electronics

MDPI

Article

GBSG-YOLOv8n: A Model for Enhanced Personal Protective Equipment Detection in Industrial Environments

Chenyang Shi, Donglin Zhu, Jiaying Shen, Yangyang Zheng and Changjun Zhou *

College of Computer Science and Technology, Zhejiang Normal University, Jinhua 321004, China
* Correspondence: zhouchangjun@zjnu.edu.cn

Abstract: The timely and accurate detection of whether or not workers in an industrial environment are correctly wearing personal protective equipment (PPE) is paramount for worker safety. However, current PPE detection faces multiple inherent challenges, including complex backgrounds, varying target size ranges, and relatively low accuracy. In response to these challenges, this study presents a novel PPE safety detection model based on YOLOv8n, called GBSG-YOLOv8n. First, the global attention mechanism (GAM) is introduced to enhance the feature extraction capability of the backbone network. Second, the path aggregation network (PANet) structure is optimized in the Neck network, strengthening the model's feature learning ability and achieving multi-scale feature fusion, further improving detection accuracy. Additionally, a new SimC2f structure has been designed to handle image features and more effectively improve detection efficiency. Finally, GhostConv is adopted to optimize the convolution operations, effectively reducing the model's computational complexity. Experimental results demonstrate that, compared to the original YOLOv8n model, the proposed GBSG-YOLOv8n model in this study achieved a 3% improvement in the mean Average Precision (mAP), with a significant reduction in model complexity. This validates the model's practicality in complex industrial environments, enabling a more effective detection of workers' PPE usage and providing reliable protection for achieving worker safety. This study emphasizes the significant potential of computer vision technology in enhancing worker safety and provides a robust reference for future research regarding industrial safety.

Keywords: industrial safety; PPE detection; computer vision; YOLOv8n; SimC2f; GhostConv

Citation: 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*, 4628. https://doi.org/10.3390/electronics12224628

Academic Editor: Dah-Jye Lee

Received: 8 October 2023
Revised: 1 November 2023
Accepted: 7 November 2023
Published: 12 November 2023

1. Introduction

As the global industrial scale continues to expand, incidents involving industrial safety have evolved into a severe global challenge, posing a serious threat to life, property, and the sustainable development of society. In an industrial environment, ensuring workers' safety is paramount. However, with the increasing number of workers and limited supervisory personnel, it becomes challenging to effectively oversee the safety conditions of many workers. Furthermore, insufficient safety awareness among workers significantly increases the risk of industrial safety incidents [1]. As of 28 July 2023, a total of 12,070 various production safety incidents have occurred in China in 2023 alone, resulting in 10,527 fatalities. Therefore, more effective measures are urgently needed to ensure the safety of workers.

Wearing PPE [2] effectively reduces hazards and mitigates safety risks in industrial environments. PPE includes safety helmets, safety goggles, safety vests, face masks, etc. [3]. Safety helmets contribute to head protection, preventing injuries from falling objects and electrical hazards [4]. Safety goggles reduce the risk of eye injuries [5]. Safety vests enhance worker visibility, reducing the risk of collisions with machinery and other objects [6]. Also, face masks are crucial in blocking harmful substances from being inhaled and protecting the respiratory system [7]. However, ensuring that workers wear PPE and use it correctly is a complex task. Therefore, the timely and accurate detection of PPE usage is of paramount importance [8].

In the early industrial environments, researchers typically employed traditional methods to ascertain whether workers were correctly wearing PPE in compliance with existing regulations. These methods required that factory supervisory personnel routinely engage in manual observations and patrols, manually scrutinizing surveillance videos to detect workers' unsafe behaviors, and requiring workers to conduct self-examinations daily and report any issues. While these conventional methods contribute to the monitoring of PPE use, to some extent, they include a series of issues: (1) regulatory personnel are susceptible to external interference, which may lead to oversight and incorrect judgments; (2) subjective factors, such as emotions and psychological states, can affect the objectivity of judgment; (3) manually reviewing all monitoring images is a highly labor-intensive task; and (4) workers may lack sufficient safety awareness. Therefore, to solve these problems, it has become particularly urgent to realize the automation and intelligence of PPE inspection in industrial environments.

Initially, researchers primarily focused on implementing PPE detection using sensor technology. However, these technologies require expensive equipment, increasing the cost of industrial production and posing potential health risks to workers. As technology advances, the application of artificial intelligence in industrial automation is increasingly prevalent. Before developing deep learning technology, the principal approach to the detection of correct PPE usage by workers involved image processing and machine learning. However, these methods did not perform well in complex scenarios with many interferences and could only identify limited types of PPE. As the demand for industrial production increased, the variety of PPE also gradually expanded. The emergence of deep learning technology has led to the utilization of techniques like object detection for use in the field of PPE detection. YOLO, recognized for its exceptional performance characterized by improved detection accuracy and faster processing speeds, has emerged as a prominent representative in object detection. In recent years, YOLO has undergone multiple iterations [9–14] and has been widely applied in various domains [15–19]. It has also begun to be used for PPE detection in industrial environments. However, the current research primarily focuses on construction sites and covers only a limited range of PPE types. With the increasing demands of industrial production, the variety of PPE required by workers is constantly expanding, presenting new detection challenges.

Hence, this study proposes the GBSG-YOLOv8n model, built upon the YOLOv8n framework, for swift and precise PPE detection in industrial settings. The key contributions of this study include the following:

- To enhance the PPE detection model's performance, we have established a new dataset called PPES, which comprises many images captured by cameras in industrial settings, providing ample data resources for research.
- By introducing GAM and embedding it into the model's backbone network, we enhance the focus on PPE targets, suppress interference from non-target background information, and significantly improve the feature extraction capability of the backbone network.
- To effectively integrate feature information from different scales and prevent the loss of PPE feature details, we optimized the PANet structure within the Neck network. This optimization facilitated efficient bidirectional cross-scale connections and feature-weighted fusion, further enhancing detection accuracy.
- We have innovatively designed the SimC2f structure to significantly enhance the performance of the C2f module, resulting in the more efficient processing of image features and an improvement in overall detection efficiency.
- To satisfy the real-time PPE detection and the light weight of the model, we use GhostConv to optimize the convolution operation in the backbone network, which significantly reduces the amount of model computation and parameters, while ensuring high detection accuracy.

The subsequent sections of this paper are organized as follows: Section 2 offers a concise introduction to the pertinent background knowledge. Section 3 presents the GBSG-

YOLOv8n model. In Section 4, we substantiate the effectiveness and success of GBSG-YOLOv8n through comprehensive experiments and analyses. Finally, Sections 5 and 6 are dedicated to discussing and summarizing our work.

2. Related Work

Currently, PPE detection methods fall into two main categories: sensor-based and computer vision-based. Sensor-based methods use installed sensors to analyze signals and assess proper PPE usage by workers. Kelm et al. [20] utilized RFID sensor technology, deployed at the entrances and exits of workplaces, to inspect the safety of PPE. In addition, Bauk et al. [21] proposed an RFID worker safety model tailored to specific workplace requirements, embedding RFID in workers' PPE. Furthermore, Dong et al. [22] introduced an innovative method for automated remote monitoring and evaluation of PPE by integrating pressure sensors and positioning technology. They also developed a real-time locating system (RTLS) to track workers' positions, ensuring the proper use of PPE. Additionally, Hayward et al. [23] presented a PPE access control system prototype that integrates PPE with indoor and outdoor personnel location monitoring systems to ensure that employees and visitors wear PPE correctly. While sensor-based methods assist in detecting the wearing of PPE by workers, these methods incur substantial costs and require intricate deployment and maintenance. Additionally, they are restricted by particular environments and object categories, consequently diminishing detection accuracy.

As technology advances, contemporary industrial production is shifting toward higher speeds, greater precision, and automation. Computer vision-based methods are gradually emerging In PPE detection. These methods fall into two categories: the approach combining image processing with machine learning [24], mainly used for solving issues like image segmentation, feature extraction, and classification, and the method which performs tasks such as target detection and image generation with the help of deep learning techniques [25]. In traditional methods, it is typical to employ image processing techniques to initially locate areas of interest, then extract image features and utilize machine learning methods to train classifiers to determine whether these areas contain PPE. Li et al. [26] utilized the background modeling algorithm to detect moving objects in the field of view of surveillance cameras. After identifying regions of interest linked to motion, they employed the histogram of oriented gradient (HOG) method to extract features. Subsequently, they used the extracted HOG features to train a support vector machine (SVM) for worker classification. Wu et al. [27] presented a color-based hybrid descriptor that combines local binary patterns (LBP), Hu moments invariants (HMI), and color histograms (CH) to capture a wide range of colors. They then implemented a hierarchical support vector machine (H-SVM) for feature classification, facilitating safety helmet detection. Although traditional methods combining image processing and machine learning have been widely applied, the need for manual feature design and extraction poses challenges, especially when dealing with large-scale data, which may lead to issues related to computational resources and memory limitations.

The continuous advancement of deep learning technology has effectively resolved these challenges, prompting numerous researchers to incorporate techniques like object detection into PPE detection. Ross et al. [28] introduced the concept of R-CNN, laying the foundation for deep-learning object detection. Subsequently, object detection algorithms based on candidate boxes were proposed and applied in the PPE detection field. Zhang et al. [29] proposed a safety management framework for on-site construction utilizing computer vision and real-time positioning systems. They employed Fast R-CNN to analyze image data from on-site cameras, enabling object detection and classification, assessing proper PPE usage by workers. Additionally, Fan et al. [30] explored the mechanisms and effectiveness of different target detection algorithms in the context of helmet detection. The findings highlighted the remarkable accuracy achieved by Faster R-CNN. They further improved the helmet detection algorithm by integrating models, resulting in enhanced detection capabilities. Although these two-stage object detection algorithms excel

in accuracy, they must first locate candidate regions in the image and then employ a classifier to categorize each candidate region. These two steps individually require substantial computational resources, resulting in relatively lower efficiency for two-stage algorithms. In contrast, by using CNN for end-to-end processing, one-stage object detection algorithms directly generate multiple bounding boxes in the image and predict classification probabilities for each bounding box. This approach achieves simultaneous localization and classification, reducing the demand for computational resources and enhancing efficiency. SSD [31] and YOLO [32] algorithms are notable examples known for their capability to achieve higher detection accuracy and faster processing speeds. Han et al. [33] introduced an object detection algorithm based on a cross-layer attention mechanism and multi-scale perception. This approach effectively detects the use of safety helmets, building upon the foundation of the SSD algorithm, and it demonstrates a significant improvement in accuracy. Wang et al. [34] utilized the YOLOv3 model at a construction site to detect workers and the presence of safety helmets. Jiang et al. [35] enhanced YOLOv3 by incorporating squeeze-and-excitation (SE) blocks between convolution layers in Darknet53, substituting the mean squared error (MSE) with GIoU loss, and employing focal loss to mitigate the significant foreground–background class imbalance issue, thereby more effectively achieving the real-time monitoring of mask-wearing. Ji et al. [36] introduced a residual feature enhancement module based on YOLOv4, reducing the loss of valuable information in high-level feature maps, enhancing object detection accuracy, and enabling the timely detection of workers who not wearing safety helmets or clothing in industrial environments. Wang et al. [37] tested the performance of YOLOv3, YOLOv4, and YOLOv5 on a custom dataset. The findings demonstrate that the YOLOv5 model outperformed the others. Zhang et al. [38] introduced shallow detection heads tailored for small object detection within the YOLOv5 algorithm. These heads are combined with SENet channel attention modules to effectively condense global spatial information. Additionally, they added a denoising module to the backbone network to ensure feature clarity and accuracy, significantly improving helmet detection accuracy. Tai et al. [39] introduced a new dynamic anchor box mechanism based on YOLOv5 for safety helmet detection, improving the model's accuracy in handling target changes. Sun et al. [40] integrated the MCA module into YOLOv5 to obtain more comprehensive feature map data. By employing strategies such as sparse training and channel pruning, they notably improved safety helmet detection performance. Ali et al. [41] assessed the performance of different YOLOv5 and YOLOv7 versions in detecting students' PPE compliance in a laboratory setting using a self-created safety-related dataset. Based on YOLOv7, Wang et al. [42] improved computational efficiency by introducing the CPC structure and combining it with the SA mechanism, enabling the model to concentrate on localized image information at a reduced computational cost, enhancing accuracy and improving it to better respond to the need for the real-time detection of masks in complex scenarios. These studies have demonstrated the effectiveness of deep learning in the field of PPE detection. Table 1 displays models frequently employed in PPE detection, including Faster R-CNN, SSD, YOLOv3, YOLOv4, YOLOv5, and YOLOv7.

In summary, many scholars have conducted extensive research on PPE detection. While specific achievements have been made, the diverse on-site conditions and challenges of detecting small targets limit the field, necessitating high adaptability and generalization. Trim marks are often subject to background interference, and overlapping targets make accurate detection and recognition more complex. Therefore, the primary objective of this study is to address the issues mentioned earlier and to propose a high-performance method for the rapid and accurate detection of PPE in an industrial setting.

Table 1. Comparison analysis of common models in the PPE field.

Model	Algorithm Description	Pros	Cons
Faster R-CNN	Faster R-CNN represents an enhanced iteration within the R-CNN series, improving processing speed and precision. The method involves feature extraction through convolutional neural networks, followed by generating region proposals utilizing the region proposal network (RPN). It ultimately conducts region classification and performs bounding box regression.	Faster R-CNN uses convolutional neural networks to extract features, capturing a wide range of visual attributes of PPE objects, such as their size, shape, and appearance. Its architecture, which relies on RPN and classification networks, significantly enhances its accuracy in recognizing PPE categories.	Faster R-CNN shows slower real-time PPE detection in complex industrial environments, primarily due to its high computational demands. Furthermore, it exhibits reduced accuracy, particularly for small-sized PPE, potentially resulting in missed detections.
SSD	Compared to two-stage methods such as Faster R-CNN, SSD demonstrates superior speed and is well-suited for real-time applications. It achieves this speed by incorporating multi-scale feature maps and advanced convolutional structures.	SSD performs exceptionally well in real-time PPE detection by leveraging multi-scale feature maps to effectively handle complex scenarios and detect PPE objects of various sizes.	In the detection of small targets, SSD exhibits relatively lower accuracy, is susceptible to interference from complex backgrounds, and may lead to instances of missed detections.
YOLOv3	YOLOv3 utilizes three distinct-scale detection heads and employs Darknet-53 as the backbone network, significantly improving feature extraction capabilities.	YOLOv3 supports multi-scale PPE detection and is suitable for real-time detection of different PPE categories, making it especially ideal for industrial environments.	Compared to newer YOLO versions, YOLOv3 may display differences in detection accuracy and computational resource requirements. It shows reduced accuracy in complex environments, possibly necessitating additional training data.
YOLOv4	YOLOv4 introduces enhancements such as CIOU loss, SAM, and PANet, significantly improving detection accuracy. Simultaneously, it adopts a more powerful Darknet-53 network and additional data augmentation techniques, enhancing the model's robustness.	YOLOv4 introduces improved loss functions and network structures, leading to enhanced accuracy in PPE detection. Model optimization and a lightweight design further improve the speed and robustness of PPE detection, rendering it suitable for industrial environments with complex backgrounds.	Compared to more recent YOLO versions, the YOLOv4 model exhibits increased complexity, necessitating more significant computational resources and extended durations of training and deployment.
YOLOv5	YOLOv5 introduces adaptive feature selection and model pruning, reducing model complexity. Additionally, through a lightweight design and model optimization, it accelerates inference speed and enhances detection accuracy.	YOLOv5 balances speed and accuracy in PPE detection by implementing streamlined model pruning techniques that alleviate model complexity. It is particularly apt for application scenarios demanding instantaneous PPE detection.	In specific scenarios, YOLOv5 exhibits slightly diminished accuracy, particularly in small object detection, which requires further improve. Additionally, compared to low-complexity models, it still demands more computational resources.
YOLOv7	YOLOv7 inherits the high performance of the YOLO series, while prioritizing a balance between performance and speed. It introduces enhancements to model design, including the backbone network and loss functions. Utilizing a more advanced backbone network, YOLOv7 significantly improves feature extraction capabilities. Through optimizations in network structure and model design, it achieves a balance between detection speed and accuracy.	YOLOv7 excels in PPE object detection, maintaining high accuracy while improving detection speed, making it suitable for real-time applications. It utilizes a more advanced backbone network and model design, further enhancing the efficiency and accuracy of PPE detection.	YOLOv7 demands higher computational resources and is unsuitable for resource-constrained situations. While excelling in speed and accuracy, there is still room for improvement in PPE detection, which requires high precision and real-time performance, and more training data are needed to further improve the performance.

3. Method

3.1. YOLOv8n Model Analysis

YOLOv8 represents the most recent advancement in the YOLO series of object detection algorithms. It excels at swiftly and precisely detecting objects in images, determining their positions, and classifying their categories by learning object characteristics and shapes. Depending on the network's depth and width, YOLOv8 can be divided into YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. Given the need for PPE object detection under complex and dynamic field conditions and the challenges of detecting small objects, this study has selected the YOLOv8n network from the YOLOv8 series, which has smaller parameters, but higher accuracy. The YOLOv8n model's detection network consists of four components, as illustrated in Figure 1: Input, Backbone, Neck, and Head.

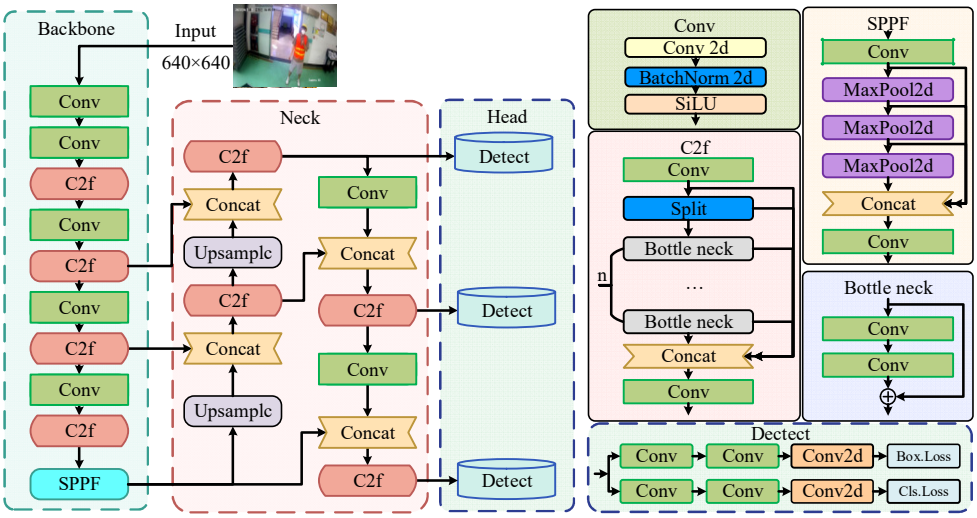

Figure 1. YOLOv8n network structure diagram.

The primary role of the Input network is to receive, preprocess, and transmit input images for feature extraction and object detection. It employs the Mosaic data augmentation to combine multiple photos into a single input sample, enhancing data diversity. Furthermore, input images are resized to a uniform dimension to ensure the model can effectively process images with consistent sizes.

The backbone network conducts feature extraction, converting input images into multiscale feature maps through deep convolutional neural networks, delivering crucial data for subsequent object detection. YOLOv8n incorporates an improved CSPDarknet53 as its backbone network. Unlike traditional backbone networks using cross-stage local CSP modules, YOLOv8n replace them with a C2f module. This module connects through gradient splitting, maintaining the network's lightweight characteristics while enriching information exchange during feature extraction. Finally, the SPPF module pools the input feature maps into fixed-size images to achieve adaptive output size.

The primary role of the Neck network is to amalgamate features across various scales to produce a feature pyramid. It utilizes the PANet structure [43], comprising the feature pyramid network (FPN) [44] and the path aggregation network (PAN). FPN acquires feature maps from the convolutional neural network, constructs a feature pyramid, and employing a top-down approach, combines multi-scale features by using up-sampling and coarser granularity feature maps, thus achieving multi-scale feature fusion. To more effectively retain target position information, the PAN module supplements FPN by adopting a bottom-up structure, fusing feature maps from different levels through convolutional layers, further enhancing detection performance.

The Head network serves as the ultimate prediction component, obtaining category and position information for objects of varying sizes from feature maps of diverse scales. This network uses a decoupled head structure, separating classification and detection

heads. It also adopts the anchor-free concept, eliminating the need for predefined anchors. Instead, it learns various object shapes and sizes to better adapt to object detection tasks in different scenarios.

3.2. Improved Model

The overall architecture of YOLOv8n has been retained, while improving or replacing some of its modules. The GBSG-YOLOv8n model is proposed, and this section will provide a detailed introduction to the model. The network structure of GBSG-YOLOv8n is shown in Figure 2.

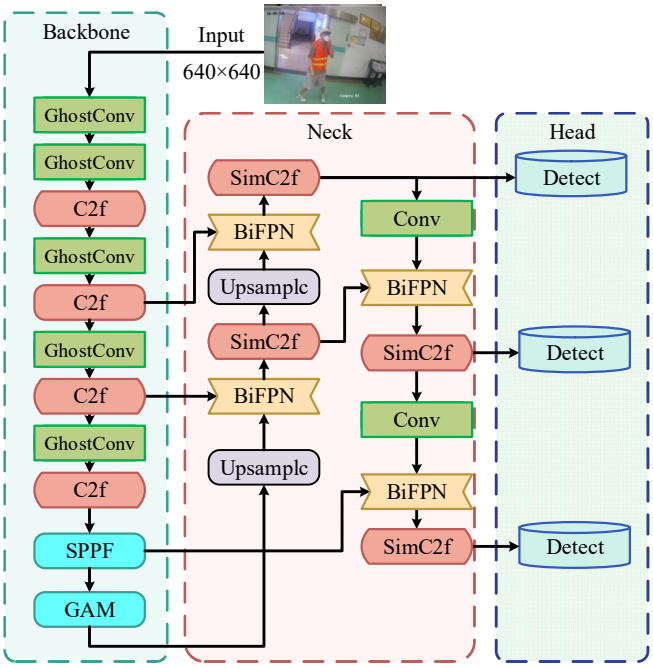

Figure 2. GBSG-YOLOv8n network structure diagram.

3.2.1. Global Attention Mechanism

In PPE safety detection, challenges arise from complex backgrounds and the potential oversight of small targets. Thus, it is vital to enhance the model's feature extraction abilities. Attention mechanisms help the model emphasize essential input information while filtering out less relevant material. In this study, we integrate GAM [45] into the backbone network to magnify the focus on PPE-specific features, strengthening the backbone network's ability to extract critical information. This enhancement ultimately results in improved detection performance and accuracy, as depicted by the GAM structure in Figure 3.

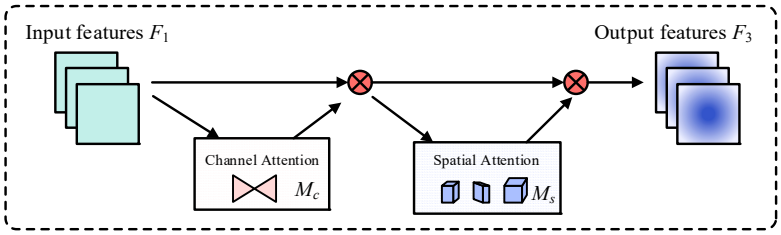

Figure 3. GAM structure diagram.

The GAM comprises two distinct submodules—the channel attention and the spatial attention submodule [46]—designed for channel and spatial attention operations. The primary aim of the channel attention submodule is to augment interactions among global features. It initially preserves information across three dimensions via a 3D permutation

operation. Subsequently, it enhances cross-dimensional channel–spatial dependencies by employing a two-layer MLP structure. Unlike the channel attention submodule, the spatial attention submodule is dedicated to improving the model's focus on spatial information. It commences with max-pooling and average-pooling operations on the input feature maps, followed by the fusion of these two pooling outcomes. The merged feature map is then subjected to convolution and processed with a Sigmoid activation function. The collaborative operation of these two submodules strengthens the interdependence between global and spatial features, allowing our model to more effectively concentrate on the requisite feature information, ultimately leading to improved performance.

3.2.2. Bidirectional Feature Pyramid Network

Various challenges are typically encountered in PPE target detection, including variations in the distance between workers and cameras, leading to inconsistent image resolutions and a wide range of PPE target sizes, ranging from small items, like safety goggles, to larger ones, like safety vests. Consequently, effectively integrating this diverse information for multi-scale targets remains challenging.

To tackle this problem, YOLOv8n utilizes the PANet structure to build a feature pyramid, facilitating the fusion of multi-scale feature information and enhancing target detection performance. Although the PANet structure effectively integrates features of different scales through top-down and bottom-up information propagation, improving its ability to detect and handle changes in scale for various target sizes, it nonetheless exhibits some inherent problems. For instance, in the PANet structure, nodes with only one input edge and those lacking feature fusion exist. This can lead to unbalanced information transmission and an increase in additional parameters and computational burden. In response to these issues, this study introduces the BiFPN structure [47] as a new Neck component. The BiFPN structure is a deep learning network architecture specifically designed for object detection tasks, and it improves the PANet structure. The BiFPN structure removes redundant nodes from the top and bottom layers of the PANet structure. It incorporates various connection methods, such as horizontal, vertical, and cross-scale connections, to enhance feature fusion efficiency. This enables the model to better adapt to multi-scale targets and reduces the risk of losing crucial information. Furthermore, the BiFPN structure can be stacked multiple times as needed, further enhancing the fusion of multi-scale features without introducing excessive redundant parameters and computational burden. These improvements significantly boost the model's performance in regards to multi-scale object detection tasks. The structure is shown in Figure 4.

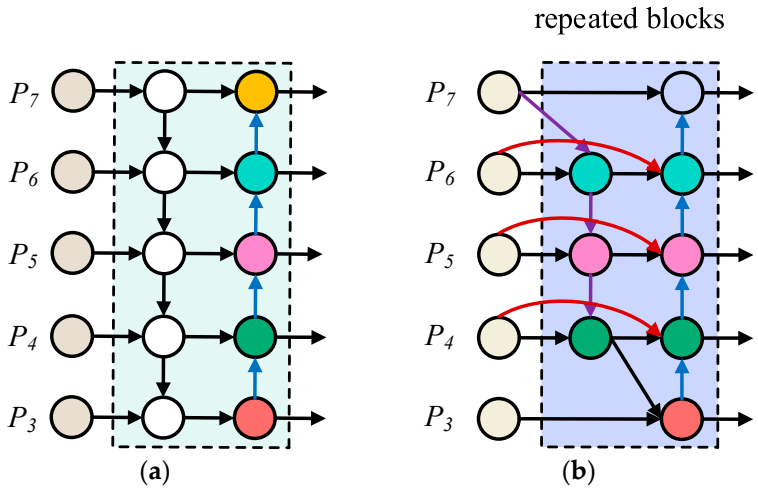

Figure 4. Schematic of the different feature fusion structures: (**a**) PANet; (**b**) BiFPN.

Compared to the PANet structure, the BiFPN structure possesses a more robust feature fusion capability, effectively integrating more feature information without increasing the

computational burden. Furthermore, its distinctive skip connection structure proficiently addresses the issue of feature loss, illustrated in Figure 5. In cases where the Neck network's initial input and output nodes coincide within the same layer, the supplementary edges are consolidated to minimize spatial information loss.

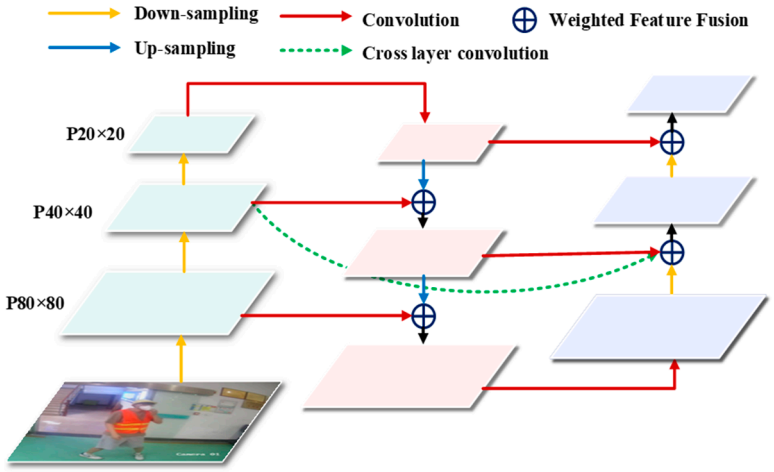

Figure 5. The architecture of BiFPN.

Due to the varying resolutions of input feature maps, BiFPN assigns weights to each additional feature layer during feature fusion, adapting them based on their contribution to the network. The model training process emphasizes learning the features with significant weight allocations and performing multi-scale feature fusion through multiple iterations. The weighted formula for BiFPN is as follows:

$$O = \sum_i \frac{\omega_i I_i}{\varepsilon + \sum_j \omega_j} \tag{1}$$

where I_i represents the input features, O is the output features, and ω_i and ω_j are learnable weights. The introduction of the ReLU activation function maps the learnable weights to the range [0, 1], and $\varepsilon = 0.0001$ is a minimal value added to ensure output stability.

In this study, we substituted the PANet structure in the model with the BiFPN structure to enhance the transmission and extraction of multi-scale PPE target features. This improvement aims to boost the PPE detection performance for targets of various scales, consequently elevating the model's mAP.

3.2.3. SimC2f Design

A significant improvement of YOLOv8 is the substitution of the original C3 module with the C2f module. The C2f module aims to augment the model's feature extraction capabilities without adding to the model's complexity, thus further enhancing its overall performance. The ELAN principle inspires the design of the C2f module. Typically, as the network reaches a certain depth, the improvement in accuracy by adding more convolutional blocks diminishes, and the model's convergence deteriorates. The ELAN module enhances performance by increasing the longest gradient path within the residual blocks to address this issue. The ELAN module enhances model performance by analyzing the gradient paths, both short and long, in each layer. Especially if the network is very deep, the ELAN module enables better control over the gradient paths, facilitating improved feature learning and overall model performance. Based on this foundation, enhancements were applied to the C3 module to achieve a more substantial gradient flow while preserving a lightweight design. The ELAN and C2f structures are shown in Figures 6 and 7.

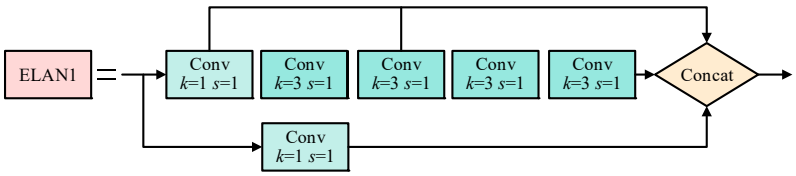

Figure 6. ELAN structure.

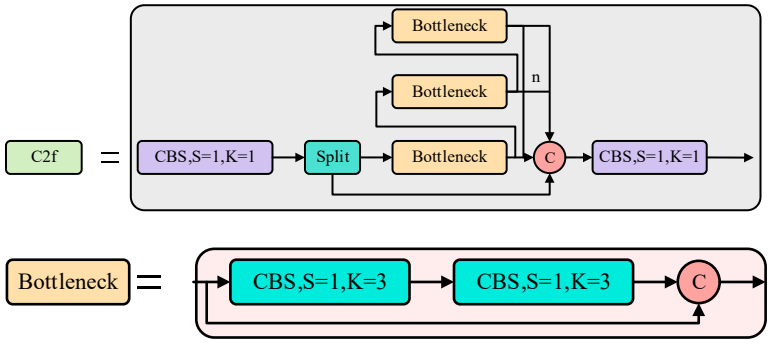

Figure 7. C2f structure.

The C2f module demonstrates relatively high accuracy in multi-object detection tasks. However, it exhibits some limitations. First, it employs fixed weights for feature fusion, which limits the adaptability to diverse targets and scenarios, resulting in inflexibility in weight assignment. Second, the simple feature summation operation may lead to information loss, particularly when addressing small targets or complex scenes, which may not adequately retain positional information and target details. In addition, for high-resolution feature maps, the C2f module requires many computations and parameters, increasing the model's complexity. To overcome these limitations in the C2f module, we designed the new SimC2f structure, which introduces the SimAM attention module [48] in C2f to assign unique weights to each neuron, thus providing a more flexible method of weight assignment. The SimC2f structure is shown in Figure 8.

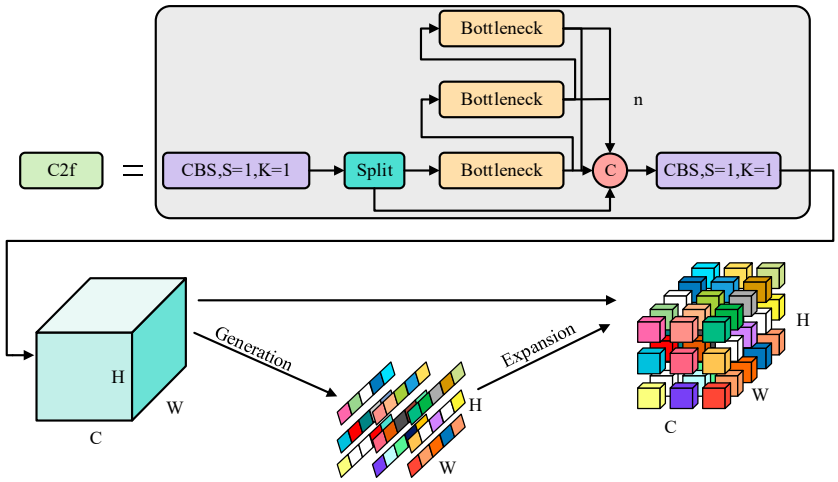

Figure 8. SimC2f structure.

The primary advantage of SimAM is its ability to allocate a unique weight to each neuron without introducing additional parameters. This capability enables the network to better adapt to various targets and scenarios, ultimately enhancing the model's performance and efficiency.

The SimAM assesses neuron importance using principles derived from neuroscience. In neuroscience, information-rich neurons typically display distinct activation patterns from

those of neighboring neurons, often inhibiting the surrounding neurons, a phenomenon known as the spatial inhibitory effects. Hence, neurons with spatial inhibitory effects are considered to have higher significance. The SimAM, drawing inspiration from spatial inhibition, assesses the significance of each neuron through an analysis of linear separability between the target neuron and its counterparts. This evaluation involves defining an energy function for each neuron, as depicted in Equation (2).

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1}\sum_{i=1}^{M-1}(-1-(w_t x_i + b_t))^2 + (1-(w_t t + b_t))^2 + \lambda w_t^2 \tag{2}$$

The above equation takes the following analytical form:

$$w_t = -\frac{2(t-\mu_t)}{(t-\mu_t)^2 + 2\sigma_t^2 + 2\lambda} \tag{3}$$

$$b_t = -\frac{1}{2}(t+\mu_t)w_t \tag{4}$$

Among these, Equations (5) and (6) can be noted as follows:

$$\hat{\mu} = \frac{1}{M}\sum_{i=1}^{M} x_i \tag{5}$$

$$\hat{\sigma}^2 = \frac{1}{M}\sum_{i=1}^{M}(x_i - \hat{\mu})^2 \tag{6}$$

Therefore, the minimum energy equation is derived from Equation (7).

$$e_t^* = \frac{4\left(\hat{\sigma}^2 + \lambda\right)}{(t-\hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \tag{7}$$

According to Equation (7), a lower energy value signifies a more pronounced distinction between neuron t and other neurons, indicating higher importance. This entire process can be represented by Equation (8).

$$\widetilde{X} = sigmoid\left(\frac{1}{E}\right) \odot X \tag{8}$$

Industrial settings often feature intricate backgrounds and diverse objects. To achieve precise PPE compliance detection, enhancing the model's feature perception and representation capabilities is imperative. Consequently, we incorporated SimAM into the C2f module, enhancing the model's capacity to perceive features in small targets while avoiding the introduction of extra parameters—this substantially improved detection performance and accuracy.

3.2.4. GhostConv

In industrial settings, achieving high precision and real-time capabilities is critical for PPE detection. Timely issue identification is essential to effectively reduce safety risks. Although various enhancements have substantially improved the accuracy and performance of PPE detection, challenges remain, particularly in regards to real-time operations and lightweight requirements. To tackle these issues, this study utilizes GhostConv [49] convolution to substitute for conventional convolutions in the network backbone.

The GhostConv convolution process involves two steps. Initially, it is employed in traditional convolution generation feature maps with fewer channels, using relatively small computational resources. Subsequently, based on these feature maps, a series of simple linear operations is applied to convolve the channel feature maps, resulting in the

acquisition of more feature maps. Ultimately, these two sets of feature maps are spliced to form the final feature map.

In Figure 9b, the input is a feature map, which undergoes an initial standard convolution operation to produce Y'. Here, $X \in R^{C \times H \times W}$, with C denoting the number of channels, H as the height, and W as the width. The $*$ symbolizes the convolution operation, and the f' represents the convolution filter for this layer.

$$Y' = X * f' \tag{9}$$

Next, the feature maps of each channel are used to generate the Ghost feature map Y_{ij} using the $\Phi_{i,j}$ operation.

$$Y_{ij} = \Phi_{i,j}(Y_{i'}) \tag{10}$$

Finally, the feature map identity is connected to yield the final feature map.

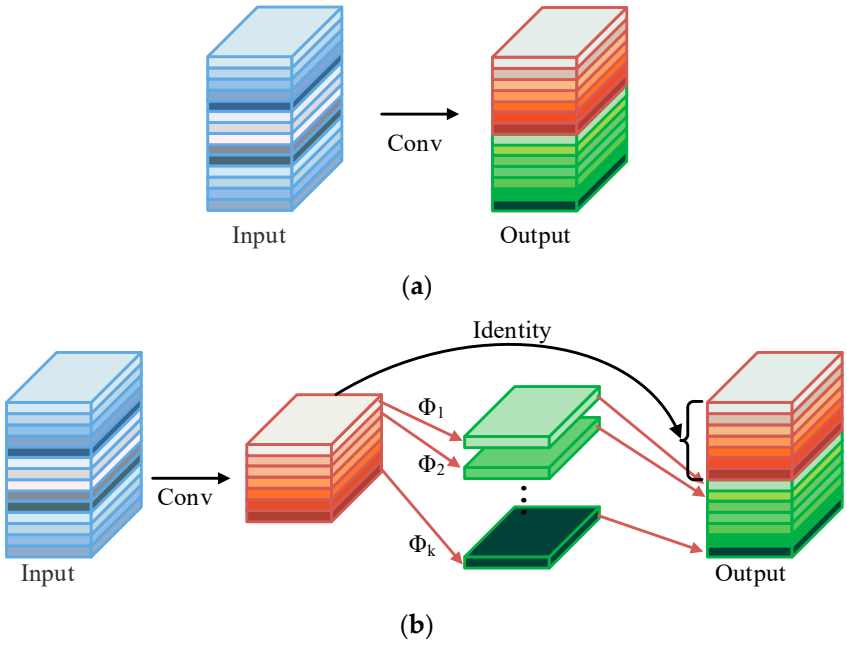

Figure 9. (**a**) Traditional convolution; (**b**) GhostConv module.

Utilizing GhostConv eliminates redundant information in feature map fusion, enhancing model performance and significantly improving the inference speed of the network model. In contrast to traditional convolution methods, GhostConv substantially reduces the cost of learning unnecessary features through cost-effective linear operations, achieving superior performance using the same computational resources.

We replace the traditional convolution of the original model with GhostConv. Although the accuracy decreases slightly, the model is more lightweight, which helps us detect problems and reduce security risks more quickly.

4. Experimental and Results

4.1. Experimental Datasets

In an industrial environment, the timely and precise detection of workers employing PPE is essential to ensure their safety. This allows us to take necessary measures promptly to mitigate potential risks. However, as of now, no publicly accessible dataset encompasses the many types of PPE. Therefore, this research focuses on detecting the usage of PPE by workers in an industrial environment. We assembled a dataset comprising four PPE categories—safety helmets, safety goggles, safety vests, and masks—by gathering pertinent images. This dataset primarily originates from two sources. The first source consists of close-up video images captured by temporarily deployed cameras. The second source includes

wide-angle video images captured by surveillance cameras in various factory workshops. These HIKVISION brand surveillance cameras all feature a 4-megapixel resolution. Images are sampled at a frequency of one frame extracted every 5 s, and the images collected from multiple distinct industrial settings. A sampling example is shown in Figure 10.

Figure 10. Sampling example.

Upon data collection, we conducted data cleaning and filtering, employing data augmentation techniques. These techniques encompassed independent object cropping, horizontal flipping, exposure adjustment, and the introduction of Gaussian noise. This entire process generated 4000 images, comprising both original and augmented variants. Subsequently, we partitioned these samples into training, validation, and testing sets at an 8:1:1 ratio, which accounted for 3200, 400, and 400 images, respectively. Before training, we utilized the LabelImg annotation software to generate text files containing image paths, annotated regions, and label types. The number of specific labeled categories and the distribution of x, y coordinates of the center point of the target box are shown in Figure 11.

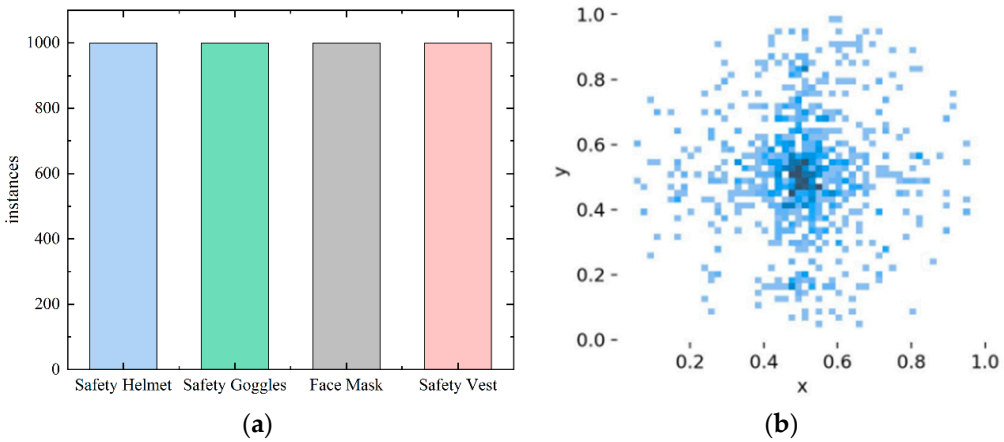

Figure 11. (**a**) PPE category diagram; (**b**) distribution plot of x and y coordinates.

4.2. Experimental Environments

All experiments in this study were performed on a Linux-based computer equipped with a 12th Gen Intel(R) Core(TM) i9-12900 K CPU and an NVIDIA GeForce RTX 4090 GPU boasting 32 GB of VRAM. The software environment was based on Python 3.8, utilizing Pytorch 2.0 as the development framework. The batch size and the number of epochs were set to 32 and 300, respectively.

4.3. Evaluation Metrics

We selected metrics such as precision, recall, F1 score, and mAP to assess the detection model's performance. The specific calculation formulas are as follows:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{11}$$

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

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

$$AP = \int_0^1 P(r)dr \tag{14}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{15}$$

TP, *FP*, *FN*, and *TN* stand for true positives, false positives, false negatives, and true negatives, respectively. *AP* signifies the average detection precision for individual defect categories, whereas *mAP* signifies the average detection precision across all defect categories.

To assess the model's lightweight characteristics, we utilized four key evaluation metrics: parameters, FLOPS, weight, and inference time. The quantity of the model's parameters directly impacts its complexity, with higher parameter counts leading to increased model complexity and greater demands on computational resources for training and inference. FLOPS quantifies the floating-point operations performed during the model's inference, and higher FLOPS values typically signify a heightened need for computational resources during inference. Weight measures the size of the model's weights, which is intimately tied to storage and transmission efficiency. Smaller weight values indicate lighter model weights, contributing to the enhanced light weight of the model. Inference time denotes the duration required for the model to process a single image, and shorter inference times are paramount for real-time applications, reflecting the model's ability to expedite object detection tasks.

4.4. Experimental Results and Analysis

4.4.1. Performance Analysis of the GBSG-YOLOv8n Model

We trained the YOLOv8n and GBSG-YOLOv8n models separately on the same training dataset and obtained the following experimental results, as shown in Tables 2 and 3.

Table 2. Comparison of model accuracy.

Model	Precision (%)	Recall (%)	F1 (%)	mAP (%)
YOLOv8n	86.6	88.4	87.5	87.8
GBSG-YOLOv8n	89.7	91.0	90.3	90.8

Table 3. Comparison of model complexity.

Model	Parameters (M)	FLOPS (G)	Weight (MB)	Inference (ms)
YOLOv8n	3.01	8.1	5.92	92.7
GBSG-YOLOv8n	2.51	7.0	4.66	88.7

The table above shows that the traditional YOLOv8n model achieved a mAP of 87.8%, while the GBSG-YOLOv8n model reached a mAP of 90.8%, representing a 3% improvement. Furthermore, compared to the traditional YOLOv8n, the GBSG-YOLOv8n model showed a 3.1% improvement in precision, a 2.6% increase in recall, and a 2.8% boost in the F1

score. These results indicate that our proposed GBSG-YOLOv8n model shows a distinct advantage, particularly in scenarios with complex backgrounds and significant variations in target sizes.

For a more intuitive representation of the detection performance, we have plotted the precision-recall curves for YOLOv8n and GBSG-YOLOv8n, which are presented in Figure 12. From these graphs, it is evident that the GBSG-YOLOv8n model significantly outperforms the traditional YOLOv8n model.

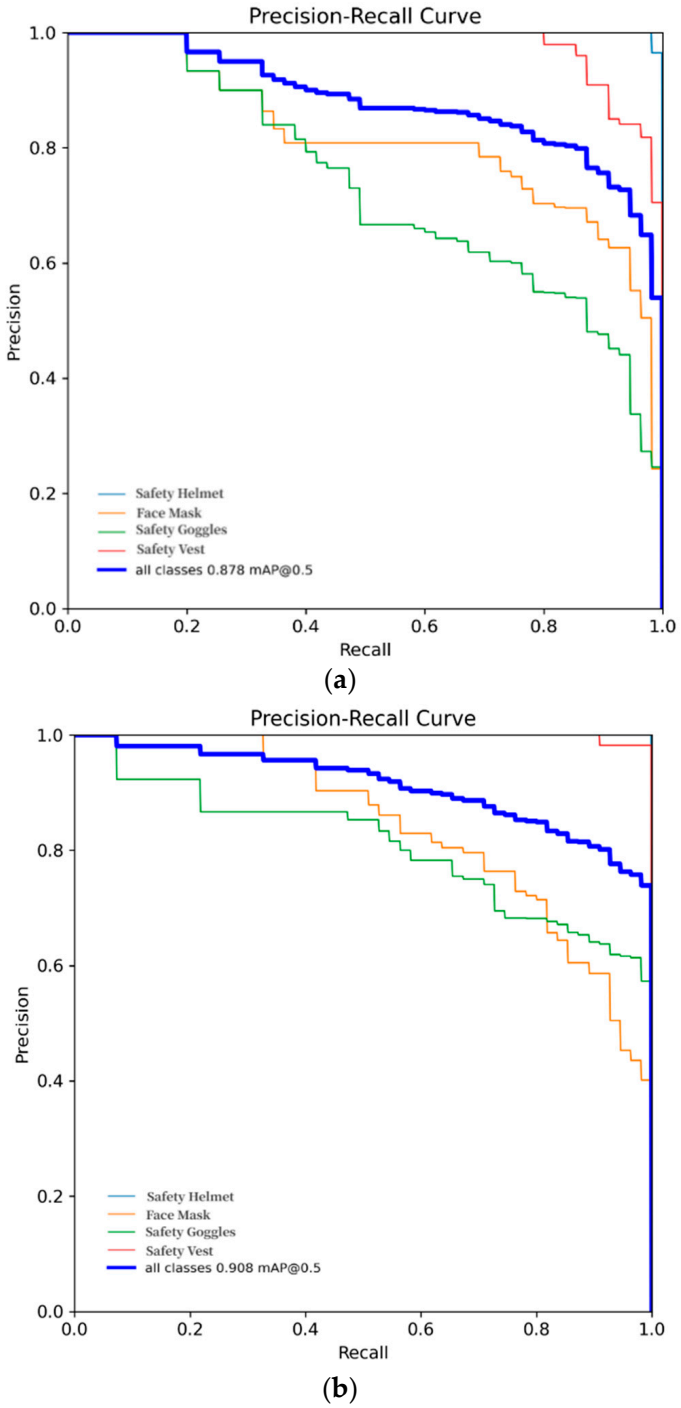

Figure 12. (**a**) YOLOv8n P-R curve; (**b**) GBSG-YOLOv8n P-R curve.

Table 3 reveals significant improvements in GBSG-YOLOv8 compared to the original YOLOv8n model, as evident in four critical metrics: parameters, FLOPS, weight, and

inference time. These results affirm GBSG-YOLOv8's lightweight nature in regards to PPE detection when contrasted with YOLOv8n. Specifically, GBSG-YOLOv8n boasts fewer parameters, reduced memory consumption, and faster computational operations. Furthermore, its more compact model size minimizes storage space requirements. These characteristics establish GBSG-YOLOv8n as a practical and efficient choice, particularly well-suited for PPE detection in industrial settings.

4.4.2. Ablation Experiment

We conducted ablation experiments to thoroughly validate the enhanced algorithm's effectiveness in optimizing the original method. Each experiment set was trained and validated using the identical PPES dataset, and the results are presented in Table 4.

Table 4. Ablation experiment result.

Model	Precision (%)	Recall (%)	F1 (%)	mAP (%)	Parameters (M)	FLOPS (G)	Weight (MB)	Inference Time (ms)
YOLOv8n	86.6	88.4	87.5	87.8	3.01	8.1	5.92	92.7
YOLOv8n+G	87.4	89.6	88.5	89.2	3.38	8.7	5.93	101.2
YOLOv8n+G+B	88.6	90.7	89.6	90.4	3.38	8.7	5.93	108.5
YOLOv8n+G+B+S	90.3	91.2	90.7	91.1	3.38	8.7	5.94	114.7
GBSG-YOLOv8n	89.7	91.0	90.3	90.8	2.51	7.0	4.66	88.7

The data in Table 4 demonstrates that introducing GAM into the backbone network has enhanced the capability to extract essential information, resulting in a 1.4% improvement in mAP compared to the original YOLOv8n. Replacing the PANet structure in the Neck network with the BiFPN structure has improved the model's feature learning ability, effectively integrating feature information across multiple scales, leading to a 2.6% improvement in mAP compared to the original YOLOv8n. The utilization of the SimC2f structure has effectively achieved adaptive feature fusion, enhancing the extraction of valuable information and resulting in a 3.3% improvement in mAP. While replacing traditional convolutions in the backbone network with GhostConv has led to a slight decrease in accuracy, the overall accuracy has significantly improved compared to that of the original model. Additionally, four key metrics, namely parameters, FLOPs, weight, and inference time, have all exhibited substantial reductions, signifying a more streamlined and lightweight model that boosts computational efficiency and reduces resource requirements. Experimental results confirm that each improvement introduced in this study has enhanced performance compared to that of the original YOLOv8n model. For a more intuitive presentation of these findings, we have created bar charts illustrating the results, which are shown in Figure 13.

Through these bar charts, it can be noted that the detection accuracy has significantly improved, and the complexity has noticeably decreased. This further validates the outstanding performance of the proposed GBSG-YOLOv8n model for real-time PPE detection.

4.4.3. Experiments Comparing GBSG-YOLOv8n to Other Models

To conduct a more in-depth evaluation of the effectiveness of the GBSG-YOLOv8n model, we performed comparative experiments using mainstream algorithms, and Table 5 presents the results regarding PPE detection performance.

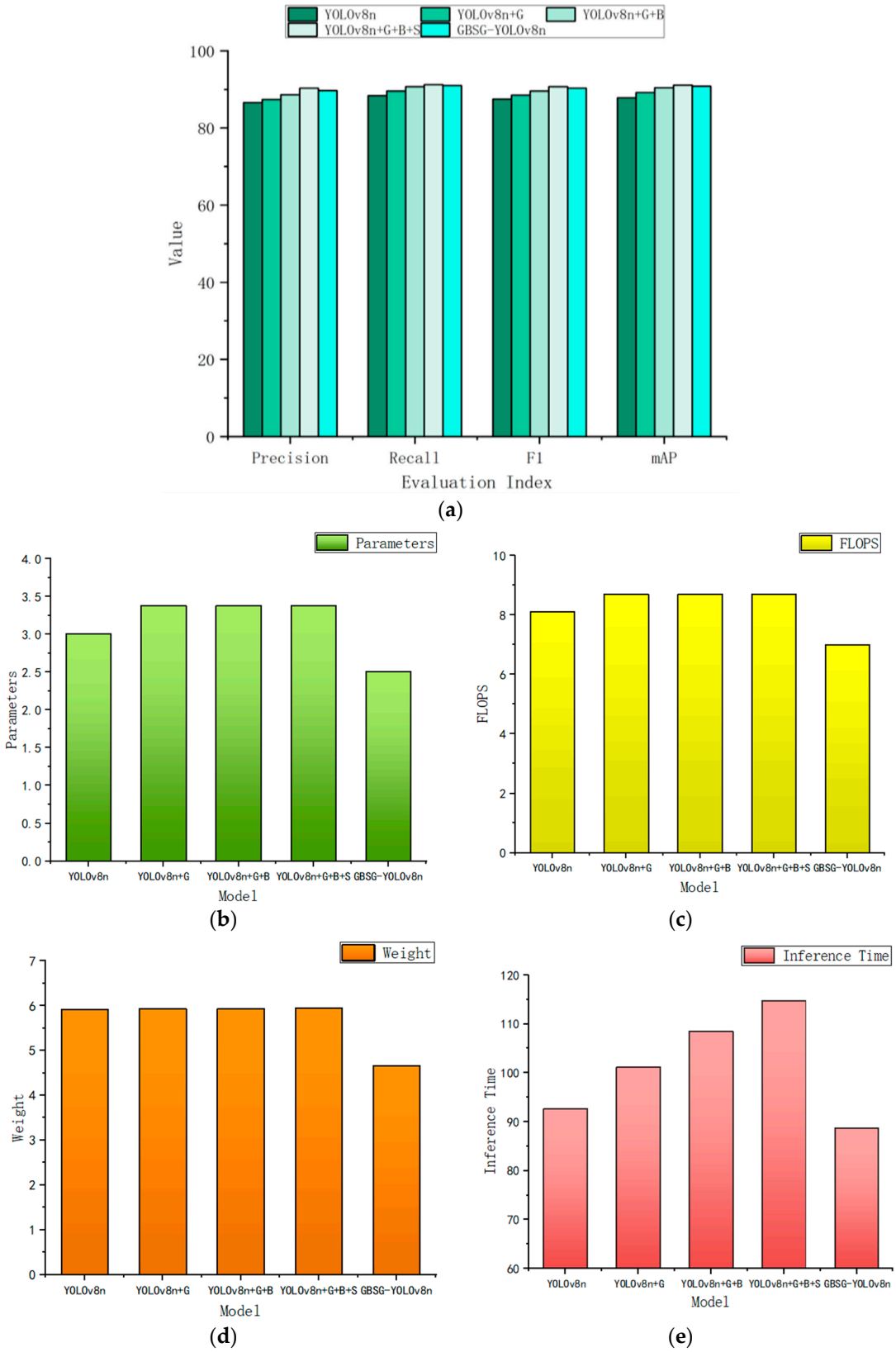

Figure 13. Ablation experiment result: (**a**) bar chart for accuracy; (**b**) bar chart for parameters; (**c**) bar chart for FLOPS; (**d**) bar chart for weight; (**e**) bar chart for inference time.

Table 5. Comparative experiment results.

Model	Precision (%)	Recall (%)	F1 (%)	mAP (%)	Parameters (M)	FLOPS (G)	Weight (MB)
Faster R-CNN	80.4	84.7	82.5	82.7	137.38	107.54	348.65
SSD	81.6	85.3	83.4	83.4	26.23	27.83	114.75
YOLOv3	83.9	86.7	85.3	84.9	60.53	35.78	59.62
YOLOv5	86.2	88.1	87.1	86.4	7.12	79.42	15.91
YOLOv7	87.1	89.7	88.4	87.3	36.52	98.34	103.24
GBSG-YOLOv8n	89.7	91.0	90.3	90.8	2.51	7.0	4.66

The analysis of these results shows that as the model versions are upgraded, the detection performance gradually improves for the parameters of accuracy, recall, F1 score, and mAP. And there are different variations in regards to the complexity aspect of the model. SSD adopts end-to-end training and multi-scale detection, showing significant improvements in performance in various aspects compared to Faster R-CNN. The YOLO series has introduced strategies such as single-stage detection, multi-scale detection, and the prediction of multiple bounding boxes, all while reducing model complexity. YOLO also offers considerable performance improvements over Faster R-CNN and SSD. YOLOv5 incorporated CSPDarkNet53 as its backbone, along with the focus module and PANet structure, enhancing detection accuracy and performance. YOLOv7 introduced E-ELAN and a deeper network structure, achieving greater speed and accuracy, but increasing model computation and parameters due to the more complex network. In contrast, our proposed model, GBSG-YOLOv8n, builds upon YOLOv8n and significantly enhances performance by introducing GAM into the backbone network, optimizing the PANet structure, utilizing the SimC2f structure, and incorporating GhostConv. Compared to other mainstream models, GBSG-YOLOv8n excels in metrics like precision, recall, F1, and mAP, surpassing those of other models. Additionally, it significantly outperforms other models in terms of parameters, FLOPS, and weight, demonstrating its lightweight nature. This reaffirms GBSG-YOLOv8n's ability to provide reliable support for PPE detection in industrial environments. To present these results more clearly, we have created a series of model comparison charts (refer to Figure 14).

4.4.4. Practical Applications of GBSG-YOLOv8n in Industrial Environments

To implement the practical application of the GBSG-YOLOv8n model in an industrial environment, we deployed the model on the cloud server of the PPE safety monitoring system. We designed and constructed the PPE safety monitoring system using a microservices architecture, with its primary task being the monitoring of the wearing of PPE by workers in industrial settings. In the factory, we deployed multiple network cameras positioned at various locations to continuously monitor the PPE usage of workers in real-time. For instance, at various entrance points within the factory (as shown in Figure 15a, where red boxes mark the different entrance points), these network cameras transmit captured real-time images of workers over the network to the cloud server. Subsequently, the GBSG-YOLOv8n model deployed on the cloud server is utilized for detection. If a worker's PPE complies with the requirements (as shown in Figure 15c), the system display on the large screen will record the worker's information, obtained through the access control system, including the worker's entry time, indicating the worker's smooth access to the factory. However, if a worker's PPE does not meet the requirements (as shown in Figure 15d), the system display on the large screen will record the worker's information, obtained through the access control system, along with the entry time. Meanwhile, it will also trigger an alarm, and the recorded violation of the worker will be logged into the system. The entire system deployment is illustrated in Figure 15b. Practical applications in industrial settings demonstrate that the use of the GBSG-YOLOv8n model has a positive impact on enhancing industrial safety.

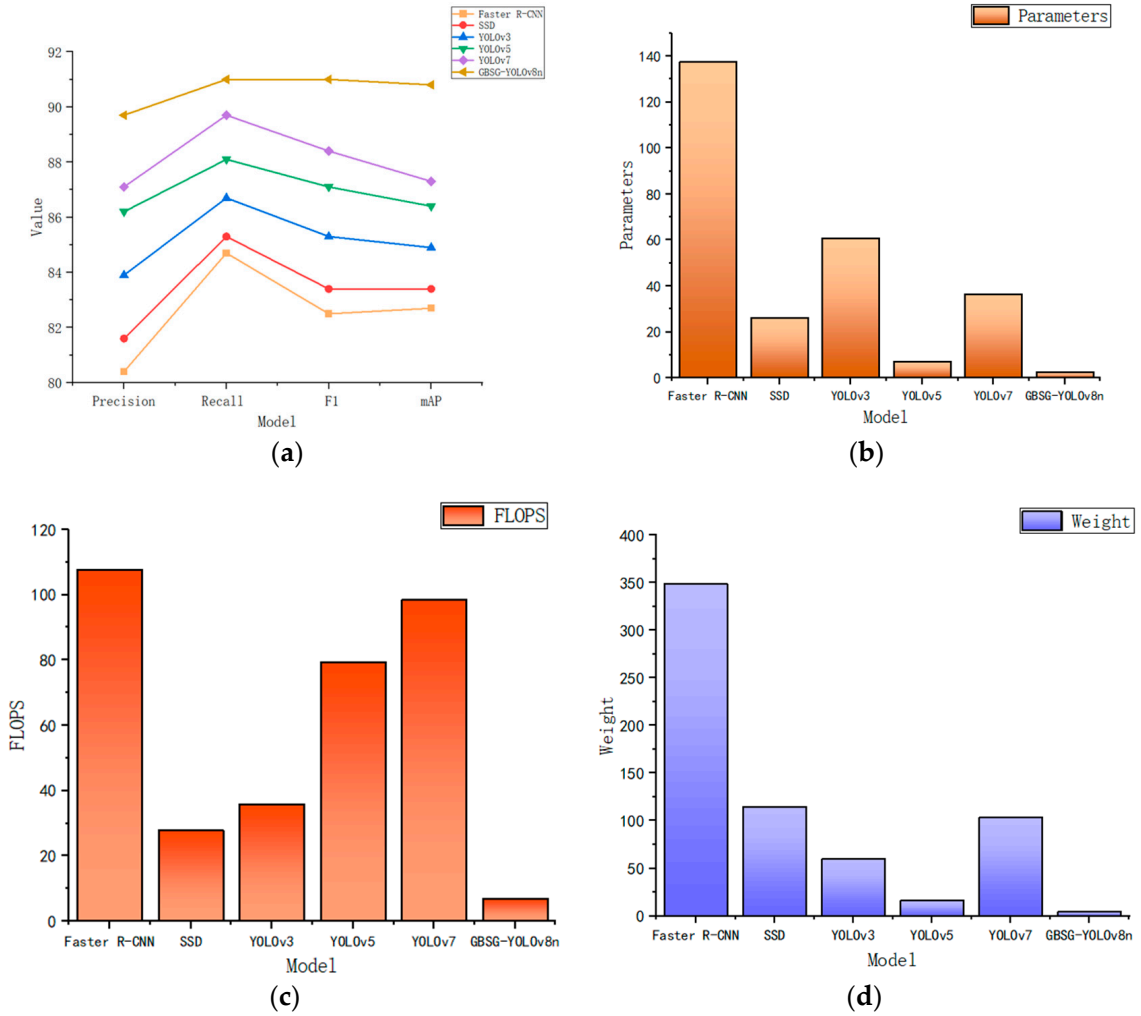

Figure 14. Comparative experimental results: (**a**) line graph showing accuracy; (**b**) bar chart for parameters; (**c**) bar chart for flops; (**d**) bar chart for weight.

While our model has been widely applied in industrial environments and has performed excellently in most cases, we inevitably faced certain limitations that have emerged through multiple rounds of experimental testing and practical applications. We have observed that, in specific situations, angle problems caused by the position of the worker and the position of the image acquisition equipment may cause our model to erroneously identify eyeglasses as safety goggles.

Figure 16 represents a series of real-time actions of workers in the detection field. The results for Figure 16a,b,d are correct; however, as shown in Figure 16c, when a worker wearing an orange safety vest turns his/her whole body to the other side, when observed from a side angle, our model will erroneously classify the ordinary glasses worn by the middle worker as safety goggles. Although eyeglasses and safety goggles may seem very similar in terms of appearance, they exhibit significant differences in terms of purpose and nature. To address this limitation, in the future, we will focus on expanding the scenarios in the PPES dataset to enhance the model's generalization capability.

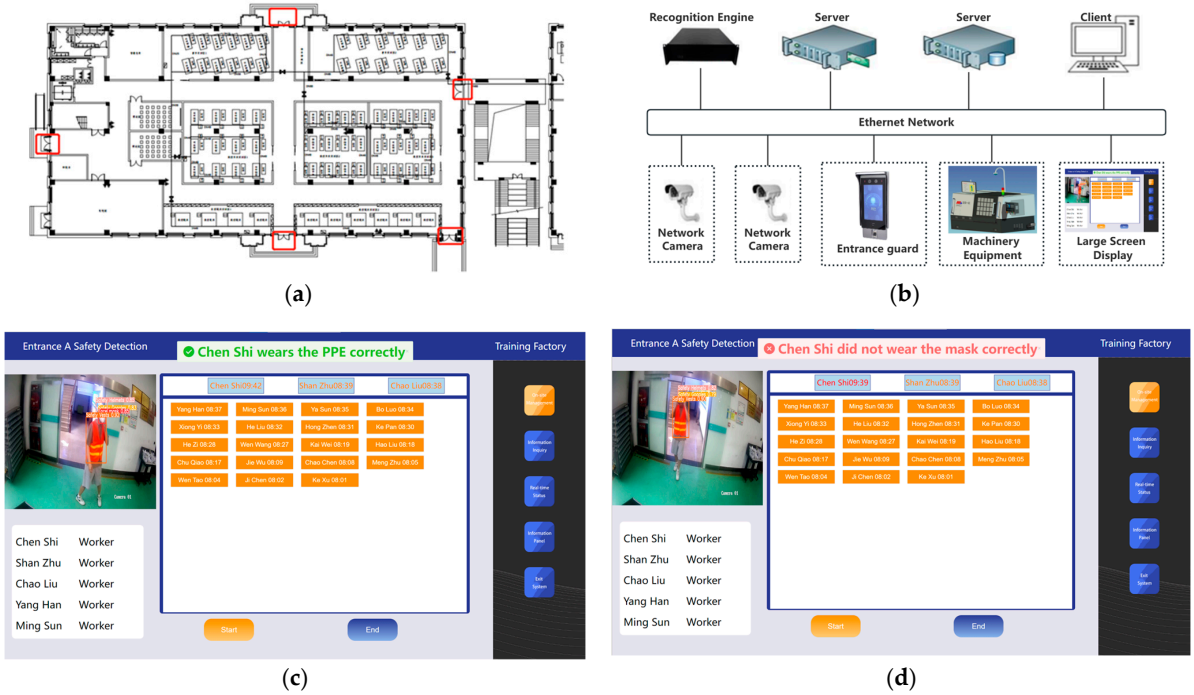

Figure 15. (**a**) Factory layout; (**b**) system deployment diagram; (**c**) system display interface for workers wearing PPE correctly; (**d**) system display interface for workers wearing PPE incorrectly.

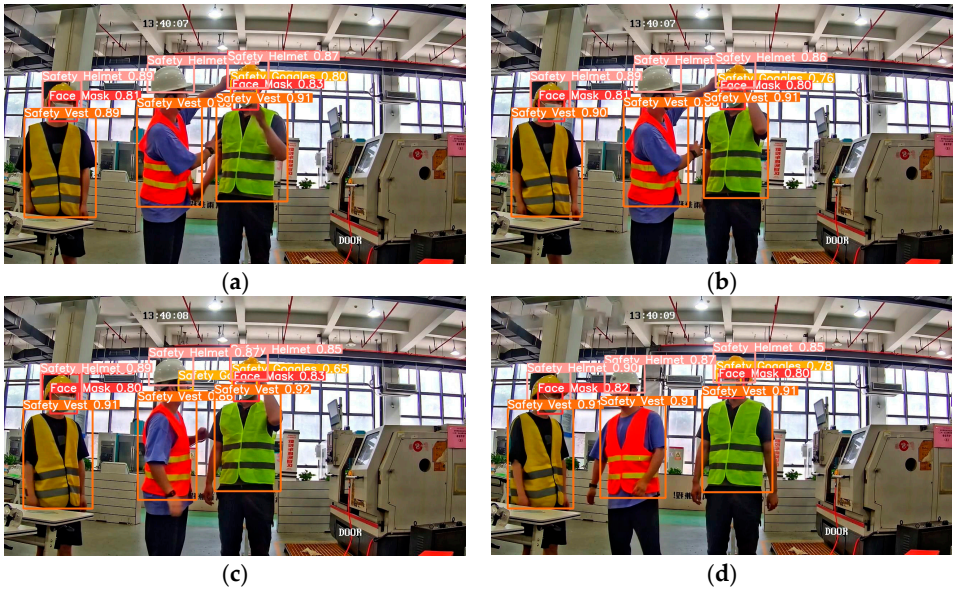

Figure 16. (**a**–**d**) are real-time monitoring images of the site.

5. Discussion

Compared to other visual detection tasks, detecting whether industrial workers are wearing PPE presents a unique set of challenges. First, the diverse types of PPE that require detection, with substantial differences in size, render the target detection task quite challenging. Moreover, within the complex and ever-changing industrial environment, various potential sources of interference further increase the complexity of detection. If workers fail to wear PPE correctly and this is not detected promptly, it may lead to severe safety issues, even tragic consequences. Hence, the accurate and timely detection of PPE compliance is vital to ensure worker safety and mitigate potential industrial risks. As a replacement for traditional manual inspections, computer vision technology has proven

to be an efficient solution. Computer vision technology offers automated and efficient detection methods. Through computer vision technology, we can accurately identify potential issues in the early stages, allowing for early warnings to mitigate potential dangers. In addition, this approach significantly reduces false alarms and omissions, enhancing detection accuracy and reliability.

In this study, we selected the YOLOv8n model from YOLOv8 due to its reduced parameter size and heightened detection accuracy. However, recognizing the task's complexity and specific YOLOv8n limitations, we proposed the GBSG-YOLOv8n model to detect whether industrial workers correctly wear PPE in an industrial environment. By refining the backbone and Neck networks, we achieved a 3% enhancement in detection performance, and the model is more lightweight.

Simultaneously, we also conducted the following research experiments on the same PPES dataset to compare the performance of the GBSG-YOLOv8n model with other transformer based models and models with exceptional performance. The specific results are shown in the Table 6 below.

Table 6. Comparison of performance results.

Model	Precision (%)	Recall (%)	F1 (%)	mAP (%)
RetinaNet	85.6	86.4	86.0	85.1
ComerNet	87.5	88.1	87.8	86.4
DETR	88.7	89.3	89.0	88.6
DINO	89.3	90.4	89.8	89.3
GBSG-YOLOv8n	89.7	91.0	90.3	90.8

These results once again highlight the outstanding performance of our proposed GBSG-YOLOv8n model in regards to target detection. The model's exceptional performance equips it with various potential applications in industrial production environments. The GBSG-YOLOv8n model enables us to quickly and accurately detect whether workers are correctly wearing PPE, effectively reducing workplace safety risks.

Furthermore, our research holds significant value, not only in ensuring worker safety in industrial environments, but also in demonstrating extensive applicability across various potential domains. In the medical field, our PPE detection technology can ensure that healthcare workers and patients correctly wear appropriate PPE, such as masks and protective gowns, especially when dealing with infectious patients. It is worth emphasizing that healthcare-associated infections are a severe concern in medical facilities, endowing our research with significant potential for reducing infection risks. Police and traffic management personnel in the field of transportation can also benefit from our PPE detection technology. This technology helps ensure their use of safety vests, helmets, and other necessary PPE, reducing the risk of road traffic accidents. This is crucial for improving traffic safety and reducing accident rates. In the military and emergency services sectors, PPE detection technology also plays a vital role in ensuring that soldiers and rescue personnel wear PPE correctly, effectively guaranteeing their safety. Furthermore, our research can be applied to environmental monitoring to ensure that researchers and workers wear appropriate PPE when handling hazardous substances or working in contaminated environments, thus reducing environmental pollution and occupational risks.

Therefore, our PPE detection technology shows broad prospects for application in various fields. It not only enhances workplace and specific environment safety, but also helps to reduce the risk of accidents. This is crucial for individual protection and improves societal safety and health.

6. Conclusions

In this study, we propose a new and improved PPE detection model, called GBSG-YOLOv8n, and construct a dedicated PPES dataset to better meet the challenges of PPE detection in industrial environments. First, we overcome the limitations in extracting PPE

target features by introducing the GAM, which maximizes the retention of channel and spatial information, enhances cross-dimensional interactions, and significantly improves the feature extraction capabilities in the backbone network, notably enhancing detection performance. Second, we optimize the fusion of multi-scale target information by replacing the original PANet structure with the BiFPN structure, effectively integrating feature information from different scales, preventing the loss of PPE feature information, and improving detection accuracy. Finally, the SimAM attention mechanism is introduced into the C2f module, and the SimC2f structure is proposed. This enhancement enables the model to more efficiently process image features, resulting in a notable improvement in detection efficiency. Finally, GhostConv is used to replace the traditional convolution in the backbone network so that the model reduces the model complexity and makes the model more lightweight, while ensuring detection accuracy.

The experimental results unequivocally illustrate the exceptional performance of the PPE detection model proposed in this study. In comparison to mainstream models, it offers substantial advantages. This model not only satisfies the demands for real-time safety monitoring in industrial settings, but also imparts significant value in safeguarding workers and mitigating potential industrial hazards. In future research, we plan to continue to enhance the model's performance, making it applicable to more complex and diverse scenarios, expanding its utility to broader domains, and smoothly deploying the model into multiple systems.

Author Contributions: Data curation, C.S. and Y.Z.; formal analysis, C.S. and J.S.; funding acquisition, C.Z.; software, C.S. and D.Z.; supervision, D.Z. and C.Z.; validation, C.Z., J.S. and Y.Z.; writing—review and editing, C.S. and D.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China (Nos. 62272418, 62102058) and the basic public welfare research program of Zhejiang Province (No. LGG18E050011).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used in this study is available on demand.

Acknowledgments: We sincerely thank Jin Hu and SANG-WOON JEON of Hanyang University for their guidance.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kerr, W.A. Accident Proneness of Factory Departments. *J. Appl. Psychol.* **1950**, *34*, 167. [CrossRef] [PubMed]
2. Fall Protection—Overview. Occupational Safety and Health Administration. Available online: https://www.osha.gov/fall-protection (accessed on 16 September 2023).
3. Chughtai, A.A.; Khan, W. Use of Personal Protective Equipment to Protect against Respiratory Infections in Pakistan: A Systematic Review. *J. Infect. Public Health* **2020**, *13*, 385–390. [CrossRef] [PubMed]
4. Hulme, A.; Gilchrist, A. Industrial Head Injuries And The Performance Of Helmets. In Proceedings of the 1995 International IRCOBI Conference on the Biomechanics of Impact, Brunnen, Switzerland, 13–15 September 1995.
5. De la Hunty, D.; Sprivulis, P. Safety Goggles Should Be Worn by Australian Workers. *Aust. N. Z. J. Ophthalmol.* **1994**, *22*, 49–52. [CrossRef] [PubMed]
6. Arditi, D.; Ayrancioglu, M.A.; Shi, J. Effectiveness of safety vests in nighttime highway construction. *J. Transp. Eng.* **2004**, *130*, 725–732. [CrossRef]
7. Kyung, S.Y.; Jeong, S.H. Particulate-Matter Related Respiratory Diseases. *Tuberc. Respir. Dis.* **2020**, *83*, 116. [CrossRef]
8. Hung, H.M.; Lan, L.T.; Hong, H.S. A Deep Learning-Based Method For Real-Time Personal Protective Equipment Detection. *Le Quy Don Tech. Univ.-Sect. Inf. Commun. Technol.* **2019**, *13*, 23–34.
9. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
10. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
11. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.

12. Kuznetsova, A.; Maleva, T.; Soloviev, V. Detecting Apples in Orchards Using YOLOv3 and YOLOv5 in General and Close-up Images. In Proceedings of the Advances in Neural Networks–ISNN 2020: 17th International Symposium on Neural Networks, ISNN 2020, Cairo, Egypt, 4–6 December 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 233–243.
13. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. *arXiv* **2022**, arXiv:2209.02976.
14. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
15. Choi, J.; Chun, D.; Kim, H.; Lee, H.-J. Gaussian YOLOv3: An Accurate and Fast Object Detector Using Localization Uncertainty for Autonomous Driving. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27–28 October 2019; pp. 502–511.
16. Vats, A.; Anastasiu, D.C. Enhancing Retail Checkout through Video Inpainting, YOLOv8 Detection, and DeepSort Tracking. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 5530–5537.
17. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-Captured Scenarios. IEEE Conference Publication. IEEE Xplore. Available online: https://ieeexplore.ieee.org/document/9607487 (accessed on 28 September 2023).
18. Liu, K.; Sun, Q.; Sun, D.; Peng, L.; Yang, M.; Wang, N. Underwater Target Detection Based on Improved YOLOv7. *J. Mar. Sci. Eng.* **2023**, *11*, 677. [CrossRef]
19. Real-Time Growth Stage Detection Model for High Degree of Occultation Using DenseNet-Fused YOLOv4—ScienceDirect. Available online: https://www.sciencedirect.com/science/article/pii/S0168169922000114?via%3Dihub (accessed on 28 September 2023).
20. Kelm, A.; Laußat, L.; Meins-Becker, A.; Platz, D.; Khazaee, M.J.; Costin, A.M.; Helmus, M.; Teizer, J. Mobile Passive Radio Frequency Identification (RFID) Portal for Automated and Rapid Control of Personal Protective Equipment (PPE) on Construction Sites. *Autom. Constr.* **2013**, *36*, 38–52. [CrossRef]
21. Bauk, S.; Schmeink, A.; Colomer, J. An RFID Model for Improving Workers' Safety at the Seaport in Transitional Environment. *Transport* **2018**, *33*, 353–363. [CrossRef]
22. Dong, S.; Li, H.; Yin, Q. Building Information Modeling in Combination with Real Time Location Systems and Sensors for Safety Performance Enhancement. *Saf. Sci.* **2018**, *102*, 226–237. [CrossRef]
23. Hayward, S.; van Lopik, K.; West, A. A Holistic Approach to Health and Safety Monitoring: Framework and Technology Perspective. *Internet Things* **2022**, *20*, 100606. [CrossRef]
24. Jordan, M.I.; Mitchell, T.M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349*, 255–260. [CrossRef] [PubMed]
25. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
26. Li, J.; Liu, H.; Wang, T.; Jiang, M.; Wang, S.; Li, K.; Zhao, X. Safety Helmet Wearing Detection Based on Image Processing and Machine Learning. In Proceedings of the 2017 Ninth International Conference on Advanced Computational Intelligence (ICACI), Doha, Qatar, 4–6 February 2017; pp. 201–205.
27. Wu, H.; Zhao, J. An Intelligent Vision-Based Approach for Helmet Identification for Work Safety. *Comput. Ind.* **2018**, *100*, 267–277. [CrossRef]
28. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
29. Zhang, J.; Zhang, D.; Liu, X.; Liu, R.; Zhong, G. A framework of on-site construction safety management using computer vision and real-time location system. In Proceedings of the International Conference on Smart Infrastructure and Construction 2019 (ICSIC) Driving Data-Informed Decision-Making, Cambridge, UK, 8–10 July 2019; pp. 327–333.
30. 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*, e311. [CrossRef]
31. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2016; Volume 9905, pp. 21–37, ISBN 978-3-319-46447-3.
32. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
33. Han, G.; Zhu, M.; Zhao, X.; Gao, H. Method Based on the Cross-Layer Attention Mechanism and Multiscale Perception for Safety Helmet-Wearing Detection. *Comput. Electr. Eng.* **2021**, *95*, 107458. [CrossRef]
34. Wang, J.; Zhu, G.; Wu, S.; Luo, C. Worker's Helmet Recognition and Identity Recognition Based on Deep Learning. *Open J. Model. Simul.* **2021**, *9*, 135–145. [CrossRef]
35. Jiang, X.; Gao, T.; Zhu, Z.; Zhao, Y. Real-Time Face Mask Detection Method Based on YOLOv3. *Electronics* **2021**, *10*, 837. [CrossRef]
36. Ji, X.; Gong, F.; Yuan, X.; Wang, N. A High-Performance Framework for Personal Protective Equipment Detection on the Offshore Drilling Platform. *Complex Intell. Syst.* **2023**, *9*, 5637–5652. [CrossRef]

37. Wang, Z.; Wu, Y.; Yang, L.; Thirunavukkarasu, A.; Evison, C.; Zhao, Y. Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. *Sensors* **2021**, *21*, 3478. [CrossRef] [PubMed]
38. Zhang, Y.; Qiu, Y.; Bai, H. FEFD-YOLOV5: A Helmet Detection Algorithm Combined with Feature Enhancement and Feature Denoising. *Electronics* **2023**, *12*, 2902. [CrossRef]
39. Tai, W.; Wang, Z.; Li, W.; Cheng, J.; Hong, X. DAAM-YOLOV5: A Helmet Detection Algorithm Combined with Dynamic Anchor Box and Attention Mechanism. *Electronics* **2023**, *12*, 2094. [CrossRef]
40. Sun, C.; Zhang, S.; Qu, P.; Wu, X.; Feng, P.; Tao, Z.; Zhang, J.; Wang, Y. MCA-YOLOV5-Light: A Faster, Stronger and Lighter Algorithm for Helmet-Wearing Detection. *Appl. Sci.* **2022**, *12*, 9697. [CrossRef]
41. Ali, L.; Alnajjar, F.; Parambil, M.M.A.; Younes, M.I.; Abdelhalim, Z.I.; Aljassmi, H. Development of YOLOv5-Based Real-Time Smart Monitoring System for Increasing Lab Safety Awareness in Educational Institutions. *Sensors* **2022**, *22*, 8820. [CrossRef]
42. Wang, J.; Wang, J.; Zhang, X.; Yu, N. A Mask-Wearing Detection Model in Complex Scenarios Based on YOLOv7-CPCSDSA. *Electronics* **2023**, *12*, 3128. [CrossRef]
43. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
44. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
45. Liu, Y.; Shao, Z.; Hoffmann, N. Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. *arXiv* **2021**, arXiv:2112.05561.
46. Zhou, S.; Zhao, Y.; Guo, D. YOLOv5-GE Vehicle Detection Algorithm Integrating Global Attention Mechanism. In Proceedings of the 2022 3rd International Conference on Information Science, Parallel and Distributed Systems (ISPDS), Guangzhou, China, 22–24 July 2022; pp. 439–444.
47. Zhang, C.; Tian, Z.; Song, J.; Zheng, Y.; Xu, B. Construction Worker Hardhat-Wearing Detection Based on an Improved BiFPN. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; IEEE Computer Soc: Los Alamitos, CA, USA, 2021; pp. 8600–8607.
48. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. Simam: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual, 24 July 2021; pp. 11863–11874.
49. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More Features from Cheap Operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1580–1589.