuracy. Compared with the original YOLOv8 model, the mapped value of this paper's model is improved by 6.2%, and the total model parameters are reduced by 53.4%; this indicates that the model in this paper exhibits superior resource utilization, higher real-time performance, and therefore is more straightforward to deploy, ultimately saving human resources in the construction site. Moreover, to visualise the performance benefits of the method proposed in this paper, several images were tested, and the results are presented in Fig. 9. These images show the actual detection results of different models on the safety wear dataset.

Additionally, relying on a single dataset is insufficient to assess the model's generalization capability. Therefore, we extended our evaluation to include another publicly available image dataset VOC2028, the dataset consists of 3248 images, each with a resolution of 640 × 640 pixels. The tag types in this dataset are consistent with those in the original dataset, including four different wear types. Figure 10 provides a detailed comparison of the detection results of the proposed model and the various algorithms discussed in this study. It is clear that the improved model can adapt well to different datasets, thus validating the superiority of the proposed model.

Loss function comparison experiment

In contrast to SlideLoss, EMASlideLoss incorporates the concept of exponential moving average. EMASlideLoss applies an exponential moving average computation to the loss values of SlideLoss during training, thereby achieving smoother loss values. This approach is used to enhance the generalization and stability of the safety wear detection model, reduce noise and fluctuations, and consequently improve the accuracy and convergence speed of the training process. As a result, EMASlideLoss is employed as the loss function in this study. To validate the efficacy of the proposed enhanced loss function, comparative experiments with SlideLoss, FocalLoss, VarifocalLoss, QualityFocalLoss, and EMASlideLoss were conducted. The performance was evaluated using mAP50, precision (P), recall (R), and mean average precision (mAP) for two types of targets as the reference

Fig. 9. Comparison chart of test results.

metrics. The experimental outcomes are presented in Table 3 and Fig. 11. The findings demonstrate that the EMASlideLoss loss function utilized in this research exhibits the highest accuracy among the tested methods..

Ablation experiment

To better assess the validity and feasibility of the proposed safety wear detection model, ablation experiments were conducted using the dataset created in this study. The results are shown in Table 4. It is clear that each improvement contributes to the detection performance, with some of them focusing on accuracy and others on speed. To comprehensively evaluate the proposed model's validity and feasibility, the performance of each component was rigorously assessed through ablation experiments. Given that the proposed model in this

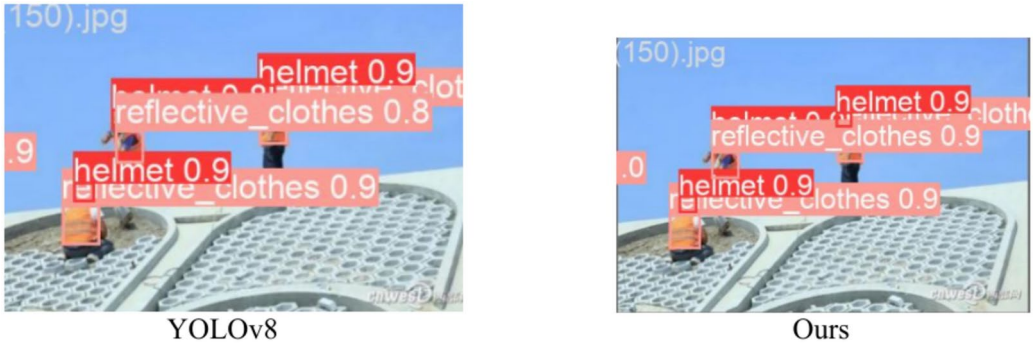

Fig. 10. New data test results.

Loss function	mAP50/%	P/%	R/%	Helmet mAP	Ref-clothing mAP
SlideLoss	90.5	91.2	86.2	97.3	**93.6**
FocalLoss	87.3	87.0	81.7	96.4	89.7
QualityfocalLoss	89.3	87.2	83.1	97.0	92.7
VarifocalLoss	88.5	86.0	81.8	96.0	91.7
Ours	91.4	91.2	86.3	97.5	92.4

Table 3. Comparative experiments with different loss functions.

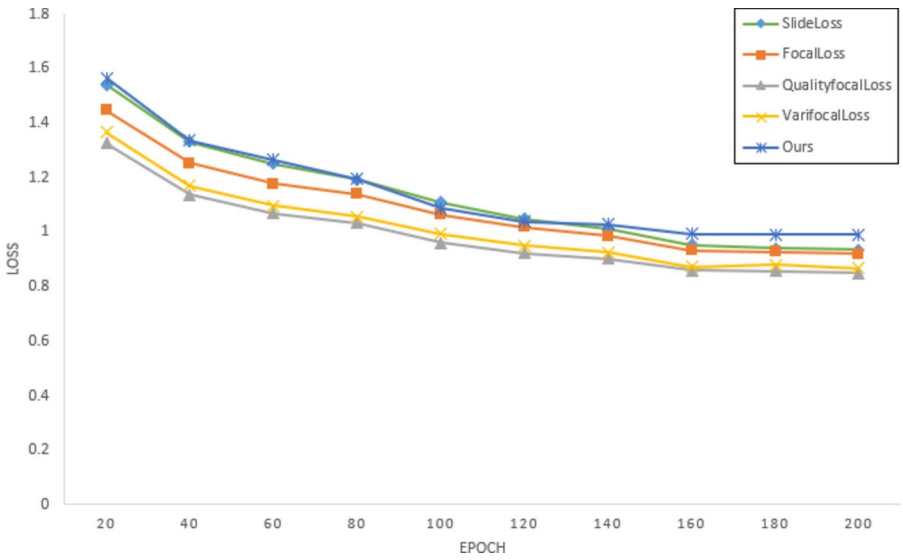

Fig. 11. The loss function compares the experimental results.

study is based on YOLOv8s, the YOLOv8s model was utilized as the benchmark for conducting the ablation experiments. The experiments utilized precision (P), recall (R), mean Average Precision (mAP), frames per second (FPS), and the total number of model parameters (GFLOPS) as evaluation metrics.

The following conclusions were derived from the ablation experiments:

1. The baseline model had the lowest detection accuracy due to the presence of shooting distances, angles and other small target detection issues in the dataset.
2. Upgrading the backbone to GhostNet resulted in a significant increase in the detection speed of the model. However, there is a slight increase in computational load due to the inclusion of the P2 detection layer in its architecture. Models B, D and E show varying degrees of improvement in detection accuracy and speed. The integration of the sub-modules shows varying degrees of improvement in detection accuracy and speed.
3. The F, G and H models that incorporate the P2 detection layer into the original architecture show higher accuracy.

Models	P2	GhostNet	Dyhead	EMASlideLoss	mAP50/%	P/%	R/%	fps/s	GFLOPS
A					86.3	88.6	87.7	62.1	8.1
B	✓				91.5	90.6	88.5	58.5	7.2
C		✓			90.1	90.6	86.9	88.7	6.7
D			✓		89.4	90.5	88.9	70.4	7.3
E				✓	91.4	91.2	86.3	70.8	7.9
F	✓	✓			91.9	93.0	90.2	77.4	4.6
G	✓		✓		90.5	89.5	89.3	60.7	6.5
H	✓			✓	90.1	93.5	88.8	64.5	7.2
I	✓	✓	✓	✓	**92.5**	**92.0**	**91.1**	**96.8**	**4.2**

Table 4. YOLO-P2-Ghost-Dyhead model ablation experiment results.

4. After integrating all the improved modules, Model I enriches the feature information capturing and fusion capability by effectively suppressing irrelevant information. By effectively suppressing irrelevant information, the feature extraction and fusion capabilities are enhanced. As a result, the overall performance of the model exceeds that of the benchmark model, especially in detecting smaller targets.

With these improvements, the model achieves significant improvements in both accuracy and speed, providing a more reliable and efficient solution for the application of safe wear detection algorithms.

Conclusion

The intricate nature of construction backgrounds and the diverse environments present at construction sites contribute to a low accuracy in detecting and recognizing workers' safety helmets and reflective clothing. This paper introduces a novel model, YOLO-P2-Ghost-Dyhead, which is an enhanced algorithm based on YOLOv8 specifically designed for the detection of safety wear at construction sites. The following conclusions are drawn from experiments conducted using an augmented dataset, along with comparative analyses of experimental results.

1. Given that safety helmets occupy fewer pixels within images, they are often overlooked and challenging to detect. To address this issue concerning small targets, we first validate our constructed dataset by incorporating a specialized P2 small target detection layer. Experimental outcomes indicate that the detection performance for safety helmets has improved by 5.2%, 8.3%, 5.8%, and 8.8% compared to baseline models including YOLOv8, YOLOv7, YOLOv5, and Fast-RCNN respectively—an overall enhancement of 17%. This clearly demonstrates that integrating the P2 detection layer significantly enhances the accuracy of helmet detection.

2. Secondly, by introducing the GhostNet lightweight network to replace it with the backbone network of Yolov8, the experimental results show that the mAP of the safety wear is improved by 3.8%, 6.9%, 4.4%, 7.4%, and 15.6% over the baseline models, YOLOv8, YOLOv7, YOLOv5, and Fast-RCNN. The number of parameters is decreased. At the same time, the FPS is improved, proving that the model's number of parameters and computational complexity are reduced. In contrast, the accuracy is improved, and also enables the model to better and faster explain the redundant information, reduces the operation cost and can be extremely meaningful to be applied to real-time detection of safety wear at construction sites, which significantly proves the superiority of the model algorithm proposed in this paper.

3. As reflective clothing can be challenging to identify against various backgrounds, it is prone to misjudgment and detection failures. The introduction of the Dyhead detection head in the Head layer, leveraging its structural characteristics to unify all types of attention within the consent architecture, significantly enhances the detection accuracy for diverse targets across a range of complex environments. Compared to benchmark models such as YOLOv8, YOLOv7, YOLOv5, and Fast-RCNN, recall rates improved by 0.2%, 2.7%, 3.1%, 16.7%, and 23.6% respectively, thereby markedly enhancing the model's generalization capability. Furthermore, by incorporating the EMASlideLoss dynamic loss function and conducting comparative experimental analyses with other loss functions, results indicate that this approach yields superior training curve convergence while simultaneously improving accuracy. This advancement contributes positively to the stability of the model.

4. Finally, the improved YOLOv8-P2-GhostNet-Dyhead-based model proposed in this paper achieves a 6.2% increase in mAP and reaches a processing speed of 96.8 fps, which is faster than the benchmark YOLOv8 model on the dataset used in this study. The analysis of the ablation experiments reveals that the mAP of the improved model is enhanced by 5.2%, 3.8%, and 5.1% when compared with the individual modules, respectively. Additionally, the results indicate that the algorithm proposed in this paper is effective and demonstrates significant performance improvements.

Although the algorithmic model presented in this paper has significantly enhanced the accuracy and speed of site safety wear detection, obtaining diverse site datasets is challenging due to the dynamic nature of construction site environments and safety concerns. Consequently, future work will focus on further refining the model to enhance its ability to detect the safety wear of site personnel across various scenarios, utilizing dynamically acquired real-time video data.

Data availability

The dataset in this paper is made up of self-picked datasets as well as SHWD's public datasets, and there is no question of any dispute of interest. The link to the open source dataset is GitHub—njvisionpower/Safety-Helmet-Wearing-Dataset: Safety helmet wearing detect dataset.

Received: 10 July 2024; Accepted: 13 December 2024
Published online: 07 January 2025

References

1. Akinlolu, M., Haupt, T. C., Edwards, D. J. & Simpeh, F. A bibliometric review of the status and emerging research trends in construction safety management technologies. *Int. J. Constr. Manag.* **22**, 2699–2711 (2022).
2. Zeng, L. & Li, R. Y. M. Construction safety and health hazard awareness in Web of Science and Weibo between 1991 and 2021. *Saf. Sci.* **152**, 105790 (2022).
3. Sanni-Anibire, M. O., Mahmoud, A. S., Hassanain, M. A. & Salami, B. A. A risk assessment approach for enhancing construction safety performance. *Saf. Sci.* **121**, 15–29 (2020).
4. Wang, H. et al. A real-time safety helmet wearing detection approach based on CSYOLOv3. *Appl. Sci.* **10**, 6732 (2020).
5. Bottlang, M., DiGiacomo, G., Tsai, S. & Madey, S. Effect of helmet design on impact performance of industrial safety helmets. *Heliyon* 8 (2022).
6. Park, M.-W., Elsafty, N. & Zhu, Z. Hardhat-wearing detection for enhancing on-site safety of construction workers. *J. Constr. Eng. Manag.* **141**, 04015024 (2015).
7. Kim, S. C., Ro, Y. S., Shin, S. D. & Kim, J. Y. Preventive effects of safety helmets on traumatic brain injury after work-related falls. *Int. J. Environ. Res. Public Health* **13**, 1063 (2016).
8. Viola, P. & Jones, M. J. Robust real-time face detection. *Int. J. Comput. Vis.* **57**, 137–154 (2004).
9. Hale, A. R., Heming, B., Carthey, J. & Kirwan, B. Modelling of safety management systems. *Saf. Sci.* **26**, 121–140 (1997).
10. Taye, M. M. Understanding of machine learning with deep learning: Architectures, workflow, applications and future directions. *Computers* **12**, 91 (2023).
11. Xu, M., Yoon, S., Fuentes, A. & Park, D. S. A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recogn.* **137**, 109347 (2023).
12. Monga, V., Li, Y. & Eldar, Y. C. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Process. Mag.* **38**, 18–44 (2021).
13. Pan, Y. & Zhang, L. Integrating BIM and AI for smart construction management: Current status and future directions. *Arch. Comput. Methods Eng.* **30**, 1081–1110 (2023).
14. Shine, L. & CV, J. Automated detection of helmet on motorcyclists from traffic surveillance videos: A comparative analysis using hand-crafted features and CNN. *Multimed. Tools Appl.* **79**, 14179–14199 (2020).
15. Girshick, R. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* 1440–1448 (2015).
16. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137–1149 (2016).
17. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* 779–788 (2016).
18. Liu, W. et al. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands*, October 11–14, 2016, Proceedings, Part I 14 21–37 (2016).
19. Shi, H., Chen, X. & Yang, Y. Safety helmet wearing detection method of improved YOLOv3. *Comput. Eng. Appl.* **55**, 213–220 (2019).
20. Wu, F., Jin, G., Gao, M., Zhiwei, H. & Yang, Y. Helmet detection based on improved YOLO V3 deep model. In *2019 IEEE 16th International conference on networking, sensing and control (ICNSC)* 363–368 (2019).
21. Song, H. Multi-scale safety helmet detection based on RSSE-YOLOv3. *Sensors* **22**, 6061 (2022).
22. Qian, S. & Yang, M. Detection of Safety Helmet-Wearing Based on the YOLO_CA Model. *Comput. Mater. Contin.* 77 (2023).
23. Wang, L. et al. Investigation into recognition algorithm of helmet violation based on YOLOv5-CBAM-DCN. *IEEE Access* **10**, 60622–60632 (2022).
24. Yung, N.D.T., Wong, W., Juwono, F.H. & Sim, Z.A. Safety helmet detection using deep learning: Implementation and comparative study using YOLOv5, YOLOv6, and YOLOv7. In *2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)* 164–170 (2022).
25. Liu, Y. et al. Helmet wearing detection algorithm based on improved YOLOv5. *Sci. Rep.* **14**, 8768 (2024).
26. Wang, Q. et al. A deep learning approach incorporating YOLO v5 and attention mechanisms for field real-time detection of the invasive weed Solanum rostratum Dunal seedlings. *Comput. Electron. Agric.* **199**, 107194 (2022).
27. Chen, J., Mai, H., Luo, L., Chen, X. & Wu, K. Effective feature fusion network in BIFPN for small object detection. *In 2021 IEEE international conference on image processing (ICIP)* 699–703 (2021).
28. Han, K. et al. Ghostnet: More features from cheap operations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* 1580–1589 (2020).
29. Dai, X. et al. Dynamic head: Unifying object detection heads with attentions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* 7373–7382 (2021).
30. Cai, S., Meng, H. & Wu, J. FE-YOLO: YOLO ship detection algorithm based on feature fusion and feature enhancement. *J. Real-Time Image Process.* **21**, 61 (2024).
31. Yu, Z. et al. Yolo-facev2: A scale and occlusion aware face detector. *arXiv preprint* arXiv:2208.02019 (2022).
32. Yang, D. et al. A streamlined approach for intelligent ship object detection using EL-YOLO algorithm. *Sci. Rep.* **14**, 15254 (2024).
33. Li, W., Solihin, M. I. & Nugroho, H. A. RCA: YOLOv8-based surface defects detection on the inner wall of cylindrical high-precision parts. *Arab. J. Sci. Eng.* 1–19 (2024).
34. Ouyang, D. et al. Efficient multi-scale attention module with cross-spatial learning. In *ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 1–5 (2023).
35. Huang, H. & Zhu, K. Automotive parts defect detection based on YOLOv7. *Electronics* **13**, 1817 (2024).

Acknowledgements

This research was funded by the National Nature Science Foundation of China (NSFC), A study of ground-space semi-physical modeling methods for predicting the water content of fine dead combustibles on the forest floor under a Grant (No.32371864), 2024 Ningxia Hui Autonomous Region Key R&D Programme Project (2024FRD05109) and Science and Technology Project of Heilongjiang Provincial Department of Transport (HJK2023B019).

Author contributions

Jian Xing—writing original draft preparation, Chenglong Zhan—conducted the experiments, JiaQiang Ma—collected data, Zibo Chao and Ying Liu—prepared figures and tables. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-83391-7.

Correspondence and requests for materials should be addressed to C.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.