



ARTICLE

An Improved Lightweight Safety Helmet Detection Algorithm for YOLOv8

Lieping Zhang^{1,2}, Hao Ma¹, Jiancheng Huang³, Cui Zhang^{4,*} and Xiaolin Gao²

¹Key Laboratory of Advanced Manufacturing and Automation Technology, Education Department of Guangxi Zhuang Autonomous Region, Guilin University of Technology, Guilin, 541004, China

²Guangxi Key Laboratory of Special Engineering Equipment and Control, Guilin University of Aerospace Technology, Guilin, 541004, China

³Guangxi Tianli Construction Engineering Co., Ltd., Guilin, 541001, China

⁴School of Information Engineering, Nanning College of Technology, Guilin, 541004, China

*Corresponding Author: Cui Zhang. Email: zhangcui@nnct.edu.cn

Received: 26 November 2024; Accepted: 23 January 2025; Published: 16 April 2025

ABSTRACT: Detecting individuals wearing safety helmets in complex environments faces several challenges. These factors include limited detection accuracy and frequent missed or false detections. Additionally, existing algorithms often have excessive parameter counts, complex network structures, and high computational demands. These challenges make it difficult to deploy such models efficiently on resource-constrained devices like embedded systems. Aiming at this problem, this research proposes an optimized and lightweight solution called FGP-YOLOv8, an improved version of YOLOv8n. The YOLOv8 backbone network is replaced with the FasterNet model to reduce parameters and computational demands while local convolution layers are added. This modification minimizes computational costs with only a minor impact on accuracy. A new GSTA (GSConv-Triplet Attention) module is introduced to enhance feature fusion and reduce computational complexity. This is achieved using attention weights generated from dimensional interactions within the feature map. Additionally, the ParNet-C2f module replaces the original C2f (CSP Bottleneck with 2 Convolutions) module, improving feature extraction for safety helmets of various shapes and sizes. The CIoU (Complete-IoU) is replaced with the WIoU (Wise-IoU) to boost performance further, enhancing detection accuracy and generalization capabilities. Experimental results validate the improvements. The proposed model reduces the parameter count by 19.9% and the computational load by 18.5%. At the same time, mAP (mean average precision) increases by 2.3%, and precision improves by 1.2%. These results demonstrate the model's robust performance in detecting safety helmets across diverse environments.

KEYWORDS: YOLO; safety helmet detection; complex environments; lightweight; WIoU loss function

1 Introduction

Construction environments in sectors such as building sites, electrical operations, and resource exploration are inherently complex and present numerous safety hazards. Helmets, as essential protective gear, significantly reduce the risk of injury and are crucial for worker safety [1]. Consequently, monitoring helmet usage is a critical safety measure in these settings. Traditionally, helmet wear detection has relied on manual inspections, which are costly in terms of both human and time resources and are characterized by low efficiency and poor timeliness. As a result, object detection-based helmet wear detection has emerged as a critical method for enhancing safety management on construction sites [2].



In recent years, the ongoing advancement of object detection algorithms driven by deep learning has divided them into two groups, distinguished by the number of detection stages involved: object detection algorithms categorized as one-stage and two-stage [3]. Algorithms with two stages, like Fast R-CNN [4] and Faster R-CNN [5], start by generating candidate regions, selecting the positions of objects within these regions, and then further classifying and locating the objects to facilitate helmet wear detection. The Tiny Face model was improved by combining it with a CNN and adding supplementary modules [6]. This approach overcame the low accuracy issues commonly associated with traditional single deep learning models in detection tasks. Espinosa-Oviedo et al. [7] enhanced Faster R-CNN using the EspNet model. Li et al. [8] introduced the Faster R-CNN-LSTM approach, an enhancement of the Faster R-CNN framework, which improved mAP performance on the dataset. Sun et al. [9] developed a safety monitoring method for factories utilizing the Faster R-CNN algorithm, which builds a feature pyramid to integrate multi-scale features from the original network, improving the detection of small objects such as helmets and safety belts. Zhang et al. [10] introduced a helmet detection method for steel mill workshops using deep learning's Faster R-CNN, enhancing detection accuracy and demonstrating good practicality and effectiveness.

Although these methods have enhanced model accuracy, their complex structures lead to high computational costs and substantial parameter sizes. Moreover, the detection scenarios, predominantly confined to single work environments, often exhibit poor real-time performance and do not adequately meet practical application requirements [11].

Safety helmet detection requires high real-time responsiveness. Single-stage object detection algorithms eliminate the need for a candidate region generation phase and accomplish object classification and location in a single step with only one detection iteration [12,13], thus offering faster detection speeds suitable for helmet wear detection. Notable examples of such algorithms include SSD and the YOLO series. Ku et al. [14] presented an enhanced helmet detection algorithm built on YOLOv4, effectively decreasing model complexity while boosting detection accuracy. Zhang et al. [15] proposed an SCM-YOLO, incorporating several optimization strategies into YOLOv4-tiny. These enhancements include integrating an SPP structure, a CBAM module, the Mish activation function, and K-Means++ clustering. The experimental results confirm the algorithm meets real-time processing and accuracy standards for safety helmet detection in challenging scenarios. Sadiq et al. [16] introduced a deep learning-driven safety helmet detection model that combines YOLOv5 with a fuzzy image enhancement module, leading to the development of the enhanced FD-YOLOv5 model. The FD-YOLOv5 outperforms YOLOv5 in complex surveillance videos, efficiently detecting helmets while demonstrating strong practicality and robustness in real-world construction environments. Shan et al. [17] introduced an enhanced YOLOv5 designed to improve detection accuracy for small, densely packed objects by incorporating weighted bidirectional feature pyramids. Yao et al. [18] proposed the AMCFY-YOLOv5s model, incorporating optimization strategies such as k-means++, BiFPN, and CBAM. This model improves the accuracy of miner status detection within cages and demonstrates strong robustness in complex environments, offering an efficient solution for coal mine safety monitoring. An et al. [19] presented an enhanced YOLOv5s model by optimizing anchor boxes, integrating attention mechanisms, adopting the SIoU, and integrating knowledge distillation techniques. Experimental findings show that the model delivers enhanced accuracy and better real-time performance. It performs exceptionally well in low-light and multi-distance conditions, and the model is highly effective for helmet detection. Song et al. [20] created an intelligent helmet detection system that combines the YOLOv5 detector with the DeepSort multi-object tracking algorithm. The system employs YOLOv5 to extract target bounding boxes and combines Kalman filtering with the Hungarian algorithm for trajectory prediction and tracking, enhancing detection accuracy in complex construction environments. Wang et al. [21] introduced a helmet recognition method built on a refined YOLOv7-tiny model, optimizing feature extraction efficiency and improving detection accuracy.

Notably, YOLOv8 incorporates a redesigned architecture and an optimized loss function, resulting in significantly faster training speeds and higher detection accuracy. These enhancements make YOLOv8 more efficient and effective for object detection tasks [22]. Lin et al. [23] the YOLOv8-SLIM-CA helmet recognition method was introduced, incorporating a coordinate attention mechanism, a streamlined neck design, and an additional layer for detecting small objects. These enhancements improve detection capabilities for small objects and complex backgrounds. The algorithm achieves real-time, efficient helmet detection performance, surpassing mainstream detection algorithms. Aboah et al. [24] introduced a helmet recognition model leveraging YOLOv8. This model utilizes a few-shot data sampling approach to reduce labeling efforts, enabling efficient video detection optimized for real-world real-time applications.

Although single-stage object detection algorithms have improved detection accuracy and speed, they face challenges, such as increased model parameters and computational costs when enhancing detection precision [25]. In practical application environments, using complex models can improve detection accuracy. However, due to factors such as network complexity, a large number of parameters, substantial physical memory requirements, and lengthy training times, these models are complex to deploy on embedded or edge devices, increasing deployment costs. Additionally, the traditional convolution process in complex models raises computational costs, wastes specific resources, and weakens the model's real-time performance. In actual detection environments, complexities such as obstructions, lighting variations, and dense worker populations often result in missed and false detections, leading to suboptimal outcomes. Additionally, the considerable variation in helmet sizes and colors complicates accurate feature extraction by these algorithms. To overcome these challenges, this study introduces an optimized version of the YOLOv8n algorithm, FGP-YOLOv8, offering a lightweight solution for helmet wear detection.

This paper's first section identifies helmet detection's challenges in complex environments and reviews recent developments in neural network-based object detection techniques. The second section describes the architecture of the proposed FGP-YOLOv8 algorithm, followed by the third section, which focuses on its crucial improvement modules. The fourth section explains the experimental setup, including datasets and parameters, and presents detailed experiments to assess the effectiveness of the proposed improvements module. Finally, the fifth section concludes the study by highlighting its contributions, exploring practical applications of the model, and offering recommendations for future research.

The main contributions of this paper can be summarized as follows:

1. The FasterNet lightweight module, utilizing PConv as the primary operator, replaces the original backbone feature extraction network. This substantially decreases the model's parameters and computational burden while maintaining detection accuracy at a comparable level, thus achieving efficient network lightweight.
2. The newly developed GSTA module reduces computational complexity while improving spatial and channel attention fusion. This improvement boosts the model's target localization and feature expression capabilities, enhancing helmet wear detection performance.
3. A new ParNet-C2f module, based on structural reparameterization, employs a parallel network structure to broaden the feature map's perceptual range. This improvement enhances the model's ability to extract features and increases detection accuracy.
4. Implementing WIoU as the loss function balances penalties between high- and low-quality anchor boxes. It includes a dynamic non-monotonic focus mechanism that preserves feature information across various scales, accelerating convergence and boosting the model's detection capabilities.

2 The Proposed FGP-YOLOv8 Algorithm

2.1 YOLOv8 Algorithm

In 2023, Ultralytics introduced YOLOv8, a new model for object detection, instance segmentation, and image classification tasks [26]. Compared to the widely used YOLOv5, YOLOv8 has achieved broader adoption in object detection due to its superior detection accuracy and faster processing speeds [27]. YOLOv8 is categorized into five models based on model depth: n, s, l, m, x [28]. YOLOv8n features the smallest parameter size and the quickest detection speed, [29] making it the chosen baseline model for enhancements and structural redesigns in this study. The YOLOv8n model comprises three primary components: backbone, neck, and head networks [30]. Although YOLOv8n excels in detection speed, it exhibits areas that require further refinement. The simultaneous detection of helmets in scenes with multiple individuals poses challenges, particularly concerning the algorithm's computational demands and parameter size, which render it less practical for deployment on construction sites [31]. In real-world applications, complex scenarios such as dense crowds, numerous small objects, and intricate backgrounds complicate feature extraction necessary for accurate helmet detection, often leading to false alarms and omissions [32].

2.2 Related Work

To resolve the challenges of limited accuracy, as well as frequent missed and false detections faced by the YOLOv8n algorithm in helmet wear detection, and considering the necessity for model lightening, this paper develops the FGP-YOLOv8 algorithm based on the YOLOv8n model. We improved and optimized recent mainstream safety helmet detection models and, through comparative experiments, ultimately selected four modules to optimize the model. This algorithm is enhanced in four major areas to achieve superior detection performance and reduce computational costs.

Firstly, to address the excessive parameters and computational demands, this study substitutes YOLOv8n's backbone network with a FasterNet model and incorporates Partial Convolution (PConv) into the backbone layers, substantially decreasing the model's parameters and floating-point computations, thereby making it more lightweight. Secondly, the newly developed GSTA module is incorporated into the neck network, boosting the model's target localization and feature representation abilities by leveraging three-dimensional attention weights in the feature maps, all while keeping the computational load unchanged. Thirdly, a novel ParNet-C2f module is designed to extract features from safety helmets of various colors and sizes more effectively. Lastly, the WIoU loss function is applied, utilizing a dynamic focusing mechanism to improve the precision of quality assessments between predicted and actual boxes, thereby accelerating model convergence and enhancing the network's detection efficiency. These enhancements improve performance, reduce computational redundancy, and increase the algorithm's efficiency. Fig. 1 presents a structural diagram illustrating the model's architecture.

3 FGP-YOLOv8 Detection Module

3.1 FasterNet Lightweight Module

In helmet-wearing detection tasks, the images often feature various colors, significant obstructions, and complicated settings. YOLOv8 employs the DarkNet-53 module as its backbone, which provides a deep network structure aiding in enhanced feature extraction but also accumulates numerous convolution operations, leading to feature redundancy. This results in a high parameter count and reduced detection speed, making it less ideal for deployment on portable devices. Introduced in 2023, FasterNet features a simpler architecture that achieves faster processing and greater accuracy across different devices. The FasterNet lightweight network is organized into four hierarchical stages [33], each stage beginning with either

an embedding layer or a merging layer to facilitate spatial downsampling and channel expansion. Each stage incorporates a FasterNet Block, utilizing a 3×3 PConv as its primary operator, followed by two 1×1 Conv layers. Batch Normalization is strategically placed between the Conv layers, complemented by the ReLU to optimize inference speed and preserve feature diversity. The complete structure of FasterNet is depicted in Fig. 2.

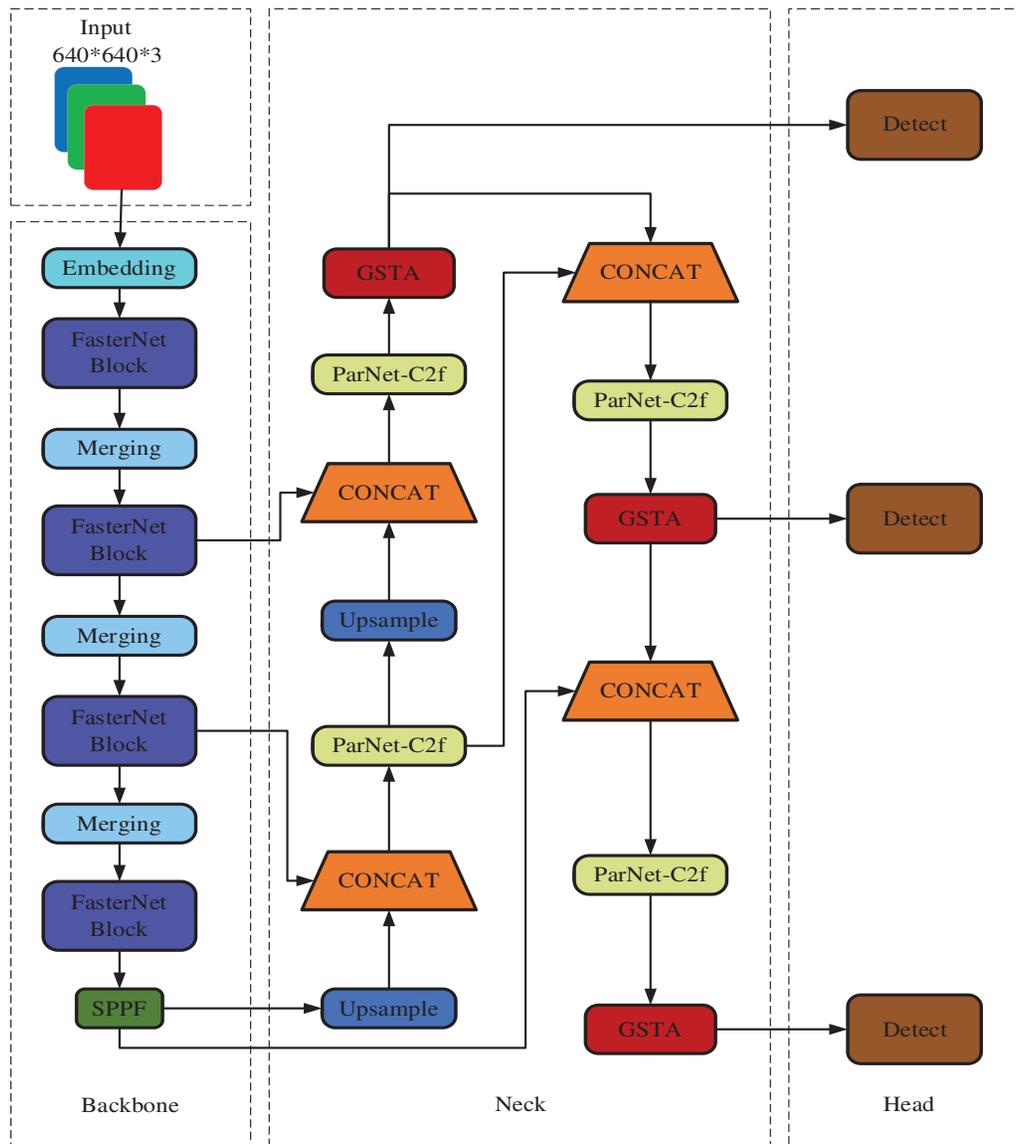


Figure 1: The network structure of FGP-YOLOv8

The FasterNet network employs Partial Convolution (PConv) as its primary operator, improving algorithm accuracy while reducing FLOPs and memory usage. Fig. 3 illustrates the operational principle of PConv.

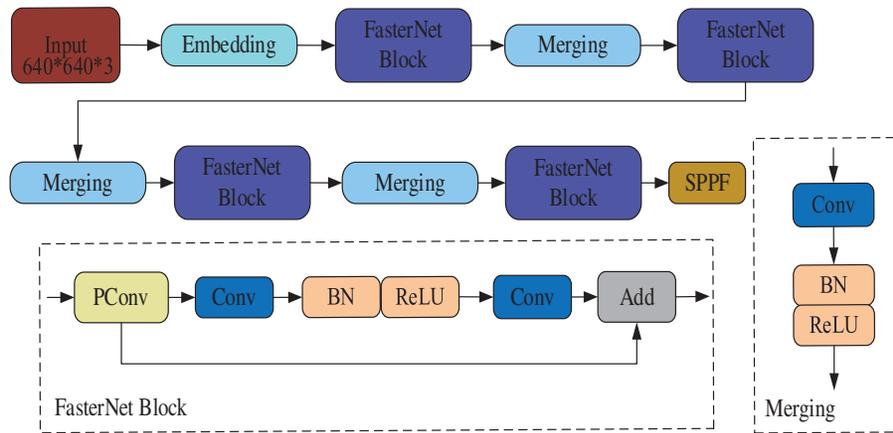


Figure 2: FasterNet network architecture diagram

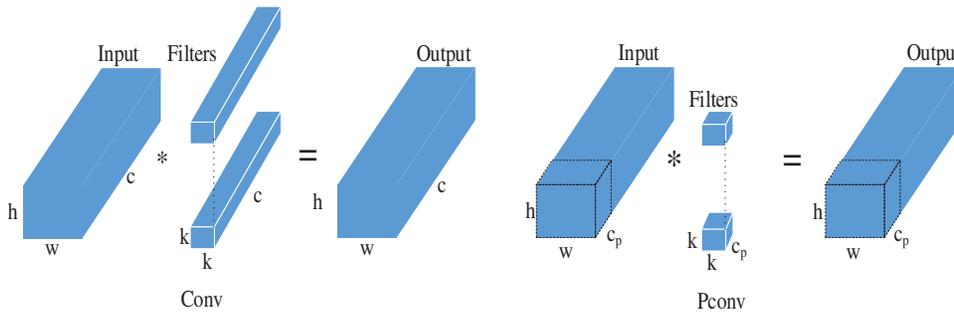


Figure 3: The operational principle of Conv and PConv

In Partial Convolution (PConv), the conventional convolution operation is selectively applied to particular input channels for spatial feature extraction, leaving the remaining channels unchanged. This approach facilitates sequential memory access by focusing on processing the feature map’s initial or final continuous channels. This technique efficiently captures spatial features while reducing computational and memory demands. The FLOPs involved in PConv are detailed in Eq. (1).

$$h \times w \times k^2 \times c_p^2 \tag{1}$$

For regular convolution (Conv), the FLOPs related to the inputs $I \in R^{c \times h \times w}$ and outputs $O \in R^{c \times h \times w}$ are outlined in Eq. (2).

$$h \times w \times k^2 \times c^2 \tag{2}$$

In the formula, w represents the width and h denotes the height of the input channels, while c and c_p represents consecutive network channel counts, and k is also used to signify the convolution kernel size. When operating under the classic ratio of $r = c_p/c = 1/4$, the FLOPs for PConv amount to just 1/16 of those for a standard Conv, dramatically lowering redundant calculations and memory usage and thus boosting computational efficiency without compromising model performance.

Integrating the FasterNet lightweight network has optimized the YOLOv8 backbone, markedly decreasing the algorithm’s computational demands and parameter count while improving detection speeds in diverse and complex environments.

3.2 GSTA Module

Attention mechanisms have demonstrated significant effectiveness in artificial intelligence and are extensively employed in object detection tasks. Traditional attention mechanisms, which compute weights for individual channels, often overlook information integration across channel and spatial dimensions. The SE attention mechanism excels at adaptively learning the weights of each channel, thereby focusing on the critical channel information [34]. The CBAM adds spatial attention as a complement to channel attention, reference [35] yet it treats them as separate entities without considering their interaction. Distinctively, the Triplet Attention module utilizes three parallel branches to focus on different input dimensions, reference [36] effectively capturing spatial-channel interactions. This method fosters the development of interdependencies between channel and spatial locations, enhancing feature interpretation capabilities, albeit with increased computational and parameter demands.

This study combines GSConv with the traditional Triplet Attention module to mitigate the excessive increases in computational and parameter demands introduced by the Triplet Attention module. It also presents the newly developed GSTA module, as depicted in Fig. 4.

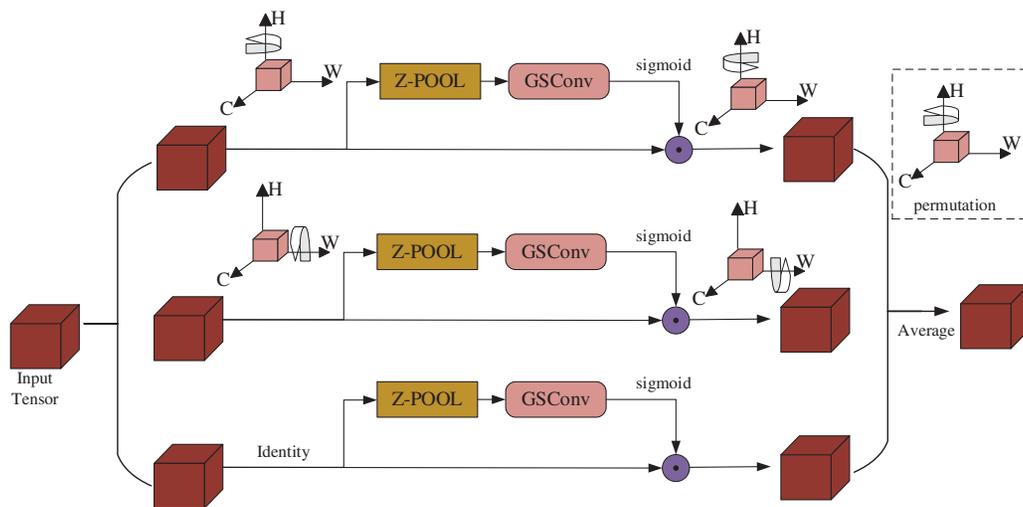


Figure 4: GSTA structure diagram

Recently introduced, GSConv represents a novel, lightweight convolution that combines standard convolution (SC), depth-wise separable convolution (DSCConv), and shuffle mixed convolution [37]. Its implementation is illustrated in Fig. 5. GSConv operates at 60% to 70% of the computational cost of standard convolution while maintaining excellent accuracy.

The GSTA module leverages three parallel branches to derive dependencies across the dimensions of the input tensor (C, H) , (C, W) , (H, W) , making spatial and channel attention processes interdependent so that they are no longer independent. The module uses rotational operations and residual transformations to set up dependencies across dimensions. First, the channel dimension C interacts with the spatial dimension H , and then the spatial dimension W interacts with the channel dimension C . Using rotational operations,

permutation can link channels with any spatial dimension. A third branch gathers traditional spatial attention weights.

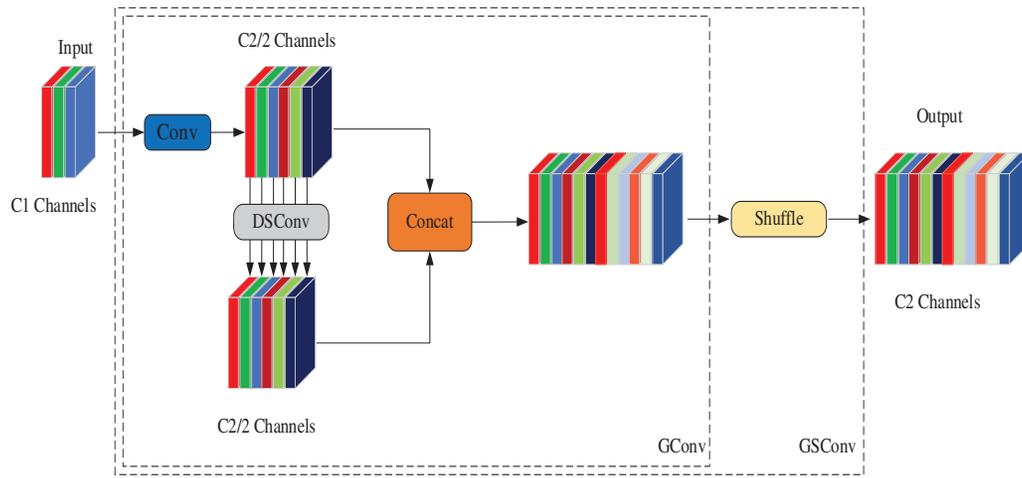


Figure 5: GSConv structure diagram

In the Z-pool layer, average and maximum pooling is applied to the zero dimensions of the input, which are subsequently concatenated. This concatenation reduces the number of channels in this dimension to two, preserving the tensor's rich representation while simplifying the computational process. The operation is detailed in the formula presented in Eq. (3).

$$\text{Z-pool} = [\text{maxpool}, \text{avgpool}] \quad (3)$$

$$O = \frac{1}{3} \begin{pmatrix} \overline{I_1 \sigma \left(\psi_1 \left(\hat{I}_1^* \right) \right)} \\ + \overline{I_2 \sigma \left(\psi_2 \left(\hat{I}_2^* \right) \right)} \\ + \overline{I \sigma \left(\psi_3 \left(\hat{I}_3^* \right) \right)} \end{pmatrix} \quad (4)$$

Assuming the input $I \in T^{C \times H \times W}$, it is directed into three parallel branches of the module. In each branch, spatial dimensions H or W interact with channel dimensions C . Permutation transformations are applied to the input of each branch. It is processed through a Z-pool and GSConv layers to produce attention weights. The permuted input tensor is restored to its original format by applying weights obtained from a Sigmoid activation layer. A final output $O \in T^{C \times H \times W}$ is obtained by combining and averaging the tensors of the three branches, as shown in Eq. (4).

3.3 ParNet-C2f Module

YOLOv8n network structure incorporates many C2f modules, which are crucial for feature extraction. Network performance quality directly relates to how well these C2f modules extract features. Detecting safety helmets, which vary widely in size and color and often appear against complex backgrounds with significant obstructions, poses a challenge for the current C2f modules to capture the boundary information of the objects accurately. This limitation hinders their ability to extract features from various helmets effectively.

Accordingly, this paper introduces the newly developed ParNet-C2f module, which replaces all Bottleneck modules within the existing C2f module with ParNet-Bottleneck modules. Fig. 6 illustrates the ParNet-C2f module design, which is based on the principles of the ParNet network.

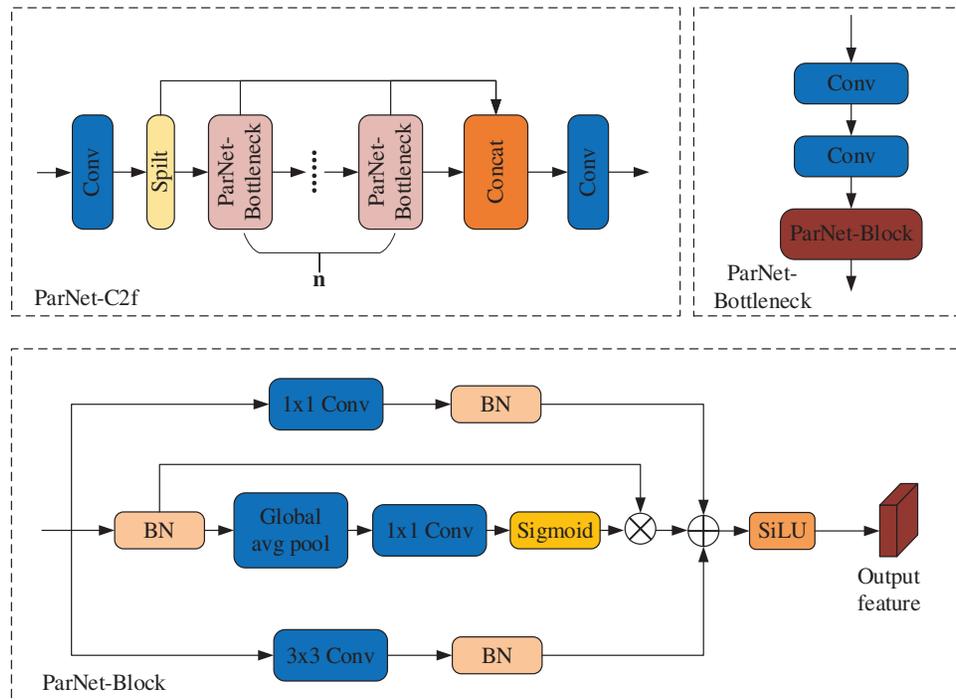


Figure 6: ParNet-C2f structure diagram

ParNet is based on structural reparameterization and employs a non-depth neural network design, offering performance benefits over deeper networks [38]. The ParNet-Block is the core component of ParNet that allows the neural network to focus on the most informative parts, enhancing its feature extraction capabilities. The ParNet-Block comprises three parallel branches: In the initial branch, input features go through a convolution layer to assign weights. The second branch utilizes a residual structure, where feature information, after passing through an average pooling layer, mitigates the convolution layer's excessive sensitivity to position while preserving background information. Weights are adjusted via a convolution layer, processed through a Sigmoid activation function, and multiplied by the original feature map. The third branch extends the perceptual field of the feature map through another convolution layer. Ultimately, feature information from all three branches is combined in the output. The ParNet-Block utilizes multi-branch fusion to fine-tune the network's focus on the image, expand the perceptual field, and bolster the backbone network's feature extraction capabilities.

3.4 WIoU Loss Function

Based on the Intersection over Union (IoU), which evaluates the overlap between predicted and truth boxes, this metric is extensively utilized in object detection tasks. By default, YOLOv8n employs the CIoU loss function, which evaluates errors between model outputs and actual labels by considering the overlap area, the distance between the centers of the target and prediction boxes, and the angle to calculate localization losses accurately. This computation is detailed in Eq. (5). In this formula, IoU represents the

overlapping area divided by the combined area of the predicted and truth boxes, P denotes the Euclidean distance between b , b and indicates the centers of the expected and actual boxes, respectively, c refers to the diagonal length of the smallest rectangle enclosing both boxes, α is a weighting factor, and ν is the similarity in aspect ratios between the actual and predicted boxes. CIoU utilizes a monotonic focusing mechanism, which could detrimentally affect the model's detection capabilities.

$$L_{\text{CIoU}} = 1 - \text{IOU} + \frac{P^2(b, b^{gt})}{c^2} + \alpha\nu \quad (5)$$

Given the limitations of traditional metrics, this paper adopts the WIoU. WIoU sets itself apart by evenly distributing penalties between high and low-quality boxes and employing a dynamic mechanism that maintains feature information across different scales, thus improving the model's overall performance. WIoU implements a new dynamic non-monotonic focusing mechanism that uses the "outlier degree" rather than traditional IoU to evaluate the quality of anchor boxes, enhancing the distribution of gradient enhancements [39]. This approach reduces the dominance of high-quality anchor boxes. It mitigates the adverse gradients produced by low-quality anchor boxes, enabling the network to focus more effectively on average anchor boxes. This setup promotes the development of distance attention and leads to the formulation of WIoU-v1, as detailed in Eqs. (6)–(8).

$$L_{\text{WIoU-v1}} = R_{\text{WIoU}} L_{\text{IoU}} \quad (6)$$

$$R_{\text{WIoU}} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)^*}\right) \quad (7)$$

$$L_{\text{IoU}} = 1 - \text{IoU} \quad (8)$$

$R_{\text{WIoU}} \in [1, e)$ is used as a distance metric to weight the location information of the target boxes, which significantly boosts the L_{IoU} of average-quality anchor boxes. $L_{\text{IoU}} \in [0, 1]$, the closer the prediction box is to perfect overlap with the actual box, the closer the L_{IoU} value gets to 1, indicating a well-sized prediction box. Conversely, a lower L_{IoU} value signifies less overlap, which triggers adjustments to shrink or expand the prediction box to better match the actual box. W_g and H_g denote the dimensions of the smallest bounding box, which includes both the predicted and actual boxes. $*$ indicates a disassociation operation, intended to lessen negative impacts on the training of the model. x_{gt} and y_{gt} denote the central coordinates of the actual box.

4 Experiments and Results Analysis

4.1 Datasets

In this experiment, a custom safety helmet-wearing detection dataset was employed. Images were sourced from surveillance video captures, on-site photography, and web scraping, with each image annotated using the LabelImg tool. The dataset comprises 8983 images from diverse, complex scenarios, including construction sites, aerial work, and offshore operations. It spans all times of day—morning, noon, and night, including 88,590 instances of wearing safety helmets (positive samples) and 13,926 cases of not wearing safety helmets (negative samples). It was partitioned into a training set of 6288 images, a validation set of 1797 images, and a test set of 898 images, adhering to a 7:2:1 ratio. Furthermore, we utilized Mosaic data augmentation to enhance the model's generalization and robustness. This technique combines multiple images into a single training sample, increasing the diversity of the dataset and enabling the model to handle

variations in object appearance and background better. This dataset includes three sample labels: Person, Head, and Helmet, each thoroughly described as shown in [Table 1](#).

Table 1: Tag name and meaning

Labels	Label names	Label meanings
0	Person	Worker
1	Head	Without a safety helmet
2	Helmet	Wearing a safety helmet

4.2 Evaluation Metrics

Network performance was assessed using several metrics, including Precision (P), Recall (R), mAP, parameters, model size, GFLOPs, and FPS. The methods for computing these metrics are outlined in [Eqs. \(9\)–\(11\)](#).

$$P = \frac{TQ}{TQ + FQ} \quad (9)$$

$$R = \frac{TQ}{TQ + FN} \quad (10)$$

$$mAP = \frac{1}{n} \sum_{l=1}^n \int_0^1 P(R) d(R) \quad (11)$$

The formulas used for evaluating network performance, TQ represents the number of actual positive samples correctly predicted, FQ denotes the number of positive samples wrongly predicted, and correspond to the number of negative samples incorrectly predicted. The total number of classes in the data is represented by n .

4.3 Model Training Parameters and Results

The experiments outlined in this article took place on a PC that runs Win11, including an RTX 3060GPU, utilizing CUDA version 12.3. Simulations were performed using the Pytorch framework. The training parameters were set with an input image resolution of 640×640 , employing an SGD optimizer to enhance the model across 300 training epochs. The batch size was established at 8, with an equivalent number of worker threads, and all other parameters were maintained at default settings. The results from training and evaluating the FGP-YOLOv8 model are displayed in [Fig. 7](#).

The Precision-Recall (PR) curve is a widely recognized method for assessing model performance. The PR curve discussed in this study is depicted in [Fig. 8](#).

The PR curve places Recall on the x -axis and Precision on the y -axis, visually depicting precision variations at various recall levels. An expanded area under the PR curve indicates an improved balance between precision and recall, indicating superior model performance.

4.4 Ablation Experiment

To assess and study the effect of various enhancement modules on the algorithm's detection capabilities, five sets of ablation experiments were carried out using different modules, with both training and testing carried out on a proprietary dataset. Using YOLOv8n as the foundational model, its backbone network

was initially swapped for a FasterNet module, producing Model A. Subsequently, the GSTA module was incorporated into the neck network, yielding Model B. Next, integrating the ParNet-C2f module resulted in the development of Model C. Lastly, substituting the CIoU with the WIoU gave rise to Model D. The results of the experiments are shown in [Table 2](#).

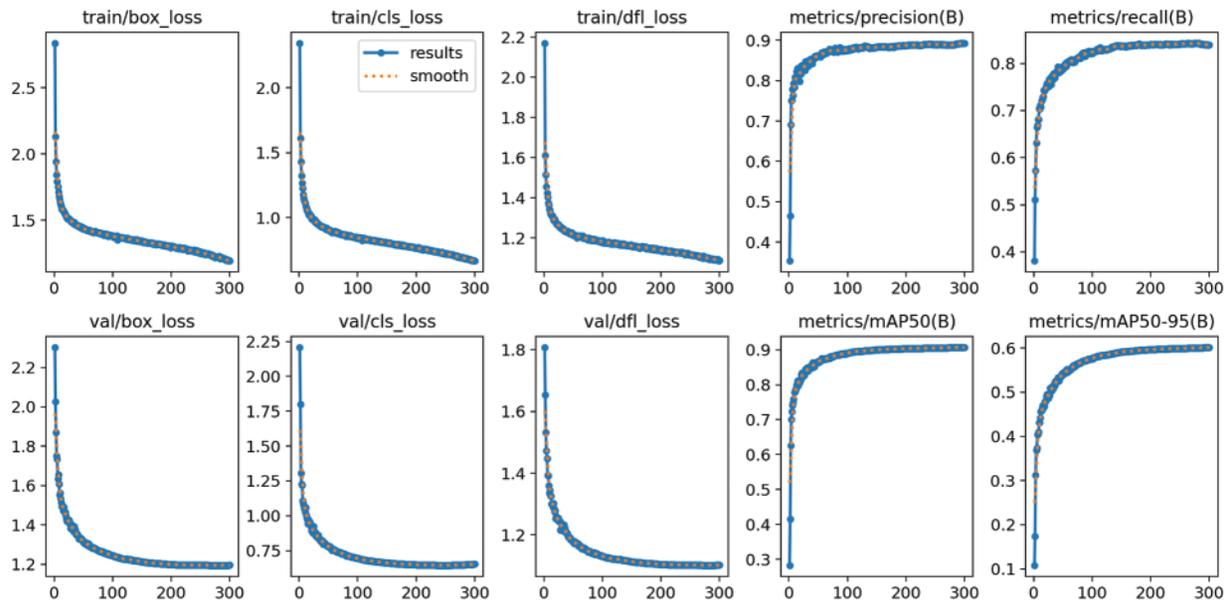


Figure 7: FGP-YOLOv8 model training evaluation results

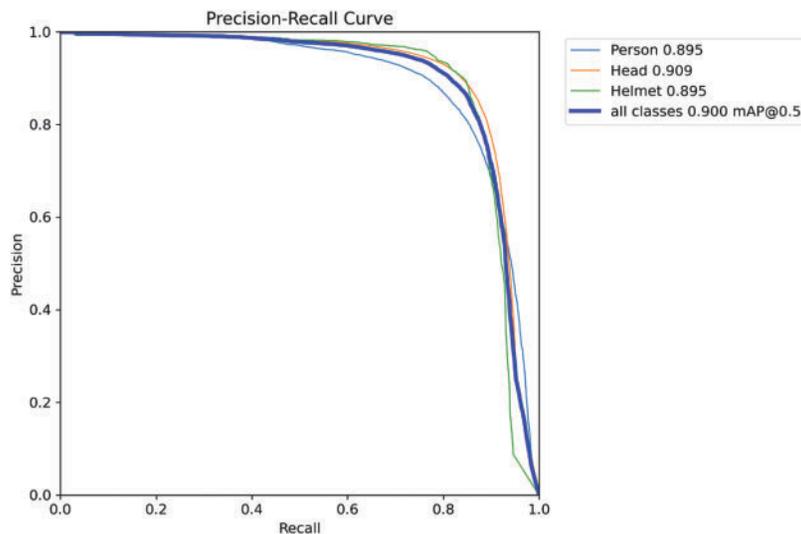


Figure 8: Precision-recall curve

Table 2: Ablation experiment

Model	P/%	R/%	mAP@0.5/%	Params/M	GFLOPs
YOLOv8n	86.3	82.1	87.3	3.01	8.1
A	85.3	81.2	86.6	2.12	5.9
B	86.2	82.8	88.0	2.07	6.1
C	87.4	83.6	89.5	2.41	6.6
D	87.5	83.8	89.6	2.41	6.6

Based on the experimental findings, this study successfully validated five improvement methods, achieving the anticipated goals. The improved FGP-YOLOv8 algorithm showed a 2.3% rise in mAP@0.5, a 1.2% increase in precision, and a 1.7% enhancement in recall compared to the YOLOv8n. Furthermore, it achieved a 19.9% decrease in the number of parameters and an 18.5% decrease in computational load. The ablation studies proved the FGP-YOLOv8 algorithm's superiority in detection accuracy and its lightweight architecture, affirming the viability of the proposed solutions for detecting safety helmet usage.

4.5 Comparative Experiments

4.5.1 Comparative Experiment of Various Lightweight Networks

To assess the improved performance of the proposed lightweight backbone network, several mainstream lightweight models were benchmarked against FasterNet, using the YOLOv8n model as a baseline. The experimental data in Table 3 reveal that all models attained substantial decreases in parameters and computational load compared to YOLOv8n. Notably, ShuffleV2 exhibited a larger reduction in mAP@0.5. PP-LCNet and GhostNet demonstrated robust detection accuracy, though improvements are needed in detection speed. FasterNet markedly improved detection speed while maintaining high accuracy, with only minor reductions of 0.9% in recall and 0.7% in mAP@0.5. It also reduced the parameter and computational burdens by 29.6% and 27.2%, respectively, and enhanced FPS 2.7 times compared to the baseline model, demonstrating that the FasterNet lightweight model can effectively balance model lightness with accuracy while ensuring robust detection performance and real-time processing capabilities.

Table 3: Lightweight network comparison experiment

Model	R/%	mAP@0.5/%	Params/M	GFLOPs	FPS
YOLOv8n	82.1	87.3	3.01	8.1	115.9
+ShuffleV2	79.8	85.9	2.12	5.9	113.8
+PP-LCNet	80.6	86.1	2.12	6	114.3
+GhostNet	81.0	86.3	2.13	5.9	111.0
+FasterNet	81.2	86.6	2.12	5.9	118.6

4.5.2 Comparative Experiment of Various Attention Modules

Using a model augmented with the FasterNet lightweight model as the baseline, this study explores the effects of different attention mechanisms incorporated into the neck network on training and performance, as illustrated in Table 4. Experimentally, the GSTA module in the neck network has been shown to vastly surpass classical attention modules in detection accuracy, achieving a mAP@0.5 of 88%. This enhancement led to increases of 0.9% in precision and 1.4% in recall while only causing a slight rise in computational

load by 0.2 G. This demonstrates that the model's detection ability was effectively enhanced with a minimal increase in computational demands. Furthermore, compared to traditional Triplet Attention (TA), the GSTA improved the mAP@0.5 by 0.7% and reduced the computational load by 0.3 G, verifying that the GSTA module enhances detection precision while reducing computational expenses, thus affirming its superiority.

Table 4: Attention module comparison experiment

Model	P/%	R/%	mAP@0.5/%	GFLOPs
Baseline	85.3	81.2	86.6	5.9
+SE	85.1	81.6	86.8	5.6
+CBAM	85.7	81.8	87.1	5.8
+TA	85.9	82.1	87.3	6.4
+GSTA	86.2	82.8	88.0	6.1

4.5.3 Comparative Experiment of Various C2f Modules

To confirm the improved performance of the proposed ParNet-C2f module, this paper employs a model augmented with the FasterNet lightweight network and GSTA module as a foundation, conducting comparative experiments with a series of well-known C2f enhancement modules. The findings, shown in Table 5, indicate that incorporating the SCConv module led to a 1.5% reduction in mAP@0.5. In contrast, the NAM and CoT modules slightly improved mAP@0.5 by 0.2% and 0.3% but were accompanied by a significant decrease in FPS. In contrast, the ParNet-C2f module developed in this study increased mAP@0.5 by 1.5%, enhanced recall by 0.8%, and achieved an FPS of 103.4, surpassing alternative approaches. This indicates that the enhancements from the ParNet-C2f module effectively satisfy the requirements for real-time, high-precision detection of helmets.

Table 5: C2f module comparative experiment

Model	R/%	mAP@0.5/%	Params/M	GFLOPs	FPS
Baseline	82.8	88.0	2.07	6.1	115.9
+SCConv-C2f	81.3	86.6	2.37	6.5	99.4
+NAM-C2f	83	88.2	2.09	6.1	97.4
+CoT-C2f	83.1	88.3	2.35	6.5	90.3
+ParNet-C2f	83.6	89.5	2.41	6.6	103.4

4.5.4 Comparative Experiment of Various Loss Functions

To confirm the effectiveness of the proposed improvements to the loss function, experiments were conducted employing four loss functions-CIoU, SIoU (Scylla-IoU), EIoU (Efficient-IoU), and WIoU- for comparative analysis. As detailed in Table 6, the results demonstrate that only WIoU improved mAP@0.5 while maintaining stable parameters and computational effort among the four loss functions.

Table 6: Loss function comparison experiment

Model	mAP@0.5/%	Params/M	GFLOPs
CIoU	89.5	2.41	6.6
SIoU	89.0	2.41	6.6

(Continued)

Table 6 (continued)

Model	mAP@0.5/%	Params/M	GFLOPs
EIoU	89.3	2.41	6.6
WIoU	89.6	2.41	6.6

4.5.5 Figure Format

To evaluate the impact of the FGP-YOLOv8 algorithm modifications, comparative tests were carried out under uniform experimental settings compared with contemporary leading algorithms, including the Faster-RCNN, the SSD, and the YOLO series from YOLOv5 to YOLOv8, as well as other enhanced algorithms. [Table 7](#) provides a detailed overview of the detection results for each algorithm.

Table 7: Comparison experiment of mainstream algorithms

Model	R/%	mAP@0.5/%	Params/M	GFLOPs	FPS
YOLOv8n (Baseline)	82.1	87.3	3.01	8.1	115.9
Faster-RCNN	–	82.2	41.3	206.4	13.1
SSD	–	77.7	24.5	87.9	21.3
YOLOv5n	81.5	86.7	2.5	7.2	99.6
YOLOv5s	83.2	88.5	9.1	24	77.7
YOLOv7-tiny	81.4	87.4	6.01	13.1	83.7
YOLOv8s	83.4	89.4	11.12	28.4	88.0
YOLOv5- BEH [17]	83.1	89.2	14.47	56.8	63.8
YOLOv8- SLIM-CA [23]	83.2	88.5	2.79	11.4	77.59
FGP-YOLOv8 (Ours)	83.8	89.6	2.41	6.6	102.7

According to the experimental outcomes, the FGP-YOLOv8 algorithm presented here achieved a mAP@0.5 of 89.6%, outperforming all other algorithms featured in the table. This algorithm shows a 7.4% improvement in mAP@0.5 over the two-stage Faster-RCNN, a 2.3% enhancement over the original YOLOv8n, and a 0.2% increase compared to YOLOv8s. It also recorded a 0.4% enhancement over YOLOv5-BEH and a 1.1% improvement relative to YOLOv8-SLIM-CA. While increases in accuracy generally lead to higher parameters and computational effort, the FGP-YOLOv8 maintains the lowest parameters and demands of computational, decreasing by 0.6 M and 1.5 G, respectively, compared to the YOLOv8n. Additionally, compared to YOLOv8s, it shows reductions of 8.71 M in parameters and 22.2 G in computational load. In terms of real-time performance, although the FPS of FGP-YOLOv8 is lower than that of the original YOLOv8n, it still reaches 102.7 frames, significantly faster than Faster-RCNN, YOLOv5n, YOLOv8s, and other algorithms, thus meeting the requirements for real-time detection. These results demonstrate that the FGP-YOLOv8 improves detection performance and preserves a lightweight profile, delivering precise detection results instantaneously and meeting real-time performance demands.

To comprehensively showcase the FGP-YOLOv8's performance in safety helmet detection, the baseline YOLOv8n and recent improvements to YOLO series algorithms were selected for detection and comparison on the test set. The outcomes are depicted in Fig. 9, and the detection performances of the various algorithms are detailed in Table 8. Fig. 9 presents a series of images where the detection outcomes of YOLOv8n, YOLOv5-BEH, YOLOv8-SLIM-CA, and FGP-YOLOv8 are displayed sequentially from left to right.

The comparison in Fig. 9a reveals that the YOLOv8n algorithm underperforms in scenarios with dense targets, missing detections, and exhibiting low accuracy. In contrast, the FGP-YOLOv8 algorithm significantly enhances target detection and positioning for safety helmets, effectively improving miss rates and resolving accuracy issues. In the scenes with obstructed targets depicted in Fig. 9b, both YOLOv8n and YOLOv5-BEH algorithms experience missed detections. However, the enhanced algorithms successfully detect all targets and increase confidence levels. Fig. 9c illustrates that the YOLOv8n algorithm misses three targets due to background clutter on the left and upper right sides, with YOLOv5-BEH and YOLOv8-SLIM-CA also exhibiting misses and false positives. Conversely, the algorithm discussed in this paper successfully detects all mentioned targets. In Fig. 9d, the original algorithm misses half of the targets in the small target detection scene. YOLOv5-BEH encounters issues with misses and false positives, and although YOLOv8-SLIM-CA reduces false positives, it still misses targets. However, the algorithm introduced in this paper successfully detects every target. Fig. 9e presents a low-light detection scenario where the YOLOv8n algorithm significantly misses detections. YOLOv5-BEH and YOLOv8-SLIM-CA slightly improve but still suffer misses due to poor lighting. The FGP-YOLOv8 algorithm introduced in this document effectively detects all targets and enhances confidence under these challenging conditions.

The experimental outcomes clearly show that the algorithm described in this paper excels in detection performance while maintaining a lightweight architecture, making it well-suited for deployment on resource-limited devices.

4.6 Discussions

Firstly, while FGP-YOLOv8 demonstrates excellent detection performance, the relatively small size of the training and testing dataset poses a limitation. This constraint may lead to missed or false detections when the model is applied to other detection scenarios or environments. To address this, further efforts are needed to collect more diverse images from various construction site environments to expand and enhance the dataset.

Additionally, the proposed FGP-YOLOv8 model achieves impressive parameter size and detection accuracy results. However, its detection speed is slightly lower compared to the original YOLOv8n. Future work will focus on optimizing the model structure to improve detection speed while maintaining its compact size and high accuracy.

Table 8: Detection effectiveness of each algorithm

Model	Dense object scenario	Occluded object scenario	Background interference scenario	Small-object detection scenario	Low-light detection scenario
YOLOv8n	Missed detection	Missed detection	Missed detection	Missed detection	Missed detection
YOLOv5-BEH	Missed detection	Missed detection	Missed detection and false detection	Missed detection	Missed detection
YOLOv8-SLIM-CA			Missed detection and false detection	Missed detection	Missed detection
FGP-YOLOv8					

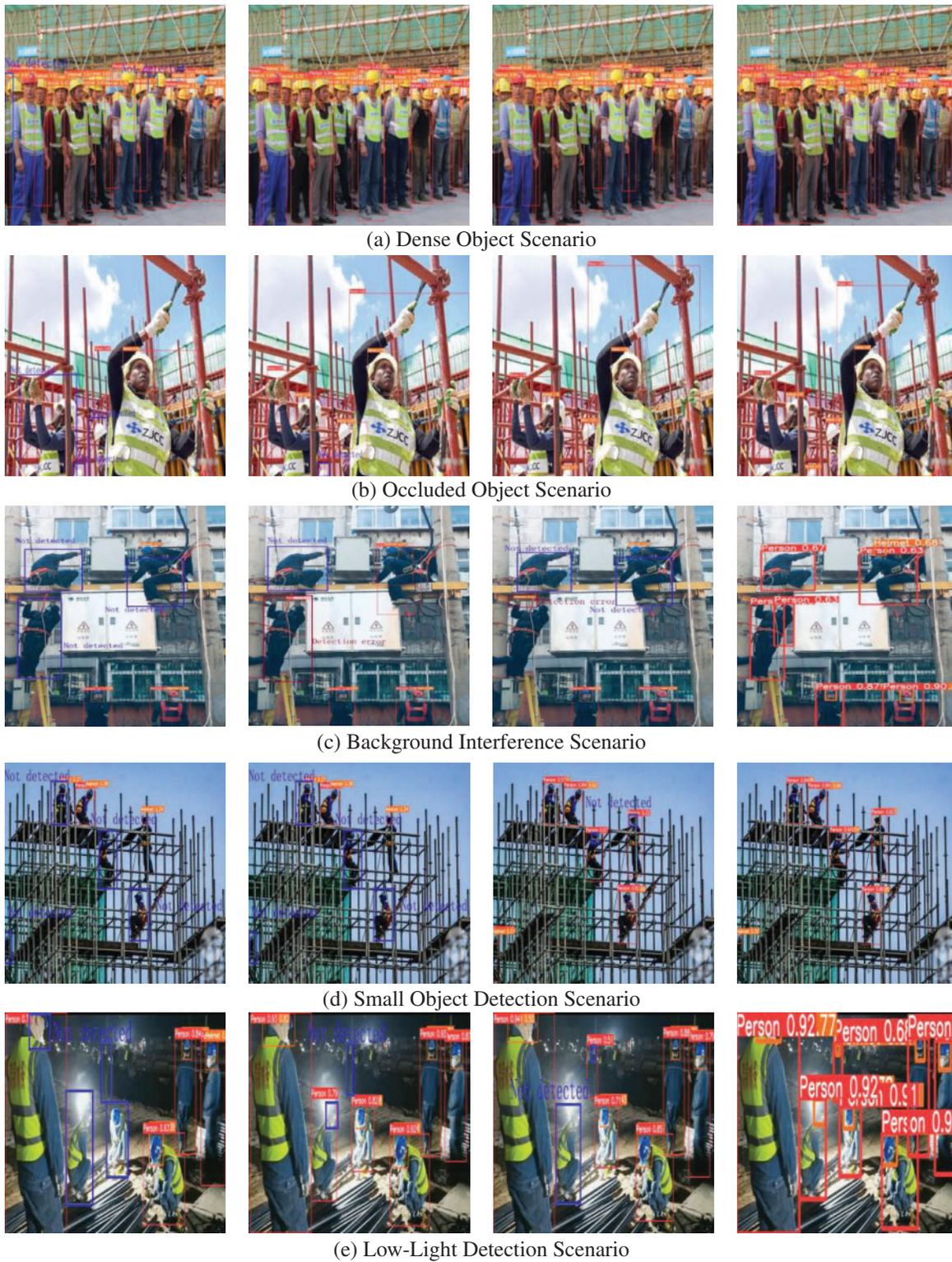


Figure 9: Comparison chart of detection effects of various algorithms

5 Conclusions

In response to challenges such as high parameter volume, low accuracy, and inadequate real-time capabilities in safety helmet detection within complex environments, this article introduces the lightweight FGP-YOLOv8 algorithm, an enhancement of the original YOLOv8n. The algorithm's key innovations include replacing the backbone network with a FasterNet lightweight network that primarily employs PConv, substantially decreasing the model's parameters and computational requirements without compromising detection accuracy. Additionally, introducing the newly proposed GSTA module in the neck network enhances the interplay between spatial and channel attentions, further reducing the computational load and improving the model's detection capacity. Subsequently, a novel ParNet-C2f module is introduced, incorporating a parallel network structure to broaden the perceptual field of the feature map, thus improving the model's feature extraction efficiency. Finally, the WIoU loss function is employed to further enhance the model's overall performance.

In conclusion, the improved model exhibits exceptional performance in detecting safety helmets in complex environments, minimizes computational costs, and fulfills the demands of real-time detection. However, there are still some shortcomings in the improvements proposed in this study. For instance, while we reduced the model size and achieved excellent detection accuracy, there is still room for improvement in detection speed. In the future, we will further optimize the model's performance to achieve faster and more accurate detection. At the same time, we will continue to advance our research in deep learning, with a particular emphasis on lightweight module and small object detection. These two directions will also be key focuses of our future work.

Acknowledgement: The authors would like to express their gratitude for the valuable feedback and suggestions provided by all the anonymous reviewers and the editorial team.

Funding Statement: This research was funded by National Natural Science Foundation of China (61741303), the Foundation Project of Guangxi Key Laboratory of Spatial Information and Mapping (No. 21-238-21-16).

Author Contributions: Hao Ma: Conceptualization, Methodology, Writing—Original Draft, Investigation, Validation, Software. Lieping Zhang: Conceptualization, Methodology, Writing—Reviewing & Editing, Investigation. Jiancheng Huang: Analysis and Interpretation of Results. Cui Zhang: Writing—Review & Editing, Investigation. Xiaolin Gao: Visualization. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author, Cui Zhang, upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Wei L, Liu P, Ren H, Xiao D. Research on helmet wearing detection method based on deep learning. *Sci Rep.* 2024;14(1):7010. doi:10.1038/s41598-024-57433-z.
2. Zhou S, Peng Z, Zhang H, Hu Q, Lu H, Zhang Z. Helmet-YOLO: a new method for real-time, high-precision helmet wearing detection. *IEEE Access.* 2024. doi:10.1109/ACCESS.2024.3443146.
3. Ahmed MIB, Sarairoh L, Rahman A, Al-Qarawi S, Mhran A, Al-Jalaoud J, et al. Personal protective equipment detection: a deep-learning-based sustainable approach. *Sustainability.* 2023;15(18):13990. doi:10.3390/su151813990.
4. Girshick R. Fast R-CNN. *arXiv:1504.08083.* 2015.
5. 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.* 2017;39(6):1137–49. doi:10.1109/TPAMI.2016.2577031.

6. Fan Z, Peng C, Dai L, Cao F, Qi J, Hua W. A deep learning-based ensemble method for helmet-wearing detection. *PeerJ Comput Sci.* 2020;6(6):e311. doi:10.7717/peerj-cs.311.
7. Espinosa Oviedo JE, Velastín SA, Branch Bedoya JW. EspiNet V2: a region based deep learning model for detecting motorcycles in urban scenarios. *Dyna.* 2019;86(211):317–26. doi:10.15446/dyna.v86n211.81639.
8. Li X, Hao T, Li F, Zhao L, Wang Z. Faster R-CNN-LSTM construction site unsafe behavior recognition model. *Appl Sci.* 2023;13(19):10700. doi:10.3390/app131910700.
9. Sun Z, Xuan Y, Fan L, Han R, Tu Y, Huang J, et al. Security monitoring strategy of distribution community operation site based on intelligent image processing method. *Front Energy Res.* 2022;10:931515. doi:10.3389/fenrg.2022.931515.
10. Zhang W, Yang CF, Jiang F, Gao XZ, Zhang X. Safety helmet wearing detection based on image processing and deep learning. In: 2020 International Conference on Communications, Information System and Computer Engineering (CISCE); 2020 Jul 3–5; Kuala Lumpur, Malaysia. p. 343–7. doi:10.1109/cisce50729.2020.00076.
11. Li J, Xie S, Zhou X, Zhang L, Li X. Real-time detection of coal mine safety helmet based on improved YOLOv8. *J Real Time Image Process.* 2024;22(1):26. doi:10.1007/s11554-024-01604-8.
12. Li T, Li G. Road defect identification and location method based on an improved ML-YOLO algorithm. *Sensors.* 2024;24(21):6783. doi:10.3390/s24216783.
13. Yu C, Liu Y, Zhang W, Zhang X, Zhang Y, Jiang X. Foreign objects identification of transmission line based on improved YOLOv7. *IEEE Access.* 2023;11:51997–2008. doi:10.1109/ACCESS.2023.3277954.
14. Ku B, Kim K, Jeong J. Real-time ISR-YOLOv4 based small object detection for safe shop floor in smart factories. *Electronics.* 2022;11(15):2348. doi:10.3390/electronics11152348.
15. Zhang B, Sun CF, Fang SQ, Zhao YH, Su S. Workshop safety helmet wearing detection model based on SCM-YOLO. *Sensors.* 2022;22(17):6702. doi:10.3390/s22176702.
16. Sadiq M, Masood S, Pal O. FD-YOLOv5: a fuzzy image enhancement based robust object detection model for safety helmet detection. *Int J Fuzzy Syst.* 2022;24(5):2600–16. doi:10.1007/s40815-022-01267-2.
17. Shan C, Liu H, Yu Y. Research on improved algorithm for helmet detection based on YOLOv5. *Sci Rep.* 2023;13(1):18056. doi:10.1038/s41598-023-45383-x.
18. Yao W, Wang A, Nie Y, Lv Z, Nie S, Huang C, et al. Study on the recognition of coal miners' unsafe behavior and status in the hoist cage based on machine vision. *Sensors.* 2023;23(21):8794. doi:10.3390/s23218794.
19. An Q, Xu Y, Yu J, Tang M, Liu T, Xu F. Research on safety helmet detection algorithm based on improved YOLOv5s. *Sensors.* 2023;23(13):5824. doi:10.3390/s23135824.
20. Song H, Zhang X, Song J, Zhao J. Detection and tracking of safety helmet based on DeepSort and YOLOv5. *Multimed Tools Appl.* 2023;82(7):10781–94. doi:10.1007/s11042-022-13305-0.
21. Wang S, Wu P, Wu Q. Safety helmet detection based on improved YOLOv7-tiny with multiple feature enhancement. *J Real Time Image Process.* 2024;21(4):120. doi:10.1007/s11554-024-01501-0.
22. Liu Y, Jiang X, Xu R, Cui Y, Yu C, Yang J, et al. A novel foreign object detection method in transmission lines based on improved YOLOv8n. *Comput Mater Contin.* 2024;79(1):1263–79. doi:10.32604/cmc.2024.048864.
23. Lin B. Safety helmet detection based on improved YOLOv8. *IEEE Access.* 2024;12(3):28260–72. doi:10.1109/ACCESS.2024.3368161.
24. Aboah A, Wang B, Bagci U, Adu-Gyamfi Y. Real-time multi-class helmet violation detection using few-shot data sampling technique and YOLOv8. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 2023 Jun 17–24; Vancouver, BC, Canada: IEEE. p. 5350–8. doi:10.1109/CVPRW59228.2023.00564.
25. Li Y, Xu H, Zhu X, Huang X, Li H. THDet: a lightweight and efficient traffic helmet object detector based on YOLOv8. *Digit Signal Process.* 2024;155(3):104765. doi:10.1016/j.dsp.2024.104765.
26. Yu C, Lu Z. YOLO-VSI: an improved YOLOv8 model for detecting railway turnouts defects in complex environments. *Comput Mater Contin.* 2024;81(2):3261–80. doi:10.32604/cmc.2024.056413.
27. Bakirci M. Enhancing vehicle detection in intelligent transportation systems via autonomous UAV platform and YOLOv8 integration. *Appl Soft Comput.* 2024;164(5):112015. doi:10.1016/j.asoc.2024.112015.
28. Lin B. YOLOv8n-ASF-DH: an enhanced safety helmet detection method. *IEEE Access.* 2024;12(17):126313–28. doi:10.1109/ACCESS.2024.3435453.

29. Shi C, Zhu D, Shen J, Zheng Y, Zhou C. GBSG-YOLOv8n: a model for enhanced personal protective equipment detection in industrial environments. *Electronics*. 2023;12(22):4628. doi:10.3390/electronics12224628.
30. Hu L, Ren J. YOLO-LHD: an enhanced lightweight approach for helmet wearing detection in industrial environments. *Front Built Environ*. 2023;9:1288445. doi:10.3389/fbuil.2023.1288445.
31. Yatheendra K, Hemalatha AN, Dhanasekhar Reddy A, Pavani K. Model for detecting safety helmet wearing using improved yolo-M. *Int J Inform Technold Comput Eng*. 2024;12(3):688–700.
32. Song X, Zhang T, Yi W. An improved YOLOv8 safety helmet wearing detection network. *Sci Rep*. 2024;14(1):17550. doi:10.1038/s41598-024-68446-z.
33. Chen J, Kao SH, He H, Zhuo W, Wen S, Lee CH, et al. Run, don't walk: chasing higher FLOPS for faster neural networks. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2023 Jun 17-24; Vancouver, BC, Canada: IEEE. p. 12021–31. doi:10.1109/CVPR52729.2023.01157.
34. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA: IEEE. p. 7132–41. doi:10.1109/CVPR.2018.00745.
35. Woo S, Park J, Lee JY, Kweon IS. CBAM: convolutional block attention module. In: Proceedings of the European Conference On Computer Vision (ECCV); 2018; Munich, Germany. p. 3–19.
36. Misra D, Nalamada T, Arasanipalai AU, Hou Q. Rotate to attend: convolutional triplet attention module. In: 2021 IEEE Winter Conference on Applications of Computer Vision (WACV); 2021; Waikoloa, HI, USA. p. 3138–47. doi:10.1109/WACV48630.2021.00318.
37. Li H, Li J, Wei H, Liu Z, Zhan Z, Ren Q. Slim-neck by GSConv: a lightweight-design for real-time detector architectures. *J Real Time Image Process*. 2024;21(3):62. doi:10.1007/s11554-024-01436-6.
38. Goyal A, Bochkovskiy A, Deng J, Koltun V. Non-deep networks. arXiv:2110.07641. 2021.
39. Tong Z, Chen Y, Xu Z, Yu R. Wise-IoU: bounding box regression loss with dynamic focusing mechanism. arXiv:2301.10051. 2023.