applied sciences

MDPI

Article

MEAG-YOLO: A Novel Approach for the Accurate Detection of Personal Protective Equipment in Substations

Hong Zhang [1], Chunyang Mu [2,3,*], Xing Ma [1,3], Xin Guo [1] and Chong Hu [1]

[1] College of Electrical and Information Engineering, North Minzu University, Yinchuan 750021, China; 20227419@stu.nmu.edu.cn (H.Z.); maxing@nmu.edu.cn (X.M.); 20227407@stu.nmu.edu.cn (X.G.); 20217312@stu.nmu.edu.cn (C.H.)
[2] College of Mechatronic Engineering, North Minzu University, Yinchuan 750021, China
[3] Ningxia Provincial Key Laboratory of Intelligent Information and Big Data Processing, North Minzu University, Yinchuan 750021, China
* Correspondence: muchunyang@nmu.edu.cn

Abstract: Timely and accurately detecting personal protective equipment (PPE) usage among workers is essential for substation safety management. However, traditional algorithms encounter difficulties in substations due to issues such as varying target scales, intricate backgrounds, and many model parameters. Therefore, this paper proposes MEAG-YOLO, an enhanced PPE detection model for substations built upon YOLOv8n. First, the model incorporates the Multi-Scale Channel Attention (MSCA) module to improve feature extraction. Second, it newly designs the EC2f structure with one-dimensional convolution to enhance feature fusion efficiency. Additionally, the study optimizes the Path Aggregation Network (PANet) structure to improve feature learning and the fusion of multi-scale targets. Finally, the GhostConv module is integrated to optimize convolution operations and reduce computational complexity. The experimental results show that MEAG-YOLO achieves a 2.4% increase in precision compared to YOLOv8n, with a 7.3% reduction in FLOPs. These findings suggest that MEAG-YOLO is effective in identifying PPE in complex substation scenarios, contributing to the development of smart grid systems.

Keywords: PPE detection; substation safety management; feature fusion efficiency; YOLOv8n; EC2f

Citation: Zhang, H.; Mu, C.; Ma, X.; Guo, X.; Hu, C. MEAG-YOLO: A Novel Approach for the Accurate Detection of Personal Protective Equipment in Substations. *Appl. Sci.* **2024**, *14*, 4766. https://doi.org/10.3390/app14114766

Academic Editor: Andrea Prati

Received: 8 April 2024
Revised: 21 May 2024
Accepted: 27 May 2024
Published: 31 May 2024

1. Introduction

The power energy sector is growing fast, with more large substations than ever. Keeping these substations running smoothly and safely is crucial for the power grid's stability [1]. Staff violations in operations significantly increase the likelihood of accidents [2]. Statistics show that from 2011 to 2021, 967 electrical workers in the United States died in accidents [3]. From 2020 to 2021, there were 70 power-related personal injury accidents in China, resulting in 82 deaths and over 1 million yuan in direct economic losses [4–6]. Substation workers' limited safety awareness jeopardizes both their safety and the reliability of business operations and urban energy provision. Therefore, timely detection of PPE is of utmost importance.

As the electric power industry develops, the importance of safety management in substations is increasingly emphasized. The correct use of PPE is fundamental to ensuring the safety of workers in the substation environment. However, despite the significant risk reduction provided by wearing PPE, challenges persist in PPE recognition and detection within substations. First, the complex and variable working conditions in substations can impair the effectiveness of PPE detection. Additionally, in emergencies, workers might overlook essential gear like insulated gloves or helmets, raising the risk of violations. Furthermore, current PPE detection technologies lack precision and speed, particularly in complex situations. Therefore, it is vital to develop effective PPE detection models that

quickly and accurately evaluate PPE use in diverse conditions, improving substation safety management.

Although PPE detection technology has significantly improved, detecting various scale targets in complex and dynamic substation environments remains challenging due to numerous model parameters. Therefore, this study proposes the lightweight MEAG-YOLO model for detecting PPE. The model incorporates several key innovations:

- This study integrated MSCA into the backbone network. This integration enhances the efficiency of target feature extraction. As a result, it improves the overall detection performance.
- The EC2f structure was developed to streamline the model. It reduces the number of parameters. This structure also enhances feature fusion efficiency. Furthermore, it simplifies calculations by using one-dimensional convolution. This is used for cross-channel interactions.
- The PAN was optimized. This bolstered the model's feature fusion capabilities. It ensures a more effective integration of multi-scale features.
- YOLOv8n was incorporated with GhostConv. This merger significantly reduces the amount of calculations. It has a slight impact on precision.

These innovations collectively make the MEAG-YOLO model effective for detecting various scales of PPE in substation environments.

2. Related Work

PPE significantly mitigates safety hazards for substation workers [7]. This protective equipment typically includes safety helmets, insulated gloves, and operating rods [8]. The algorithm for early detection of PPE mainly relied on sensor networks. Sensor technology detects PPE by installing sensors and analyzing signals. For instance, Barro-Torres et al. [9] employed RFID sensors to monitor how workers wear protective gear. However, this system is limited by its numerous hardware components and restricted coverage area. Kelm et al. [10] used scanners to identify RFID tags on PPE components, verifying the correct usage by workers. Dong et al. [11] used pressure sensors and Real-Time Location Systems (RTLS) to check if workers were following safety rules, like wearing helmets. Kang et al. [12] developed a drone-based system equipped with RFID detectors to identify tags on ground materials. Yet, this system's effectiveness is significantly hindered by the environment. It often results in delayed and imprecise feedback, and its deployment is complex. Zhang et al. [13] created an IoT-based helmet detection system incorporating infrared and thermal sensors. However, its utility is limited by incomplete internet coverage across construction sites. Moreover, Hayward et al. [14] developed an RFID-based system for PPE monitoring at workplace entrances, which notably lacks the capability for continuous surveillance. While sensor-based methods can identify PPE usage, their deployment and maintenance are expensive and complex. Furthermore, environmental factors and specific target categories restrict the accuracy of such detection methods. Consequently, real-time and efficient computer vision approaches have gained widespread adoption.

Early computer vision relies on machine learning technology. Machine learning enables computers to learn from data and make decisions and predictions. For example, Support Vector Machines (SVMs) effectively handle nonlinear features by finding the optimal hyperplane in high-dimensional space. Wu et al. [15] proposed a safety helmet recognition method combining color space and multilayer SVM. Similarly, Cai et al. [16] developed a multi-feature visual anti-occlusion framework. However, SVM is sensitive to lighting and background, limiting its detection capability in complex scenes. The K-Nearest Neighbors (KNNs) algorithm classifies based on the nearest neighbor labels. Wu et al. [17] used the KNN to identify moving objects in videos, then classified pedestrians and safety helmets with Convolutional Neural Network (CNN). However, KNN is easily affected by data noise. Bayesian algorithms use conditional probabilities for classification purposes. Chan et al. [18] developed a Bayesian Network (BN) model to identify and mitigate factors contributing to accident risks. Nonetheless, Bayesian methods struggle with complex or

nonlinear data relationships. K-means clustering is another approach, categorizing data into K-distinct clusters to minimize background noise and interference. Al-Bayati et al. [19] used fuzzy sets and K-means to evaluate factors of PPE non-compliance. However, K-means is limited in processing features like color and texture, which are unsuitable for complex scenes. Despite its widespread use, the combination of image processing and machine learning often requires manually extracting complex features. This can limit the analysis of large datasets and burden computational resources. Due to these limitations, especially in detecting PPE, deep learning methods in computer vision for object detection have become more popular recently.

Deep learning techniques are categorized into single-stage and two-stage methods. Two-stage methods incorporate RPN and merge candidate region generation with object detection. This blending improves both the speed and precision of detection. For example, Ahmed et al. [20] used a two-stage method with R-CNN for detecting PPE, achieving a mean average precision (mAP) of 96%. Bouhayane et al. [21] created a helmet detection system using Cascade R-CNN. Lee et al. [22] designed a safety helmet detection technique with Faster R-CNN that may overlook small targets. Though two-stage methods are more accurate, identifying and then classifying regions, they also demand more computational power, making them less efficient. In contrast, single-stage methods like SSD [23] and YOLO [24] have made significant progress in the field of object detection. They use CNN to create many bounding boxes and predict classes' probabilities at once. Meanwhile, this process makes detection and classification faster and more efficient. Zhao et al. [25] improved YOLOv3 for better performance in complex environments. However, this improvement made the network structure more complex and increased the number of parameters. Fang et al. [26] proposed an improved YOLOv4 model by integrating attention mechanisms with multi-scale features. However, it struggles to detect small objects. Ji et al. [27] improved the YOLOv4 model by integrating a residual feature enhancement network. This network preserves critical information in high-level feature maps, enhancing object detection accuracy. However, the model is limited in the range of categories it can detect. Lo et al. [28] trained and tested their custom PPE dataset using YOLOv3, YOLOv4, and YOLOv7 without any modifications to these networks. Qiao et al. [29] enhanced the YOLOv4 network by adding a global context module, which improved scene understanding and increased detection accuracy by 3.1%. Nevertheless, this model's effectiveness in complex scenarios remains untested. Gallo et al. [30] developed a real-time PPE detection system utilizing YOLOv4-tiny, designed to process video streams on embedded devices. However, its recognition performance was 10.4% less effective compared to the standard YOLOv4. Wu et al. [31] implemented an upsampling enhancement module (USEM) and a feature fusion module to optimize efficiency and accuracy. However, the model's large parameter count makes it impractical for deployment in operational environments. Zeng et al. [32] enhanced small object detection in YOLOv4 by replacing standard convolutional ones with cross-level hierarchy modules. However, this algorithm requires more computational resources and a longer inference time. Zhao et al. [33] integrated a bidirectional feature pyramid network (BiFPN) in the YOLOv5 framework, with additional detection heads for various scales, improving detection in complex situations. Li et al. [34] enhanced the fitting ability of YOLOv5 using hierarchical positive sample selection (HPSS) and box-density post-processing techniques, reducing the probability of false detection. However, their research was limited to helmet detection and did not cover other PPE like insulated gloves and rods. Hayat et al. [35] optimized YOLOv5x, enhancing the detection of small targets under low-light conditions, although its accuracy in complex backgrounds remains limited. Han et al. [36] enhanced YOLOv5 with super-resolution and CSP modules, minimizing information loss and gradient confusion. However, this may lower small object detection efficiency and require more computational resources. Chen et al. [37] developed a lighter YOLOv5s safety helmet detection model with a Bi-FPN. In comparison with YOLOv5s, this model cut down parameters by 35.7% but had slightly reduced accuracy. Wang et al. [38] applied YOLOv5s and YOLOv5x networks to their custom CHV dataset, observing that

YOLOv5s was 7% less accurate in detecting safety helmets compared to YOLOv5x. Du et al. [39] enhanced the YOLOv5 model by incorporating the GhostConv lightweight network and bidirectional feature pyramid modules. This integration significantly improved real-time detection performance, achieving a mAP of 91.8%. Nguyen et al. [40] integrated the seahorse optimization (SHO) algorithm with the YOLOv5 model for automated PPE detection at construction sites. However, the model experienced prolonged inference times and showed limited effectiveness in complex environments. Liu et al. [41] enhanced the YOLOv5 network by incorporating a coordination attention mechanism. This addition significantly boosted the detection accuracy of reflective clothing while also minimizing false positives. Despite these enhancements, the model's real-time processing efficiency remains constrained. Samma et al. [42] incorporated a contrastive loss branch into the YOLOv7 network, which increased the model's mAP by 2% compared to the YOLOv7. Wang et al. [43] enhanced the YOLOX backbone by integrating the ConvNeXt module to extract deeper features. They also added a new detection head to improve multi-scale predictions, resulting in a 4.23% increase in detection accuracy. However, the dataset used was limited to small targets in low-light conditions. Chen et al. [44] added new feature extraction and content-aware recombination modules to YOLOv7, increasing detail feature capture but at the expense of detection accuracy and higher computational and training demands. Lee et al. [45] used MobileNetV3 as the backbone to lighten YOLOACT, effectively improving PPE detection accuracy, though its performance in complex backgrounds is limited. Shi et al. [46] proposed an improved PPE detection model based on YOLOv8n, enhancing accuracy and reducing complexity. However, the generalization ability of this model in complex scenarios has not been verified. Di et al. [47] replaced the backbone network of YOLOv8s with MobileOne-S0, aiming to enhance PPE detection accuracy. They integrated R-C2F and ASFF modules to fuse features of different sizes, improving the model's overall performance. However, the model sometimes inaccurately detects or misses targets in complex environments, and recognition accuracy for specific targets still requires further enhancement.

3. Method

3.1. YOLOv8n Model Structure

YOLOv8 marks a significant advancement in the YOLO model series, offering improved accuracy and computational speed over its predecessor. The model introduces four architectural variants: YOLOv8n, YOLOv8s, YOLOv8m, and YOLOv8x. These network architectures vary in terms of model size, computational complexity, and detection accuracy. YOLOv8n is the lightest model in the series, designed for rapid processing and suited for resource-constrained devices. It achieves faster processing by reducing the number of network layers and parameters, but at the cost of some accuracy. YOLOv8s strikes a balance between speed and accuracy, making it suitable for standard hardware. It has more layers and parameters than YOLOv8n, which enhances its accuracy and processing speed. YOLOv8m, a medium-complexity model, is positioned between YOLOv8s and YOLOv8x. It aims to balance accuracy and speed, improving detection precision by increasing network depth and width. This model is appropriate for applications that demand accuracy but do not face severe computational limits. YOLOv8x, the most complex and accurate model, is tailored for high-end applications. It features the deepest network structure and the most parameters, which slow its processing speed but ensure optimal performance in complex tasks. The YOLOv8 series caters to a wide spectrum of application requirements, from edge devices to high-performance servers. The primary differences among the models lie in the trade-offs between processing speed and accuracy.

In response to the complex conditions of electrical substations, this research selects YOLOv8n, noted for its streamlined parameterization, as the initial model for enhancement. The architecture of YOLOv8n is systematically divided into four main components, as depicted in Figure 1.

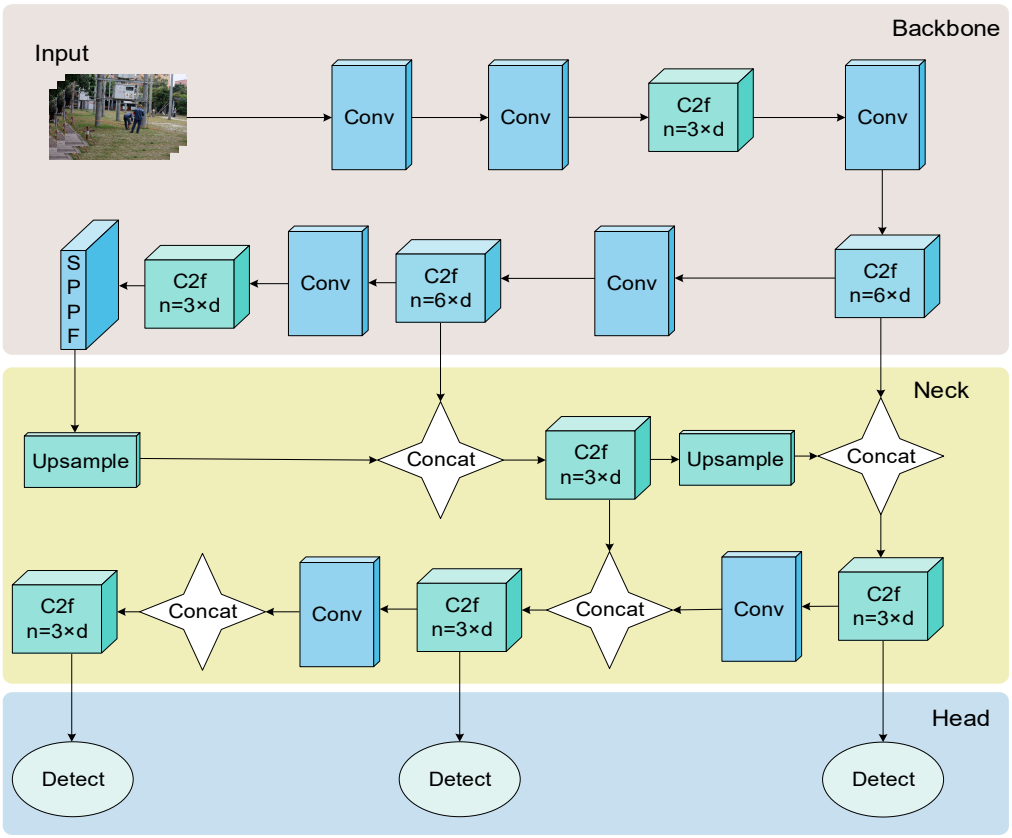

Figure 1. YOLOv8n model structure diagram.

In the YOLOv8n model, the input phase employs Mosaic data augmentation and an anchor-free mechanism. This approach focuses on direct target center localization, effectively reducing the reliance on anchor boxes and subsequently accelerating the NMS [48] process. The backbone structure, essential for feature extraction, encompasses Conv, C2f, and SPPF modules. Specifically, the Conv module handles image convolution, batch normalization, and SiLU activation. YOLOv8 substitutes the C3 module with C2f. The C2f module employs bottleneck structures for multi-layer feature fusion. The SPPF module optimizes the network structure and computation, enhancing feature fusion speed. The neck consists of a feature pyramid network (FPN) [49] and PAN [50]. The FPN extracts advanced features in a bottom-to-top manner, transmitting them to lower layers through top-down lateral connections. Conversely, PAN focuses on acquiring multi-scale feature maps. It merges features through both up and down sampling paths and directs them into the detection head for precise localization and classification. The synergy of PAN and FPN is instrumental in integrating multi-scale features, thereby capturing a variety of targets and increasing the model's accuracy. The head uses an anchor-free mechanism and a decoupled head, processing different scale features separately for accurate target information prediction.

3.2. Improved MEAG-YOLO Model Structure

MEAG-YOLO has been enhanced in both its backbone and neck networks. The MSCA mechanism [51] was introduced in the backbone. In the neck, the concept of the Efficient Channel Attention (ECA) [52] mechanism was adopted, and a newly designed EC2f structure was proposed to optimize the PAN structure. In the head, it integrates the Adaptive Spatial Feature Fusion (ASFF) [53] module and restructures the detection layers. Moreover, this model replaces traditional convolutions with GhostConv [54] modules. Figure 2 illustrates the architectural design of the MEAG-YOLO model. It better retains

surface features and minimizes information redundancy, leading to increased accuracy and fewer parameters, thus requiring less computational power.

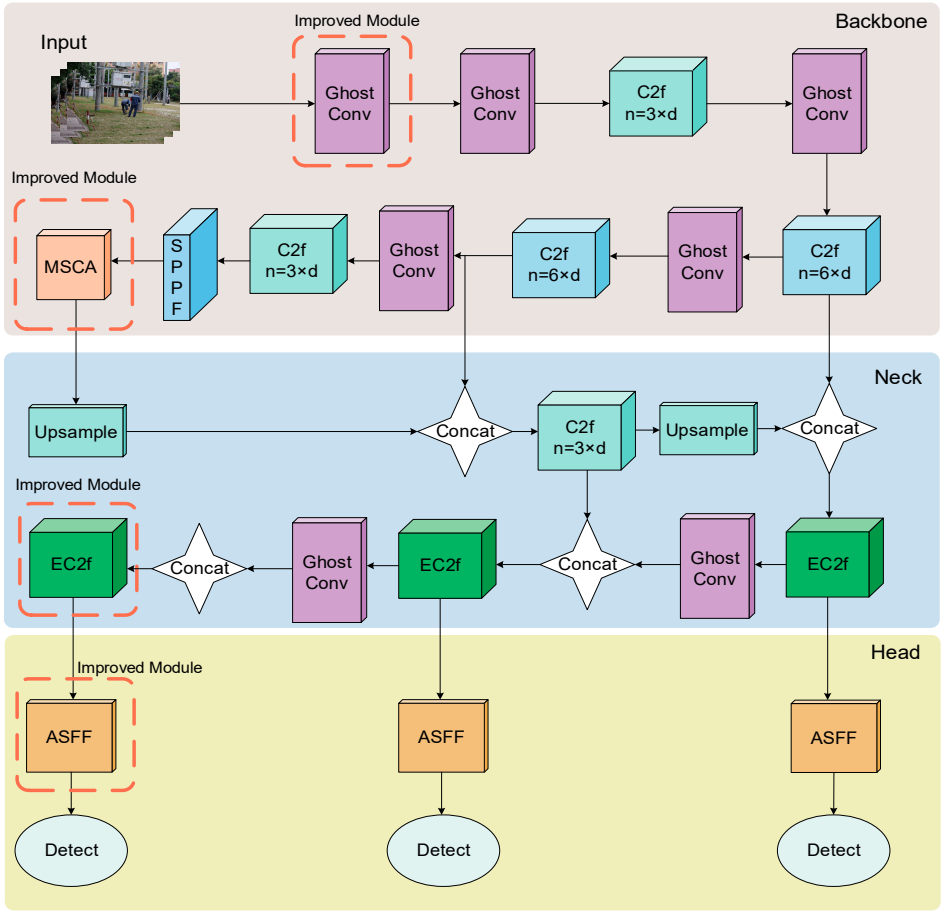

Figure 2. MEAG-YOLO model diagram.

Compared to YOLOv8n, MEAG-YOLO has enhanced feature perception capabilities. It retains more surface features and reduces cross-channel information redundancy, increasing accuracy. Additionally, it has fewer parameters, which lowers the computational resource requirements.

3.2.1. Multi-Scale Convolutional Attention

Considering the requirements for precise and efficient information extraction from complex backgrounds and multi-object detection, this study integrates MSCA into the backbone network. MSCA efficiently aggregates features from various convolutional layers. It captures extensive contextual information and incorporates feature maps of various scales. Additionally, this method improves object detection's accuracy and reliability by utilizing data from various feature layers. The core theory of MSCA is to integrate features across various model layers effectively. This mechanism excels in scenarios with significant variations in target sizes by aggregating feature maps from different scales to boost the model's expressive capacity. The MSCA module allows the model to learn features independently at each scale and improve scene understanding through cross-scale fusion. This fusion strategy greatly enhances detection accuracy, particularly for targets that vary in size, are occluded, or are in complex environments. Additionally, the spatial attention mechanism introduced by MSCA dynamically focuses on different image regions, optimizing feature representation and processing workflows. The inclusion of spatial attention helps the model concentrate on crucial areas that significantly influence the task. Overall, the MSCA module improves the model's ability to detect multi-scale targets in

complex settings and enhances its applicability in resource-limited environments. The structure of the MSCA is shown in Figure 3.

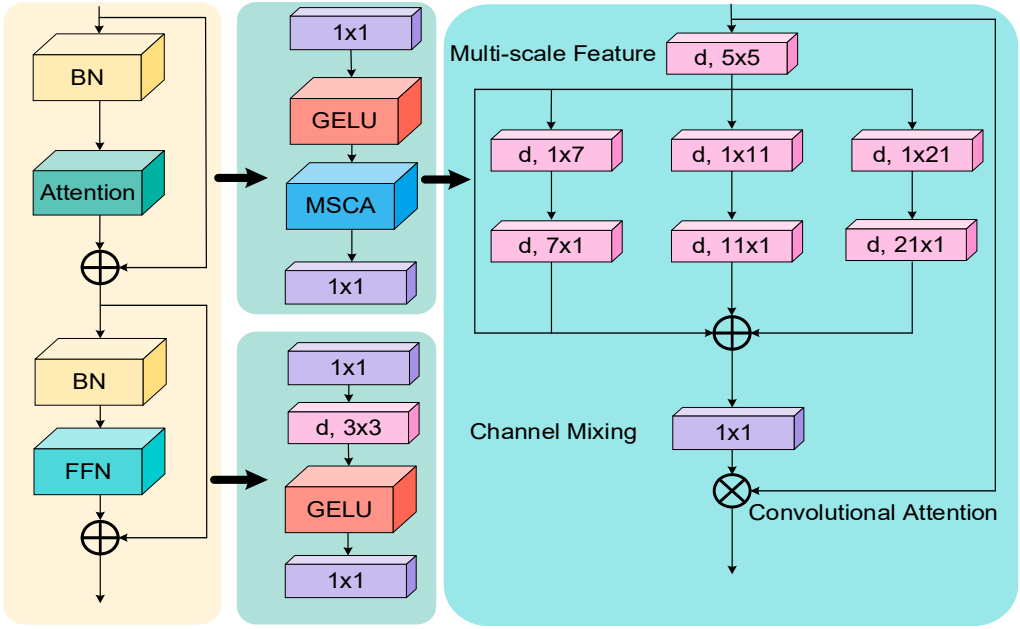

Figure 3. MSCA structure diagram.

The MSCA module is a novel approach that combines multi-scale features and spatial attention information through a convolutional attention network. It efficiently encodes contextual information using lightweight convolution and activates spatial attention by performing element-wise multiplication on the multi-scale features. During the decoding phase, MSCA aggregates multi-level features from different stages, allowing for the acquisition of multi-scale contextual information and adaptation across both spatial and channel dimensions. The mathematical framework of the MSCA module is described in detail in Equations (1) and (2):

$$\mathrm{Att} = \mathrm{Conv}_{1\times1}(\sum_{i=0}^{3} \mathrm{Scale}_i(\mathrm{DW}-\mathrm{Conv}(F))) \tag{1}$$

$$\mathrm{Out} = \mathrm{Att} \otimes F \tag{2}$$

F signifies input features, Att represents the attention map, Out refers to the output, $\otimes$ denotes matrix multiplication, $\mathrm{DW}-\mathrm{Conv}$ indicates depthwise convolution, and Scale_i identifies concatenation. The module comprises three branches, each substituting traditional depth convolutions with two depthwise strip convolutions. This modification efficiently extracts the target features.

3.2.2. EC2f Structure Design

Inspired by the ELAN structure, the C2f module first extracts input image features through the CBS module. Then, the split operation differentiates various feature levels, enabling the independent extraction of multi-scale features. Next, the bottleneck module combines coarse and fine processing stages, ensuring efficient information transmission and a lower computational burden. Finally, the CBS module is used for feature fusion. The structure of the C2f module is shown in Figure 4.

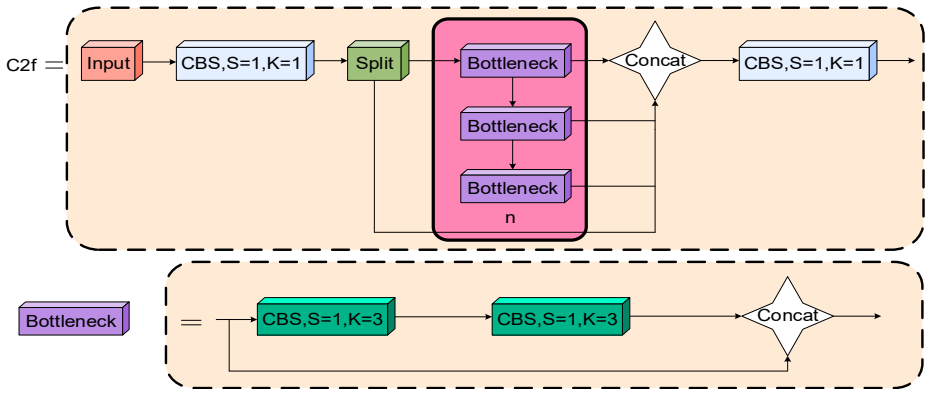

Figure 4. C2f structure.

The detection of PPE encounters numerous challenges, including variable target scales, complex scenes, and vulnerability to environmental interference. While the C2f module can integrate features across different levels, it underutilizes the output features. First, the C2f module may not adequately distinguish between the quality and information content of features at coarse and fine granularities, which can impact the model's performance. Second, it struggles to completely capture the original data's features during the weighting and merging of multi-scale features. Furthermore, the C2f module demands increased computational resources for processing high-resolution feature maps, contributing to greater model complexity. To solve these problems, this study proposes a novel EC2f structure. The EC2f structure combines C2f's layer-by-layer processing with ECA's channel weight adjustment, aiming to enhance key features. This design controls model complexity and enhances feature fusion capabilities. This mechanism optimizes feature expression and fusion by dynamically adjusting channel weights. Initially, the ECA module is tailored to identify and enhance crucial feature channels, thus enhancing the model's adaptability and accuracy with complex data. Secondly, the ECA module adaptively adjusts weights based on detailed channel information. This reduces parameters and computational demand, thereby enhancing efficiency. Additionally, the integration of the ECA module boosts the capability of the EC2f structure to overcome the feature fusion limitations observed in the C2f structure. This enhancement improves key feature detection, greatly increases accuracy and robustness, and maintains a relatively balanced computational load. The EC2f structure's design significantly enhances the model's performance, particularly in complex PPE detection tasks. The EC2f structure is illustrated in Figure 5.

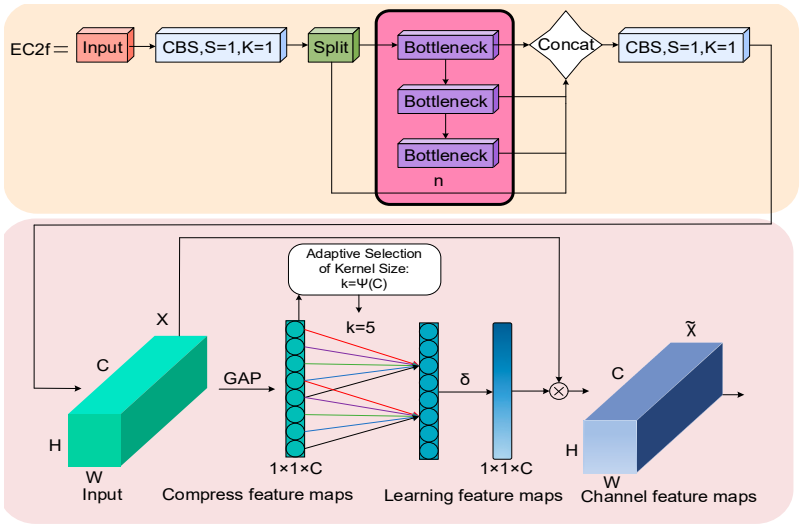

Figure 5. EC2f structure diagram.

However, the C2f module is insufficient for the information fusion of input features. To enhance this structure, the ECA module has been incorporated into this study. The ECA module replaces SENet's fully connected layers with 1×1 convolution layers, preventing the reduction of feature dimensions.

The ECA module begins by performing global average pooling. This step involves extracting feature maps from the backbone network and calculating the average value of each channel, thereby capturing global spatial information. The structure of Global Average Pooling (GAP) is shown in Figure 6.

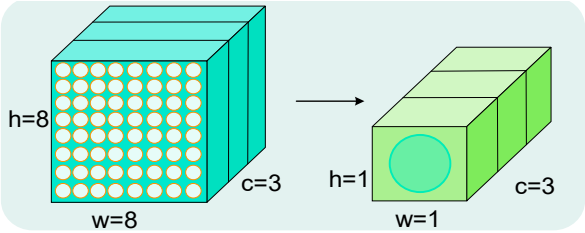

Figure 6. GAP structure diagram.

The GAP structure, depicted in Figure 6, aggregates the averages of pixel points to create a compressed feature map of dimensions $1 \times 1 \times C$. This map is then processed with a one-dimensional convolution kernel of size 5, as detailed in Equation (3):

$$k = \psi(C) = \left| \frac{log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{3}$$

The kernel size in Equation (3) is adaptively determined based on the channel dimension by selecting the closest odd number to C. γ and b are set to two and one, respectively. This approach allows the model to directly learn inter-channel local relationships without requiring additional parameters. The output from the one-dimensional convolution is then transformed into channel weights through an activation function. These weights are used to recalibrate the original feature map, effectively highlighting important channels and reducing the impact of less relevant ones. The final result is a feature map refined with channel attention.

3.2.3. Adaptive Spatial Feature Fusion Module

The size of the PPE in the substation scenario can vary greatly. The original YOLOv8n network utilizes the FPN + PAN structure to merge information, effectively integrating high-level and low-level features. However, the network's weight allocation may lead to the loss of critical features, thereby affecting detection performance. To address this issue, this study incorporates the ASFF module to enhance the effectiveness of merging features from different scales. The ASFF module is a sophisticated mechanism that dynamically adjusts the weights of feature maps across different scales to achieve fusion. The structure of the ASFF module is illustrated in Figure 7.

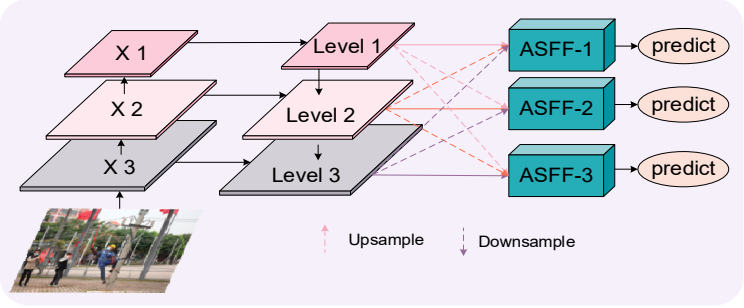

Figure 7. ASFF module structure diagram.

The ASFF module uses an adaptive learning system to change the weights of feature maps on various scales, depending on the task and data traits. Before combining them, it assigns these adjusted weights to the feature maps. This adaptive weight distribution allows the network to automatically adapt fusion strategies based on the importance of features across different spatial positions and scales. The ASFF module dynamically adjusts the fusion weights of feature maps across different scales, enhancing the efficiency of feature utilization. It fine-tunes the enhancement or suppression of features, thereby improving the model's capacity to recognize multi-scale targets in complex scenes. This mechanism enables the model to tailor its fusion strategy based on the spatial location and scale significance of features, leading to more precise target detection. With the integration of the ASFF module, the YOLOv8n model's ability to effectively process multi-scale targets is enhanced, particularly in intricate substation environments.

The fusion process of ASFF-1 is demonstrated using X1, X2, and X3 as examples. This method uses three feature maps from the YOLOv8n backbone. They are then handled by combining FPN and PAN at three levels: Level 1, Level 2, and Level 3. ASFF-1 is formed by combining these three levels. The process of feature fusion at Level l is described by Equation (4):

$$\mathbf{y}_{ij}^{l} = \alpha_{ij}^{l} \cdot \mathbf{x}_{ij}^{1\to l} + \beta_{ij}^{l} \cdot \mathbf{x}_{ij}^{2\to l} + \gamma_{ij}^{l} \cdot \mathbf{x}_{ij}^{3\to l} \tag{4}$$

The vector at the $(i,\ j)$ position in the inter-channel output feature mapping is $\mathbf{y}_{ij}^{l}$. α_{ij}^{l}, β_{ij}^{l}, and γ_{ij}^{l} are the spatial importance weights of features from different levels during the forward propagation of the first layer. $\mathbf{x}_{ij}^{n\to l}$ represents the feature vector at position $(i,\ j)$ from level n to l. After processing with the softmax function, $\alpha_{ij}^{l}+\beta_{ij}^{l}+\gamma_{ij}^{l} = 1$, $(\alpha_{ij}^{l}, \beta_{ij}^{l}, \gamma_{ij}^{l} \in [0,\ 1])$. This module adaptively aggregates features from three distinct levels. They operate on their respective scales. After this aggregation, the fused features are passed to the head structure. The head structure is instrumental in classifying and detecting PPE.

3.2.4. GhostConv Module

The precise and prompt detection of PPE in substations is crucial. While existing deep learning models have improved detection accuracy, they often lack lightweight and real-time capabilities. The YOLOv8n model consists of numerous convolutions and has a high parameter count and complex computations. To streamline the model, this study incorporates the GhostConv module into the YOLOv8n framework. This integration aims to simplify computations and enhance model efficiency. By adopting the GhostConv module, this research focuses on reducing complexity while maintaining performance. This can meet the requirements of high-detecting efficiency substation scenarios. The GhostConv module uses grouped convolution technology instead of traditional convolution. This significantly reduces the model's parameter count and computational complexity. The method maintains performance while reducing computational resource consumption. This reduction is crucial for real-time processing environments with limited resources. The module breaks down standard convolutions into lighter operations, which reduces redundant computations in convolutional layers. This reduction eases the computational load. The design simplifies the network structure and reduces energy consumption. It also enhances computational efficiency, making the model more suitable for edge devices. There may be a slight performance drop, but this is acceptable in substation environments where real-time responsiveness is crucial. Overall, the introduction of the GhostConv module increases processing speed and decreases resource use. This enables the YOLOv8n model to meet the needs of modern monitoring systems that require fast and efficient computation while maintaining detection accuracy. The structure of GhostConv is shown in Figure 8.

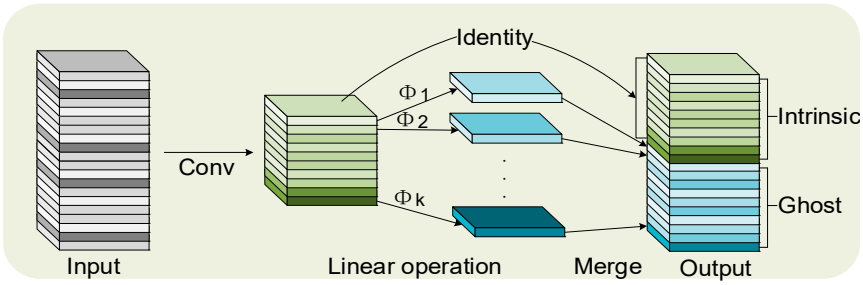

Figure 8. GhostConv module structure diagram.

The GhostConv convolution process occurs in three steps. Firstly, the input image generates an original feature map through standard convolution.

$$Y' = X * f' \tag{5}$$

In Equation (5), the input data are $X \epsilon R^{c \times h \times w}$, where c represents the number of input channels, and h and w are the height and width, respectively. Y' represents the Ghost output feature map, $*$ denotes the convolution operation, and f' is the size of the convolution kernel. Next, the original feature map is transformed into multiple Ghost feature maps through a linear transformation.

$$y_{ij} = \Phi_{i,j}(y'_i) \tag{6}$$

Next, the original and Ghost feature maps merge to create the final output feature map, with $\Phi_{i,j}$ indicating the feature mapping and y_{ij} being the output, as shown in Equation (6).

This study utilized the GhostConv module to optimize the convolution layers in both the backbone and neck networks of the YOLOv8n model. The implementation of lightweight linear operations significantly decreases the parameter count and enhances the speed of inference. While this optimization slightly reduces detection accuracy, the resulting lightweight model is better suited for real-time detection in resource-constrained environments.

4. Experimental and Results

4.1. Experimental Datasets

Detecting PPE in substations quickly and accurately is vital. However, there is a shortage of open-source datasets, including a range of this equipment. Therefore, this study created a dataset featuring six categories: individuals, insulating gloves, safety helmets, armbands, insulating rods, and regular gloves. This dataset combines online resources and on-site photography to ensure it represents the diversity of real application scenarios. The photography captures various times and backgrounds, simulating different working environments at substations. This study prepared for training by using LabelImg software (v.1.8.1) to create text files with image paths, annotations, and labels. Moreover, to enhance the model's generalization capabilities, this study employed techniques such as shifting, modifying brightness and saturation, and introducing noise. These steps increased the dataset to 2800 images. These techniques increase the dataset's coverage of different scenes and realistically mimic changes in lighting and background, enhancing the model's robustness. Each category of samples is extensively represented, both in quantity and scene complexity. The dataset promotes diversity by including various types of PPE and different wearers, capturing the range of conditions in substation work environments. Furthermore, it is noted that many workers use insulated tools but do not wear insulated gloves. This diversity is secured through a meticulously planned collection strategy, designed to capture a wide range of anomalies to train a more resilient model. Examples of the enhanced data are shown in Figure 9.

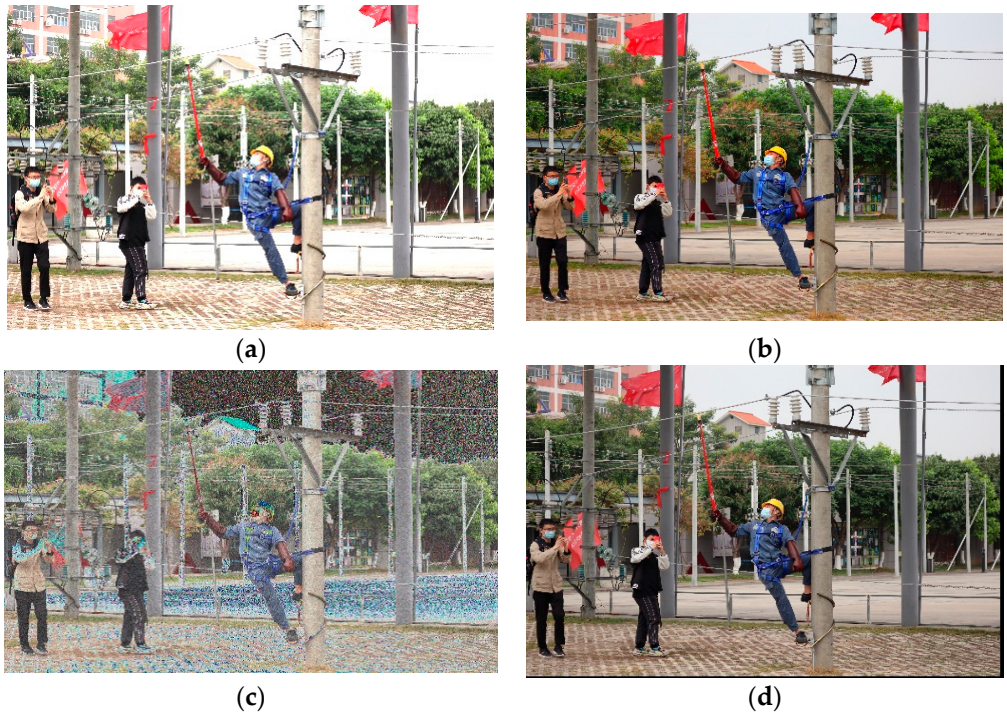

Figure 9. Data-enhancement diagram. (**a**) Adjusting brightness; (**b**) adjusting saturation; (**c**) adding noise; and (**d**) random panning.

The dataset is split into training and testing sets using an 8:2 ratio to ensure that the model performs well on new data. Figure 10a displays a heat map depicting the density distribution of data points across a two-dimensional plane. Darker colors indicate a higher probability of the target's presence. The graph's x and y axes extend from 0 to 1, indicating the normalized coordinates of object centers within bounding boxes. A notable clustering of data points towards the center implies a frequent central placement of target objects in images. In contrast, Figure 10b shows a bar chart detailing the occurrence rates of various categories in the dataset.

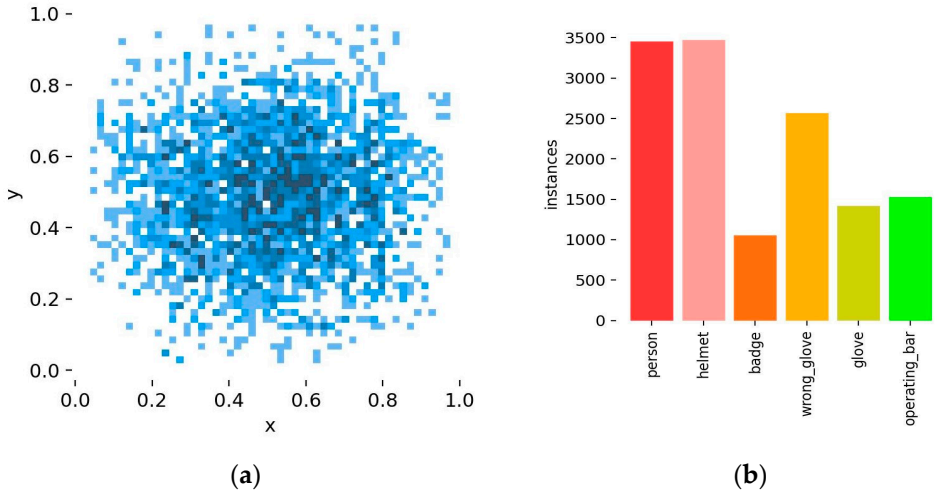

Figure 10. (**a**) PPE category diagram; (**b**) distribution plot of x and y coordinates.

4.2. Experimental Environments and Evaluation Metrics

All experiments were conducted in a Linux environment using an NVIDIA GeForce RTX 4090 Ti graphics card (32 GB of video memory) (NVIDIA, Santa Clara, CA, USA),

the deep learning framework PyTorch version 1.7.0, and the Python 3.8 environment. To ensure the accuracy and reproducibility of the findings, a thorough analysis was conducted during the selection and optimization of hyperparameters. The initial learning rate was set at 0.01 to balance model convergence speed and stability. A lower learning rate helps prevent premature convergence to local optima, thus enhancing efficiency and maintaining accuracy. The training duration was established at 300 epochs, allowing ample time for the model to learn complex data features. This duration is chosen based on the behavior of the loss function and validation set performance, ensuring training continues until the model stabilizes. The momentum parameter was set at 0.937 to increase the optimizer's inertia, accelerating convergence and reducing training oscillations. To prevent overfitting, a weight decay of 0.0005 was applied, helping the model learn generalizable features by minimizing the weights. This setting balances the model's complexity with the data volume, effectively managing overfitting with a smaller weight decay. Considering the limits of graphics card memory and parallel processing capabilities, the batch size was determined to be 16. Experiments showed that this batch size optimally balances efficiency and performance.

To objectively evaluate the improved MEAG-YOLO algorithm, this study uses mAP, $Precision$, $Recall$, and $F1$ scores as evaluation metrics.

$$mAP = \frac{1}{6}\sum_{i=1}^{6} AP_i \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{9}$$

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

Six is the total number of categories in the test dataset. For each category i, the average accuracy is calculated and recorded as AP_i. The mean average precision (mAP) is obtained from the AP_i of each category. Key indicators for the performance evaluation include True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN). These indicators demonstrate the model's performance in classification tasks. For example, FP refers to the model misclassifying a negative sample as positive. $Precision$ assesses the proportion of correctly labeled positive samples. FN is when a positive sample is misjudged as negative. $Recall$ measures the proportion of correctly identified real positive samples. To comprehensively evaluate efficiency and performance, this study also considers the model's total number of parameters and FLOPs. The total number of parameters indicates the complexity of a model, with more parameters usually requiring higher computing resources. FLOPs reflect the computational load of the model during inference by measuring the number of floating-point operations performed.

5. Experimental Results and Analysis

5.1. Analysis of Results before and after Improvement

The article contrasts the MEAG-YOLO model before and after improvements to ascertain its efficacy, as presented in Table 1.

Table 1. Comparison of PPE detection results.

Model	Precision (%)	Recall (%)	F1 (%)	mAP (%)	Parameters (M)
YOLOv8n	96.1	93.3	94.7	96.1	3.0
MEAG-YOLO	98.4	94.2	96.3	96.5	2.8
Δ	+2.4%	+0.9%	+1.7%	+0.4%	−6.7%

In comparison to YOLOv8n, the MEAG-YOLO model achieves a precision increase of 2.4% and a 6.7% reduction in the parameter count. These advancements are primarily attributed to the MSCA module, which enhances feature extraction in complex backgrounds. Moreover, the model's parameter count has been effectively controlled by integrating the lightweight design of GhostConv and the innovative EC2f module. The EC2f module focuses on extracting key features while simplifying background complexities. The inclusion of the ASFF module further improves the model's performance by amalgamating multi-scale features, thereby enhancing the overall effectiveness of the model. Consequently, the MEAG-YOLO model has achieved significant improvements in detection accuracy and has efficiently reduced the number of parameters.

To more clearly demonstrate the detection capabilities, precision–recall curves for the YOLOv8n and MEAG-YOLO models are displayed in Figure 11. The graphs reveal the MEAG-YOLO model's superior performance over the conventional YOLOv8n model. Furthermore, MEAG-YOLO exhibits greater precision as recall rises.

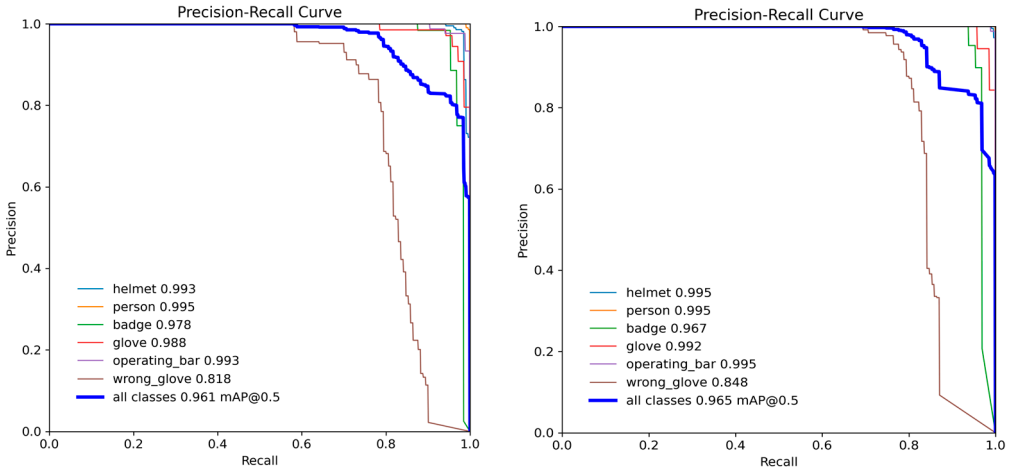

Figure 11. (**a**) YOLOv8n P-R curve; (**b**) MEAG-YOLO P-R curve.

5.2. Ablation Experiment and Analysis of Results

To assess the impact of each enhancement, this paper performed ablation experiments. The outcomes are presented in Table 2.

Table 2. Ablation experiment results.

Model	Precision (%)	Δ_1	Recall (%)	mAP (%)	FLOPs (G)	Δ_2
YOLOv8n	96.1		93.3	96.1	8.1	
YOLOv8n + MSCA	96.4	**+0.3%**	93.5	96.4	8.2	
YOLOv8n + MSCA + EC2f	97.9	**+1.9%**	93.8	97.3	8.2	
YOLOv8n + MSCA + EC2f + ASFF	98.7	**+2.7%**	95.1	98.1	8.2	
MEAG-YOLO	**98.4**	**+2.4%**	**94.2**	**96.5**	**7.6**	**−7.3%**

Table 2 and Figure 12 show that incorporating the MSCA module into the YOLOv8n backbone network improved precision by 0.3%. This improvement is attributed to the mechanism's optimization of feature extraction. The MSCA module boosts detection performance by dynamically concentrating on multi-scale features and refining feature maps. This process minimizes background noise and irrelevant details, allowing for the effective identification and processing of information-rich visual features. The newly designed EC2f structure significantly improves the extraction of essential information, leading to a precision increase of 1.9%. Meanwhile, the ECA module efficiently captures channel features' global information, adaptively tuning the feature response. This enhancement bolsters the

network's representational capacity without extensive dimensionality reduction. Furthermore, the ASFF module in the Head network greatly strengthens multi-scale information fusion. As a result, the precision rose by 2.7% compared to the original YOLOv8n. This module dynamically manages feature fusion, optimizing the use of spatial information across various resolutions. By integrating features from distinct spatial dimensions, this fusion technique improves scene comprehension and the accuracy of target localization and identification. It is particularly effective in scenarios with considerable target size variations and where precise spatial details are paramount.

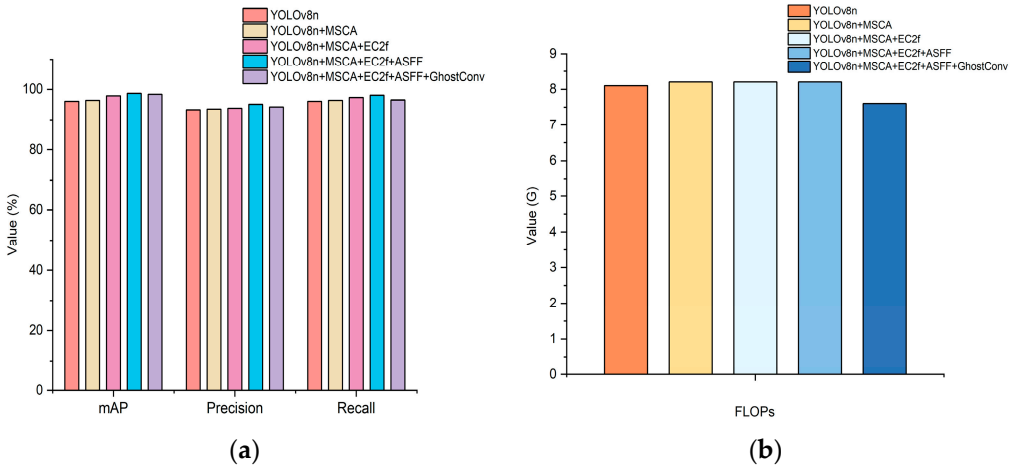

Figure 12. Ablation experiment results: (**a**) comparison of the accuracy of each model; (**b**) comparison of the FLOPs.

Figure 13 clearly shows the ablation experimental results. Replacing standard convolution with GhostConv led to a 0.9% decrease in precision. The minor precision loss is compensated by a substantial reduction in computational demand, creating a reasonable trade-off. This is particularly valuable in substation monitoring scenarios where high real-time responsiveness is essential. However, MEAG-YOLO achieved a 7.3% reduction in FLOPs across all test datasets, signaling a marked increase in computational efficiency. This improvement stems from the GhostConv module, which produces feature maps using fewer convolutions and creates additional "ghost" feature maps via linear transformations. Specifically, GhostConv uses grouped convolutions to replace some standard convolutions, significantly reducing unnecessary computations. It substitutes some operations in standard convolutional layers with lightweight linear operations, cutting computational requirements by about 50%. This reduction decreases hardware performance demands and improves the model's suitability and inference speed in resource-limited environments. This module reduces the number of direct convolutions, cutting computational complexity and resource usage. It efficiently captures the vital features needed for detection tasks, illustrating the module's capability to balance efficiency and performance. The outcomes confirm that the model is both lighter and more computationally efficient. Despite a slight sacrifice in accuracy, the introduction of GhostConv crucially enhances computational efficiency, which is vital for practical deployment.

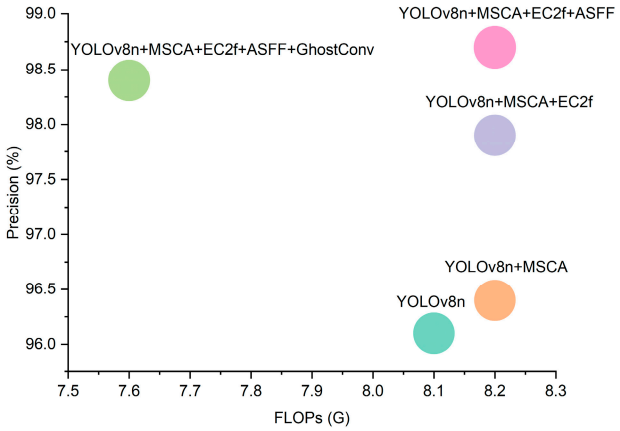

Figure 13. Comparison of precision with different model FLOPs.

5.3. Comparative Experiment and Analysis of Results

In this study, the MEAG-YOLO model was evaluated alongside similar models. The results are shown in Table 3.

Table 3. Performance comparison of object detection models.

Model	Precision (%)	Recall (%)	F1 (%)	mAP (%)
Faster R-CNN	89.7	87.8	88.7	88.6
SSD	90.6	89.4	89.9	90.1
YOLOv3	92.3	90.2	91.2	91.8
YOLOv5	94.2	91.5	92.8	93.5
YOLOv7-tiny	94.9	92.3	93.6	94.3
MEAG-YOLO	**98.4**	**94.2**	**96.3**	**96.5**

Table 3 and Figure 14 highlight significant improvements in crucial performance indicators, including mAP, accuracy, recall rate, and F1 score, for the enhanced model. SSD surpasses faster R-CNN in detection accuracy by directly detecting objects on multi-scale feature maps, adeptly managing targets of various sizes. SSD concurrently processes multi-scale feature maps, streamlining the detection workflow. It increases detection efficiency by directly predicting categories and bounding box offsets with anchor boxes of different sizes on each feature map layer. The YOLO series outstrips both faster R-CNN and SSD in performance metrics. YOLOv3 employs a multi-scale prediction mechanism to maintain high accuracy across targets of various scales, enhancing the overall mAP. It incorporates Darknet-53, a complex feature extractor with residual connections. These connections prevent gradient vanishing and enhance feature representation. YOLOv5's improved architecture allows adaptive and automatic adjustments. It employs an effective feature extraction network and advanced multi-scale prediction, enhancing adaptation to various target shapes and sizes. YOLOv7-tiny boasts the innovative E-ELAN network structure, refining feature extraction and utilization to boost detection accuracy. It employs cross-scale feature fusion to enhance feature capture and integration. Additionally, it utilizes efficient data augmentation and optimization algorithms during training, improving generalization and robustness. Last but not least, MEAG-YOLO integrates MSCA and optimizes the detection head. Through the innovative design of the EC2f structure and GhostConv module integration, feature extraction and fusion efficiency. It maintains a high mAP value, achieving a balance between detection accuracy and lightweight design.

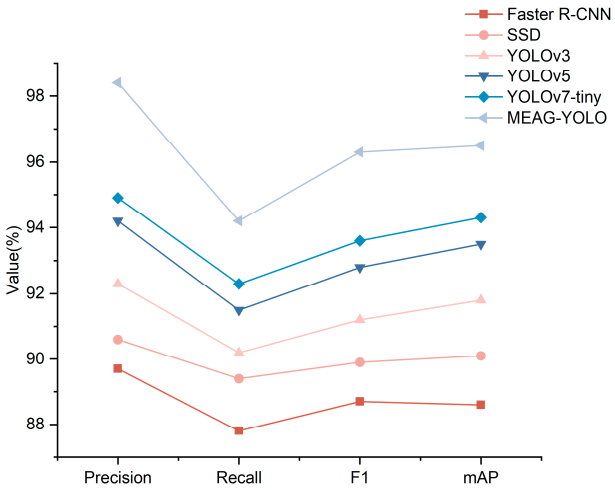

Figure 14. Comparison of experimental results.

Finally, Figures 15 and 16 provide a comparative analysis of the detection results between the YOLOV8n and MEAG-YOLO models.

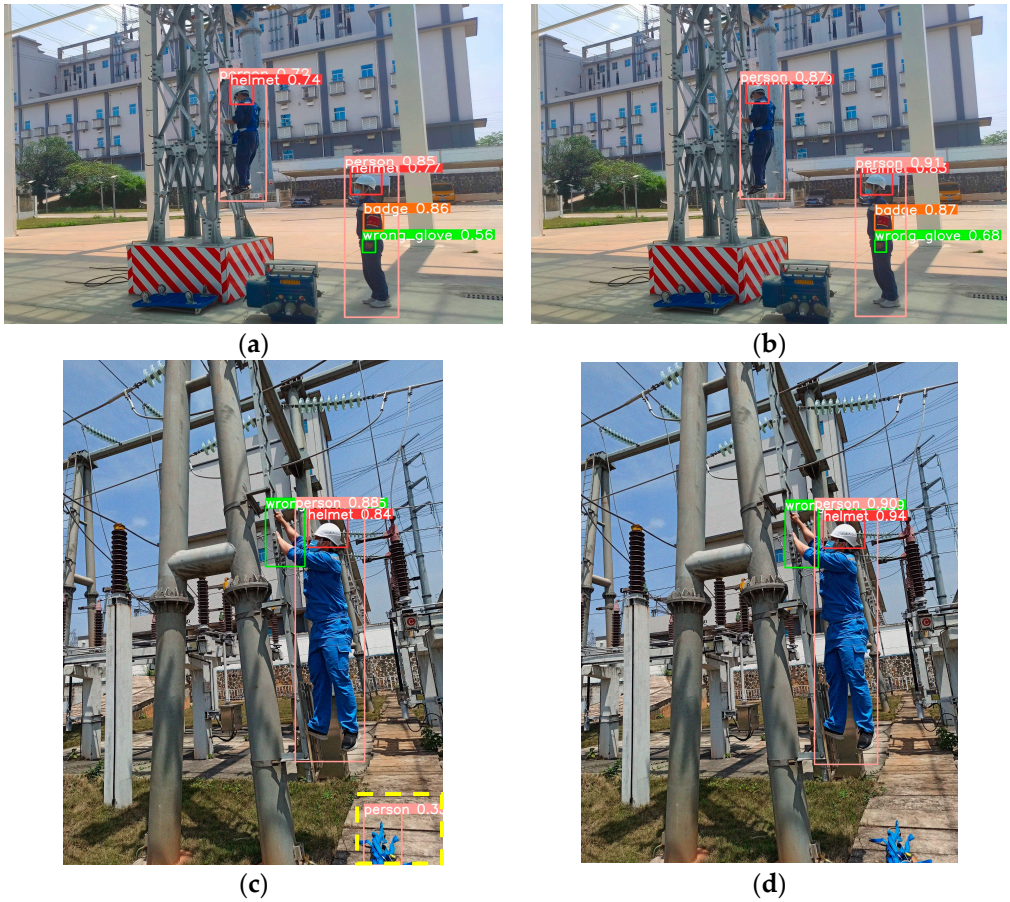

Figure 15. Comparison of model detection results. (**a**) YOLOv8n's detection results; (**b**) MEAG-YOLO's detection results. (**c**) YOLOv8n's detection results; (**d**) MEAG-YOLO's detection results.

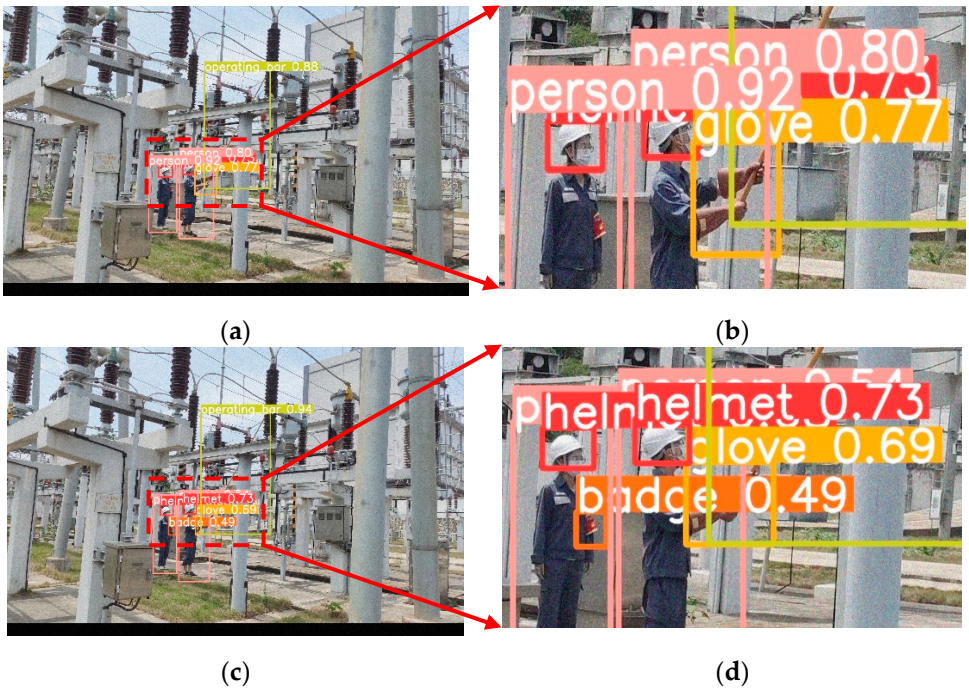

Figure 16. Comparison of model detection results. (**a**) YOLOv8n's detection results; (**b**) MEAG-YOLO's detection results; (**c**) Local magnification of YOLOv8n's detection results; (**d**) Local magnification of MEAG-YOLO's detection results.

Figure 15a shows YOLOv8n correctly identifying persons, safety helmets, ordinary gloves, and armbands, with confidence levels of 0.77 for helmets and 0.56 for gloves. Figure 15b highlights MEAG-YOLO's improvements in detecting small items in complex settings, increasing confidence to 0.83 and 0.68 for helmets and gloves, respectively. In Figure 15c, YOLOv8n accurately detects people, helmets, and gloves. But this model incorrectly identifies a background object as a "person". The yellow dashed box highlights this situation. Figure 15d shows MEAG-YOLO accurately identifying targets and avoiding false alarms.

Figure 16a,c compare YOLOv8n's and MEAG-YOLO's detection abilities. Figure 16b,d zoom in on YOLOv8n's results and MEAG-YOLO's results, respectively. It shows that MEAG-YOLO correctly identified the missed armband, thereby addressing issues in the original model.

6. Discussion

This paper proposes the MEAG-YOLO model for accurately and rapidly identifying the PPE of workers in substations. The model employs specific feature extraction and optimization strategies. This enhances detection accuracy and makes the model lightweight. Consequently, it signifies substantial progress in the development of lightweight deep learning models tailored for particular application scenarios.

This study delves deeply into the problem of detecting the PPE of substation workers and proposes an improved version of the YOLOv8n model. Initially, the MSCA module is integrated into the model's backbone network. This approach significantly improves the model's ability to recognize PPE in substations. Next, to meet the requirements for lightweight design, the EC2f structure was innovatively designed. This architecture streamlines feature processing while minimizing parameter count and complexity. Furthermore, the ASFF module was introduced to enhance detection performance. This module strengthens feature fusion, improving accuracy and efficiency. Considering the need for efficient computation, GhostConv was combined with YOLOv8n. This method enhances computational efficiency and reduces the number of parameters by minimizing convolution operations. The GhostConv module fundamentally reduces the computational complexity

of convolutional neural networks by modifying standard convolutional layer operations. Specifically, it splits traditional convolutions into pointwise convolutions for linear transformations and cost-effective depthwise separable convolutions for nonlinear transformations. This approach decreases the model's parameters and computational load, boosting efficiency on edge devices. However, the simplified computation may compromise feature extraction, particularly with datasets in complex weather conditions. Although standard convolutions gather extensive feature information, the simplified approach of the GhostConv module may struggle to interact fully with all input features, potentially missing critical details. This shortfall could marginally lower the model's performance under severe conditions. To address this issue, future research could consider two improvement strategies. One strategy involves using both the GhostConv module and traditional convolutions in critical parts of the model to maintain sensitivity to complex features. Additionally, introducing attention mechanisms and dynamically adjusting weight distributions can optimize feature extraction, enhancing the capture of important information. Overall, the GhostConv module significantly lowers computational demands, making it ideal for resource-limited, real-time processing environments. As shown in Table 3, MEAG-YOLO excels in detecting PPE in substation scenarios. Compared to previous studies, two-stage target detection algorithms, such as Faster R-CNN, have shown improved performance. However, they suffer from low computational efficiency, limited detection capability for small targets, a large number of parameters, and slow inference. On the other hand, single-stage algorithms like SSD can directly predict and are simpler to train, but they have lower accuracy and limited feature extraction capability in complex scenarios. Therefore, MEAG-YOLO aims to enhance detection accuracy in complex scenarios while reducing the number of parameters and complexity.

However, the model proposed in this paper still has some limitations. Firstly, the model needs large quantities of accurately annotated training data. Acquiring these data is both time-consuming and costly, especially in substations with variable weather conditions. Although the MSCA and ASFF modules enhance feature extraction and fusion, their robustness under extreme weather, variable lighting, and diverse backgrounds still needs further validation.

Future research plans to broaden the dataset to encompass scenarios such as personnel wearing work attire and safety belts, and to gather data under challenging weather conditions like rain, fog, or snow. These enhancements aim to boost the model's overall performance. Additionally, future research aims to include more samples from edge positions to even out data distribution. Since each image can contain multiple targets from the same category, the number of instances per category can surpass 2800. This could result in the model performing better in these categories but worse in others. To address this issue, strategies such as oversampling, undersampling, or adjusting category weights are planned. These methods aim to balance the sample distribution and enhance the model's ability to recognize minority categories. In terms of network architecture, the MSCA module currently shows limited robustness in handling drastic target scale variations. The ASFF module may increase the computational load when processing high-dimensional features, limiting the model's deployment and application. The GhostConv module aims to reduce computational resource consumption but may be insufficient in feature extraction when dealing with complex features. Furthermore, the model design prioritizes lightweight and efficiency, yet computational resource limitations are a major consideration in practical deployments. In edge computing environments, the limited processing power and storage of processors can further affect model performance. While the GhostConv module reduces computational resource use, its ability to extract features from large, complex images may be insufficient, potentially impacting application outcomes. Despite the MEAG-YOLO model's strong test performance, future work will explore additional optimization strategies, such as enhancements to algorithms and network structures, to boost the model's generalization capabilities and practical utility.

In the future, the research will incorporate imagery captured in diverse weather conditions. This will be performed by generating images with GAN networks. The study also intends to develop layered or recursive attention structures to increase the efficiency of feature extraction. A key focus will be on devising effective multi-scale feature fusion strategies, aiming to more adeptly amalgamate information across various scales. Meanwhile, the research plans to optimize grouped convolution strategies and introduce efficient activation functions to enhance feature extraction efficiency. Furthermore, this study will consider knowledge distillation strategies to reduce parameter counts and computational requirements. This approach will enhance the model's applicability with limited hardware resources.

The model proposed in this study demonstrated outstanding performance in testing datasets. However, its practical value hinges not just on performance but also on scalability across different environments, ease of deployment, and integration capabilities with existing systems. Future studies will assess the model's scalability on large datasets and multiple computing platforms. This evaluation will illuminate how the model performs under increasing data loads and computing demands, providing insights for further refinements. Deployment challenges include ensuring hardware compatibility, managing software dependencies, and simplifying maintenance. It is also essential to safeguard user data privacy and security during operation. Moreover, effective integration of the model is critical. It enhances practical deployment and fosters collaboration with other systems, boosting security measures.

The MEAG-YOLO model proposed in this study performed excellently on a custom dataset. It also reduced complexity. Future studies will comprehensively evaluate the model's practicality and applicability. They will test its scalability and efficiency on different dataset sizes and hardware configurations. Larger datasets require longer training times, but parameter adjustments help maintain the model's consistent performance. This shows strong adaptability. Optimizing algorithms and refining the model's structure can greatly enhance computational efficiency and reduce resource demands. The model's adaptability allows it to operate efficiently across various hardware settings, even with limited resources. These efforts will provide vital insights for further development and optimization of the model. They will also lay a solid foundation for future research into the diversity and challenges of real deployment scenarios.

7. Conclusions

This study proposes a lightweight and efficient MEAG-YOLO model to enhance the accuracy of detecting the PPE of substation workers. Initially, the model integrates the MSCA module, optimizing feature extraction for targets of various scales in complex backgrounds. This approach improves detection performance. Secondly, inspired by the ECA mechanism, this study innovatively designed the EC2f structure, effectively enhancing the efficiency of key information extraction and detection accuracy. Furthermore, improvements were made to the model's PAN structure to further boost performance. This effectively integrates information from targets of different scales, enhancing feature fusion capabilities. Finally, the combination of GhostConv with YOLOv8n reduces the model's complexity and makes it lightweight. Experimental results demonstrate that MEAG-YOLO improves mAP by 2.4% and reduces model parameters by 6.7%. Compared to other mainstream target detection algorithms, this model demonstrates superior efficiency and accuracy. These qualities are essential for improving safety monitoring in substations and reducing accident risks among power grid workers. However, it is important to note that the dataset used in this study did not cover complex weather conditions. Future research will concentrate on enhancing both accuracy and speed of detection, as well as applying the model to devices such as inspection robots. Future efforts will aim to improve the MEAG-YOLO model's detection abilities in different weather conditions. They will also work on integrating the model more effectively into automation robots used in the power industry. Additionally, research will address the balance between model complexity and

detection efficiency in environments with limited resources to ensure the model performs reliably and can be practically deployed.

Author Contributions: Conceptualization, H.Z. and C.M.; software, H.Z. and X.G.; validation, X.G. and C.H.; formal analysis, H.Z. and X.M.; writing—original draft preparation, H.Z.; writing—review and editing, C.M. and X.M.; funding acquisition, C.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Autonomous Region Science and Technology Innovation Leading Talent Training Project (grant number 2021GKLRLX08), Yinchuan Science and Technology Innovation Project (grant number 2022GX04), and Key Scientific Research Project of North MinZu University (grant number 2023ZRLG10).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Written informed consent has been obtained from the patient(s) to publish this paper.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by corresponding author on request.

Acknowledgments: The authors would like to thank the editors and anonymous reviewers for their helpful suggestions. We sincerely thank Hongyuan Gu of the University of Science and Technology of China for his suggestions. Additionally, we sincerely thank Chuang Li of North Minzu University for his guidance.

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

References

1. Gammon, T.; Lee, W.-J.; Intwari, I. Reframing Our View of Workplace "Electrical" Injuries. *IEEE Trans. Ind. Appl.* **2019**, *55*, 4370–4376. [CrossRef]
2. Meira de Andrade, P.H.; Villanueva, J.M.M.; Macedo Braz, H.D. An Outliers Processing Module Based on Artificial Intelligence for Substations Metering System. *IEEE Trans. Power Syst.* **2020**, *35*, 3400–3409. [CrossRef]
3. Electrical Fatalities in the Workplace: 2011–2021. Available online: https://www.esfi.org/electrical-fatalities-in-the-workplace-2011-2021 (accessed on 11 November 2023).
4. National Energy Administration: Accident Analysis Report for 2020. Available online: https://www.nea.gov.cn/2021-03/25/c_139834495.htm# (accessed on 5 March 2021).
5. National Energy Administration: Accident Analysis Report for the First Quarter of 2021. Available online: http://www.nea.gov.cn/2021-08/10/c_1310119210.htm (accessed on 10 August 2021).
6. National Energy Administration: Accident Analysis Report for the Second Quarter of 2021. Available online: http://www.nea.gov.cn/2021-09/09/c_1310177594.htm (accessed on 9 September 2021).
7. Zhao, M.; Barati, M. Substation Safety Awareness Intelligent Model: Fast Personal Protective Equipment Detection Using GNN Approach. *IEEE Trans. Ind. Appl.* **2023**, *59*, 3142–3150. [CrossRef]
8. 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]
9. Barro-Torres, S.; Fernández-Caramés, T.M.; Pérez-Iglesias, H.J.; Escudero, C.J. Real-Time Personal Protective Equipment Monitoring System. *Comput. Commun.* **2012**, *36*, 42–50. [CrossRef]
10. 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]
11. 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]
12. Kang, S.; Park, M.-W.; Suh, W. Feasibility Study of the Unmanned-Aerial-Vehicle Radio-Frequency Identification System for Localizing Construction Materials on Large-Scale Open Sites. *Sens. Mater.* **2019**, *31*, 1449. [CrossRef]
13. Zhang, H.; Yan, X.; Li, H.; Jin, R.; Fu, H. Real-Time Alarming, Monitoring, and Locating for Non-Hard-Hat Use in Construction. *J. Constr. Eng. Manag.* **2019**, *145*, 04019006. [CrossRef]
14. 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]
15. Wu, H.; Zhao, J. An Intelligent Vision-Based Approach for Helmet Identification for Work Safety. *Comput. Ind.* **2018**, *100*, 267–277. [CrossRef]

16. Cai, D.; Bamisile, O.; Zhang, W.; Chang, Z.; Li, J.; Zhang, Z.; Wu, J.; Huang, Q. Anti-Occlusion Multi-Object Surveillance Based on Improved Deep Learning Approach and Multi-Feature Enhancement for Unmanned Smart Grid Safety. *Energy Rep.* **2023**, *9*, 594–603. [CrossRef]
17. Wu, H.; Zhao, J. Automated Visual Helmet Identification Based on Deep Convolutional Neural Networks. In *Computer Aided Chemical Engineering*; Eden, M.R., Ierapetritou, M.G., Towler, G.P., Eds.; 13 International Symposium on Process Systems Engineering (PSE 2018); Elsevier: Amsterdam, The Netherlands, 2018; Volume 44, pp. 2299–2304.
18. Chan, A.; Wong, F.; Hon, C.; Choi, T. A Bayesian Network Model for Reducing Accident Rates of Electrical and Mechanical (E&M) Work. *Int. J. Environ. Res. Public Health* **2018**, *15*, 2496. [CrossRef] [PubMed]
19. Jalil Al-Bayati, A.; Rener, A.T.; Listello, M.P.; Mohamed, M. PPE Non-Compliance among Construction Workers: An Assessment of Contributing Factors Utilizing Fuzzy Theory. *J. Saf. Res.* **2023**, *85*, 242–253. [CrossRef] [PubMed]
20. Ahmed, M.I.B.; Saraireh, L.; Rahman, A.; Al-Qarawi, S.; Mhran, A.; Al-Jalaoud, J.; Al-Mudaifer, D.; Al-Haidar, F.; AlKhulaifi, D.; Youldash, M.; et al. Personal Protective Equipment Detection: A Deep-Learning-Based Sustainable Approach. *Sustainability* **2023**, *15*, 13990. [CrossRef]
21. Bouhayane, A.; Charouh, Z.; Ghogho, M.; Guennoun, Z. A Swin Transformer-Based Approach for Motorcycle Helmet Detection. *IEEE Access* **2023**, *11*, 74410–74419. [CrossRef]
22. Lee, J.-Y.; Choi, W.-S.; Choi, S.-H. Verification and Performance Comparison of CNN-Based Algorithms for Two-Step Helmet-Wearing Detection. *Expert Syst. Appl.* **2023**, *225*, 120096. [CrossRef]
23. 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]
24. Du, Y.; Liu, X.; Yi, Y.; Wei, K. Optimizing Road Safety: Advancements in Lightweight YOLOv8 Models and GhostC2f Design for Real-Time Distracted Driving Detection. *Sensors* **2023**, *23*, 8844. [CrossRef]
25. Zhao, B.; Lan, H.; Niu, Z.; Zhu, H.; Qian, T.; Tang, W. Detection and Location of Safety Protective Wear in Power Substation Operation Using Wear-Enhanced YOLOv3 Algorithm. *IEEE Access* **2021**, *9*, 125540–125549. [CrossRef]
26. Fang, J.; Li, X. Object Detection Related to Irregular Behaviors of Substation Personnel Based on Improved YOLOv4. *Appl. Sci.* **2022**, *12*, 4301. [CrossRef]
27. 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]
28. Lo, J.-H.; Lin, L.-K.; Hung, C.-C. Real-Time Personal Protective Equipment Compliance Detection Based on Deep Learning Algorithm. *Sustainability* **2022**, *15*, 391. [CrossRef]
29. Qiao, R.; Cai, C.; Meng, H.; Wu, K.; Wang, F.; Zhao, J. An Improved Personal Protective Equipment Detection Method Based on YOLOv4. *Multimed. Tools Appl.* **2024**, 1–19. [CrossRef]
30. Gallo, G.; Rienzo, F.D.; Garzelli, F.; Ducange, P.; Vallati, C. A Smart System for Personal Protective Equipment Detection in Industrial Environments Based on Deep Learning at the Edge. *IEEE Access* **2022**, *10*, 110862–110878. [CrossRef]
31. Wu, B.; Pang, C.; Zeng, X.; Hu, X. ME-YOLO: Improved YOLOv5 for Detecting Medical Personal Protective Equipment. *Appl. Sci.* **2022**, *12*, 11978. [CrossRef]
32. Zeng, L.; Duan, X.; Pan, Y.; Deng, M. Research on the Algorithm of Helmet-Wearing Detection Based on the Optimized Yolov4. *Vis. Comput.* **2023**, *39*, 2165–2175. [CrossRef]
33. Zhao, L.; Tohti, T.; Hamdulla, A. BDC-YOLOv5: A Helmet Detection Model Employs Improved YOLOv5. *Signal Image Video Process.* **2023**, *17*, 4435–4445. [CrossRef]
34. Li, Z.; Xie, W.; Zhang, L.; Lu, S.; Xie, L.; Su, H.; Du, W.; Hou, W. Toward Efficient Safety Helmet Detection Based on YoloV5 with Hierarchical Positive Sample Selection and Box Density Filtering. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 2508314. [CrossRef]
35. Hayat, A.; Morgado-Dias, F. Deep Learning-Based Automatic Safety Helmet Detection System for Construction Safety. *Appl. Sci.* **2022**, *12*, 8268. [CrossRef]
36. Han, J.; Liu, Y.; Li, Z.; Liu, Y.; Zhan, B. Safety Helmet Detection Based on YOLOv5 Driven by Super-Resolution Reconstruction. *Sensors* **2023**, *23*, 1822. [CrossRef] [PubMed]
37. Chen, W.; Li, C.; Guo, H. A Lightweight Face-Assisted Object Detection Model for Welding Helmet Use. *Expert Syst. Appl.* **2023**, *221*, 119764. [CrossRef]
38. Wang, Z.; Wu, Y.; Yang, L.; Thirunavukarasu, A.; Evison, C.; Zhao, Y. Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. *Sensors* **2021**, *21*, 3478. [CrossRef] [PubMed]
39. Du, Y.; Liu, X.; Yi, Y.; Wei, K. Incorporating Bidirectional Feature Pyramid Network and Lightweight Network: A YOLOv5-GBC Distracted Driving Behavior Detection Model. *Neural Comput. Appl.* **2023**, *36*, 9903–9917. [CrossRef]
40. Nguyen, N.-T.; Tran, Q.; Dao, C.-H.; Nguyen, D.A.; Tran, D.-H. Automatic Detection of Personal Protective Equipment in Construction Sites Using Metaheuristic Optimized YOLOv5. *Arab. J. Sci. Eng.* **2024**, 1–19. [CrossRef]
41. Liu, Y.; Wang, J. Personal Protective Equipment Detection for Construction Workers: A Novel Dataset and Enhanced YOLOv5 Approach. *IEEE Access* **2024**, *12*, 47338–47358. [CrossRef]
42. Samma, H.; Al-Azani, S.; Luqman, H.; Alfarraj, M. Contrastive-Based YOLOv7 for Personal Protective Equipment Detection. *Neural Comput. Appl.* **2024**, *36*, 2445–2457. [CrossRef]
43. Wang, Z.; Cai, Z.; Wu, Y. An Improved YOLOX Approach for Low-Light and Small Object Detection: PPE on Tunnel Construction Sites. *J. Comput. Des. Eng.* **2023**, *10*, 1158–1175. [CrossRef]

44. Chen, J.; Zhu, J.; Li, Z.; Yang, X. YOLOv7-WFD: A Novel Convolutional Neural Network Model for Helmet Detection in High-Risk Workplaces. *IEEE Access* **2023**, *11*, 113580–113592. [CrossRef]
45. Lee, Y.-R.; Jung, S.-H.; Kang, K.-S.; Ryu, H.-C.; Ryu, H.-G. Deep Learning-Based Framework for Monitoring Wearing Personal Protective Equipment on Construction Sites. *J. Comput. Des. Eng.* **2023**, *10*, 905–917. [CrossRef]
46. 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. [CrossRef]
47. Di, B.; Xiang, L.; Daoqing, Y.; Kaimin, P. MARA-YOLO: An Efficient Method for Multiclass Personal Protective Equipment Detection. *IEEE Access* **2024**, *12*, 24866–24878. [CrossRef]
48. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS-Improving Object Detection with One Line of Code. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; IEEE: New York, NY, USA, 2017; pp. 5562–5570.
49. Lin, T.-Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 30TH IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 21–26 July 2017; IEEE: New York, NY, USA, 2017; pp. 936–944.
50. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; IEEE: New York, NY, USA, 2018; pp. 8759–8768.
51. Guo, M.-H.; Lu, C.-Z.; Hou, Q.; Liu, Z.; Cheng, M.-M.; Hu, S.-M. SegNeXt: Rethinking Convolutional Attention Design for Semantic Segmentation. Available online: https://arxiv.org/abs/2209.08575v1 (accessed on 11 November 2023).
52. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542.
53. Liu, S.T.; Huang, D.; Wang, Y.H. Receptive field block net for accurate and fast object detection. In Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 8–14 September 2018.
54. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. GhostNet: More Features from Cheap Operations. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 1577–1586.