processes

MDPI

Article

Dual-Path Attention Network for Multi-State Safety Helmet Identification in Complex Power Scenarios

Wei Li [1], Rong Jia [1], Xiangwu Chen [2], Ge Cao [1,*] and Ziyan Zhao [3]

[1] School of Electrical Engineering, Xi'an University of Technology, Xi'an 710054, China; liweiv1998@163.com (W.L.); jiarong@xaut.edu.cn (R.J.)
[2] State Grid Shaanxi Marketing Service Center (Metrology Center), State Grid Shaanxi Electric Power Company Limited, Xi'an 710048, China; chenxiangwu@yx.sn.sgcc.com.cn
[3] School of Electronic Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710061, China; ziyan8026@163.com
* Correspondence: gcao@xaut.edu.cn; Tel.: +86-151-2923-9213

Abstract

The environment of the power operation site is complex and changeable, and the accurate identification of the wearing status of workers' safety helmets is significant to ensure personal safety and the stable operation of the power system. Existing research suffers from high rates of missed detections and limited ability to discriminate fine-grained states, especially the identification of "wrongly wearing" states. Therefore, this paper proposes an intelligent identification method of safety helmet status for power workers based on a dual-path attention network. We embed the convolutional block attention module (CBAM) in the two paths of the backbone and neck layers of YOLOv5 and enhance the feature focusing ability of the key areas of the helmet through the channel-spatial attention coordination, so as to suppress the interference of complex background. In addition, a special dataset covering power scenarios is constructed, including fine-grained state annotation under various lighting, different poses, and occlusion conditions to improve the generalization of the model. Finally, the proposed method is applied to the images of the electric power operation site for experimental verification. The experimental results show that the proposed YOLO-CBAM achieves an outstanding mean average precision of 98.81% for identifying all helmet states, providing reliable technical support for intelligent safety monitoring.

Keywords: power production safety; personal protective equipment; helmet identification; object detection; attention mechanism

check for updates

Academic Editor: Hamed Badihi

Received: 3 July 2025
Revised: 24 August 2025
Accepted: 26 August 2025
Published: 28 August 2025

Citation: Li, W.; Jia, R.; Chen, X.; Cao, G.; Zhao, Z. Dual-Path Attention Network for Multi-State Safety Helmet Identification in Complex Power Scenarios. *Processes* **2025**, *13*, 2750. https://doi.org/10.3390/pr13092750

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The power industry is an important foundation for economic development and people's livelihood protection. Electrical power production and operation environment is complex and changeable, often involving high-altitude, high-voltage, strong electromagnetic field, and other high-risk scenarios [1]. The personal safety of workers is the primary premise to ensure the stable operation of power system. Among the numerous safety protection measures, a safety helmet, correctly worn, is a key barrier that prevents injuries such as falling object blows, electric shocks, and collisions to the head [2,3]. However, in actual production sites, failure to wear or wrongly wear safety helmets due to negligence, fatigue, or illegal operations still occurs from time to time, posing a major safety hazard [4,5]. "Wrongly wearing" mainly refers to a safety helmet worn on the head but

not properly fastened with the mandibular band, and the safety helmet is severely skewed, which affects protection. This is significantly different from not wearing a helmet at all but less different than wearing a helmet correctly. Therefore, the development of an efficient and accurate automatic identification method for the safety helmet status (correctly wearing/not wearing/wrongly wearing) of power workers has important practical significance and application value for actively warning violations, improving the intelligent level of on-site safety supervision, and ensuring the safety of workers.

Recently, with the rapid development of big data processing and artificial intelligence technology, vision-based helmet detection and identification methods have made significant progress [6,7]. The existing methods can be divided into two main categories: traditional image processing methods and deep-learning-based object detection methods. Traditional image processing methods often rely on manually designed features (e.g., color features, shape features, texture features, etc.) combined with classifiers (e.g., SVM, AdaBoost, etc.) for helmet identification [8,9]. Sun et al. [10] proposed a helmet monitoring method combining multi-feature fusion and support vector machine (SVM). This method uses principal component analysis (PCA) and a SVM model based on Bayesian optimization to identify the helmet. Yue et al. [11] proposed an enhanced IBRFs algorithm for the wearing state detection of safety helmets. This method uses a histogram of oriented gradient (HOG) to extract the feature and then constructs a classifier for helmet detection based on random ferns. The traditional method has a relatively small amount of calculation, but, in complex backgrounds, lighting changes, angle changes, and the variety of helmet colors, the robustness and generalization ability are often insufficient, and the false detection rate and missed detection rate are high. The object detection method based on deep learning, represented by convolutional neural network (CNN), has gradually become the mainstream method of helmet identification due to its powerful feature extraction and end-to-end learning ability [12–14]. Representative models such as Faster R-CNN, SSD, and YOLO series are widely used in safety helmet detection tasks due to their good balance in accuracy and speed [15,16]. In order to solve the problem of difficult detection caused by the small area of the helmet in the image, Yang et al. [17] proposed a new algorithm called Yolo-Helmet. By expanding the detection layer and introducing deformable convolution, the algorithm enhances small-object detection sensitivity, and Yolo-Helmet achieves 93.1% mAP. Han et al. [18] proposed a detection algorithm based on single shot multibox detector (SSD) in response to the problem of the low accuracy of existing safety helmet detection methods. Using multi-scale features and adaptive anchors for scale robustness, the algorithm achieves 88.1% mAP on helmet data. Ji et al. [19] proposed a receptive, field-enhanced, lightweight safety helmet detection algorithm called YOLOv5s-CR. The method adopts a lightweight backbone, a feature fusion network, and a small object detection layer to improve accuracy and reduce parameters and introduces a receptive field enhancement module (RFEM) to improve multi-scale detection. Chen et al. [20] proposed a safety helmet-wearing detection algorithm based on the improved YOLOv7. The method uses 16-channel features and structured pruning to achieve rapid detection of helmets. Zhao et al. [21] proposed possible improvements to YOLOv5 and named BDC-YOLOv5 to improve the accuracy of small target recognition. While recent advancements like the lightweight LG-YOLOv8 [22] and the multi-model integrated approach based on YOLOv9 [23] have made significant progress in improving efficiency and handling occlusions for safety helmet detection, their focuses on extreme lightweighting or systemic complexity only addresses the issue of whether the helmet is worn or not. By training these models on general data sets or helmet data sets, researchers have achieved significantly better performance than traditional methods [24,25]. Although the helmet identification method based on deep learning has made great progress,

the existing methods still face many challenges and limitations in the complex power operation site:

1. Complex background interference: The background elements of the power scenarios are diverse and similar, which is easy to cause misjudgment of the model. Strong noise in the background also interferes with feature extraction.
2. Inadequate accuracy of status identification: Most of the existing studies focus on the two categories (wearing/not wearing), and there are relatively few studies on the identification of the key dangerous state of "wrongly wearing", and the accuracy is not high.
3. Inadequate scenario generalization capability: The performance of the model trained for specific scenarios may be significantly reduced in the face of different lighting conditions, weather, or different substation and line environments.

Aiming at the above problems, especially in order to improve the accuracy and generalization of safety helmet status identification (correctly wearing/not wearing/wrongly wearing) in complex power operation scenarios, this paper proposes a safety helmet status identification method for power workers based on a dual-path attention network named YOLO-CBAM. The main contributions of this paper are as follows:

1. To address the challenge of complex background interference, we introduce the convolutional block attention module (CBAM) into the YOLOv5 network, creating the YOLO-CBAM architecture. The design employs coordinated channel-spatial attention mechanisms to dynamically focus on critical head and helmet regions while actively suppressing irrelevant and noisy background elements prevalent in power operation site.
2. To overcome the inadequate scenario generalization capability, we construct a high-quality special dataset for safety helmet status identification in diverse power operation environments. This dataset captures a wide spectrum of real-world challenges, including varied lighting condition, numerous personnel poses, and all critical helmet states (correctly wearing/not wearing/wrongly wearing).
3. To tackle the inadequate accuracy of status identification, particularly for the critical "wrongly wearing" state, we apply the proposed method to electrical power operation scenarios. Extensive experiments show YOLO-CBAM achieves an outstanding mean average precision of 98.81% for identifying all three helmet states, with particular emphasis on the high accuracy attained for the critical "wrongly wearing" state.

The structure of this paper is organized as follows: Section 2 describes the helmet status identification method based on a dual-path attention network. Section 3 shows the experimental results and verifies the proposed method. Section 4 summarizes the study.

2. Proposed Method

2.1. The Overall Structure of Proposed Method

Aiming at the requirement of helmet wearing status identification in the electrical power operation scenario, especially to deal with the challenges of complex background, small object, occlusion, and so on, this paper proposes a safety helmet status identification method for power workers based on a dual-path attention network named YOLO-CBAM. The core of this method is to effectively integrate the convolutional block attention module (CBAM) into the backbone (feature extraction network) and neck (feature fusion network) layers of YOLOv5 model. YOLOv5 is known for its high maturity and controllability, and the relatively stable YOLOv5 framework allows us to clearly analyze the effects of CBAM modules introduced in different locations. The introduction of CBAM enables the model to adaptively strengthen the discriminative feature channels and spatial regions related to the

helmet, suppress irrelevant background information, and significantly improve the quality of multi-scale feature fusion and the ability to focus on key objects, so as to optimize the final identification accuracy. The network framework of the method is shown in Figure 1, which is mainly composed of the following three parts: (1) backbone network; (2) neck network; and (3) prediction identification module. Firstly, the input image is extracted through the backbone network, and the backbone network gradually extracts the basic features and semantic features of the image through multi-layer convolution operation. Then, the extracted feature map is transferred to the neck network, which integrates features of different scales through feature pyramid structure (FPN) and path aggregation network (PAN), so as to enhance the detection ability of the network for different sizes of objects. Finally, the prediction identification module completes the object classification and boundary box regression based on the fused feature map and realizes the accurate identification of the wearing status of the safety helmet of the power workers.

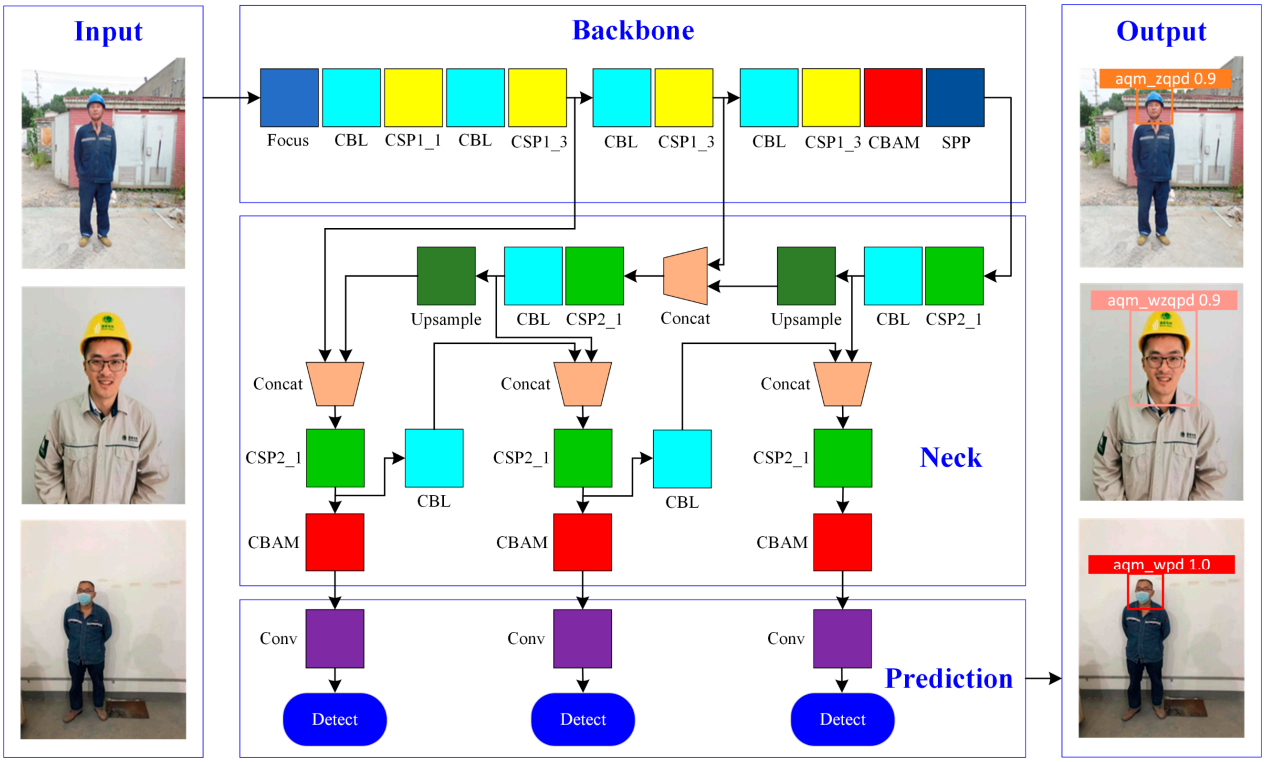


Figure 1. Framework of safety helmet status identification method based on YOLO-CBAM.

2.2. Convolutional Block Attention Module

CBAM is a lightweight and efficient attention module, which sequentially applies the channel attention module (CAM) and spatial attention module (SAM) to learn the importance weight of the feature map in the channel dimension and spatial dimension [26]. The structure of CBAM is shown in Figure 2. The CAM is used to model the dependencies between channels and generate the channel attention graph. The SAM is used to model the dependencies between spatial locations and generate the spatial attention map. The input features $F \in \mathbb{R}^{C \times H \times W}$ first generate a channel attention map $M_c \in \mathbb{R}^{C \times 1 \times 1}$ through the CAM and then generate a spatial attention map $M_s \in \mathbb{R}^{1 \times H \times W}$ through the SAM. The final output feature F'' is the result of multiplying the input feature by element of the two attention graphs. The calculation process is as follows:

$$\begin{cases} F' = M_c(F) \otimes F \\ F'' = M_s(F') \otimes F' \end{cases} \quad (1)$$

where ⊗ represents element level multiplication. F' means channel refined feature. F'' means final refined feature.

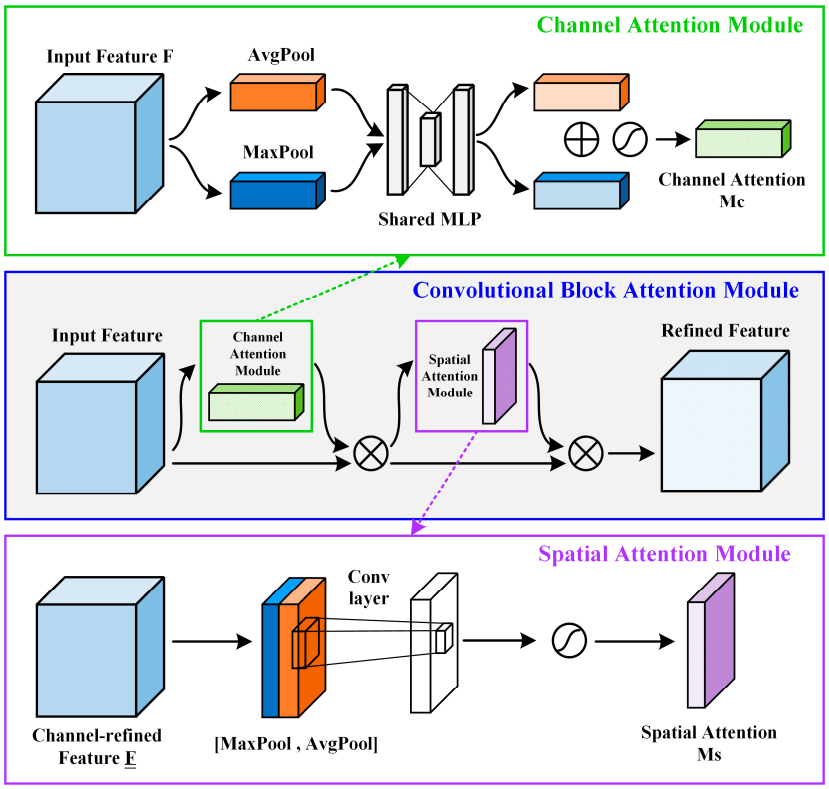


Figure 2. Structure diagram of the convolutional block attention module (CBAM) [26].

The channel attention module is used to learn the relative importance of different characteristic channels. Helmet status identification may be more dependent on some color, texture, or semantic feature channels. For the input feature F, the global average pooling (GAP) and global max pooling (GMP) are performed simultaneously to obtain two C × 1 × 1 channel descriptors F^c_{avg} and F^c_{max}. The two descriptors are sent to a shared multilayer perceptron (MLP, which usually contains a dimension reduction layer and a dimension increase layer), and then the two vectors output by the MLP are added element by element. Finally, the channel attention map M_C is generated through the sigmoid function. The calculation process is as follows:

$$\begin{aligned} M_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma\left(W_1\left(W_0\left(F^c_{\text{avg}}\right)\right) + W_1(W_0(F^c_{\text{max}}))\right) \end{aligned} \tag{2}$$

where σ is the sigmoid function, which is used to normalize the value of the attention map to the range of [0,1]. $AvgPool(F)$ and $MaxPool(F)$ represent global average pool and global max pool operations, respectively. MLP represents a multilayer perceptron, which is usually composed of two fully connected layers. $W_0 \in \mathbb{R}^{C/r \times C}$ and $W_1 \in \mathbb{R}^{C \times C/r}$ are the weights of MLP. The channel attention acts as a feature filter: amplifying critical channels (e.g., helmet yellow) while suppressing noise (e.g., background gray).

The spatial attention module is used to learn the importance of different spatial positions in the feature map. The safety helmet is usually located in the area of the worker's head, and this module helps to focus on this area. For the channel refined feature F', the average pooling and max pooling are performed along the channel dimension to obtain two channel descriptors of 1 × H × W F^s_{avg} and F^s_{max}. After splicing the results,

a 7×7 convolutional layer is passed through a sigmoid function to generate the spatial attention map M_S. The calculation process is as follows:

$$\begin{aligned} M_s(F') &= \sigma\left(f^{7\times7}([AvgPool(F'); MaxPool(F')])\right) \\ &= \sigma\left(f^{7\times7}\left(\left[F^s_{\text{avg}}; F^s_{\text{max}}\right]\right)\right) \end{aligned} \tag{3}$$

where $f^{7\times7}$ indicates convolution operation. The spatial attention works as a regional spotlight: highlighting helmet areas (even when occluded) and blurring distractions (e.g., reflective equipment).

2.3. Backbone Enhancement with CBAM

The backbone network adopts the CSPDarknet structure and integrates the gradient changes into the feature map from beginning to end through cross stage partial network (CSPNet), so as to reduce the amount of calculation and ensure the feature extraction ability of the network [27]. This paper uses two different CSP structures. CSP1_X structure is used for backbone network, and CSP2_X structure is used for neck network. CSP1_X reduces the computational complexity and improves the performance of neural network by dividing each stage into two sub stages. The structure of CSP is shown in Figure 3, mainly including convolutional layer, cross stage partial connection, residual block, concatenation layer, and linear layer. As shown in the Figure 3, this structure divides the input feature map into two parts. One part is directly passed to the next order, while the other part is subject to intensive convolution processing. The two-part features are concatenated at the end of the stage, thereby achieving richer gradient combinations and reducing computational overhead.

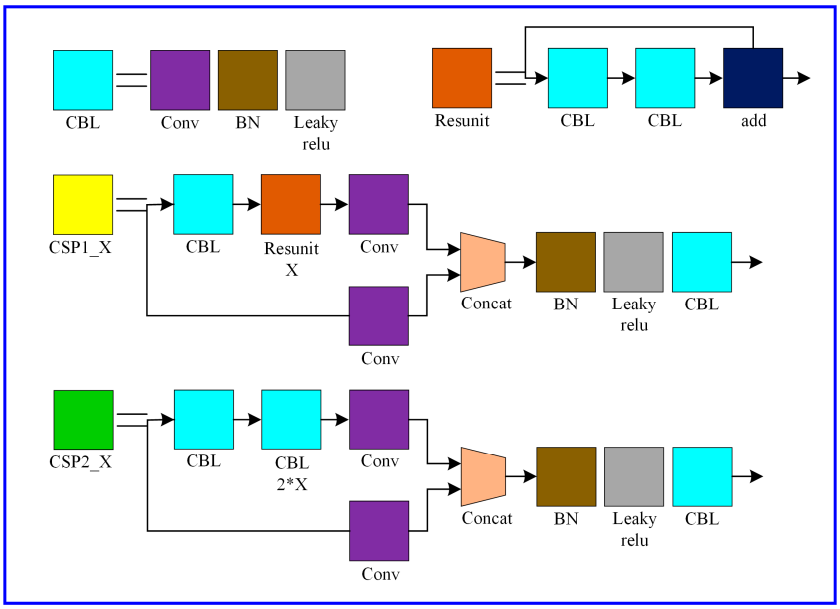


Figure 3. Structure diagram of the cross stage partial (CSP) [27].

In order to make the backbone network more focused on the key features of power workers and helmets, this paper embeds the CBAM after the last CSP1_X module. The introduction of CBAM enables the model to dynamically learn "which channel features are more important for identifying the helmet" (such as color and texture channels) and "which spatial positions in the feature map are more likely to contain the object" (such as the head area) in the feature extraction stage, giving priority to retaining key information and providing higher quality basic features for subsequent processing.

2.4. Neck Enhancement with CBAM

After collecting various features, it is necessary to fuse multiple feature information for object detection. The neck feature fusion network adopts the path aggregation network (PANet) structure, as shown in Figure 4. The network realizes multi-scale information fusion through a bidirectional feature pyramid, including three critical paths. In the top-down path, high-level semantic features (such as the overall shape of the helmet) are upsampled and fused with the underlying high-resolution features. In the bottom-up path, the underlying positioning features (such as the helmet edge) are downsampled and fused with the higher-level features. The transverse connection aligns the feature dimensions of adjacent layers through 1×1 convolution to avoid fusion distortion. PANet receives feature maps from different levels of backbone, including shallow high-resolution features and deep strong semantic features, and performs top-down and bottom-up feature fusion through upsampling, downsampling, and concatenation/addition operations to enhance the network's detection capability for multi-scale objects [28].

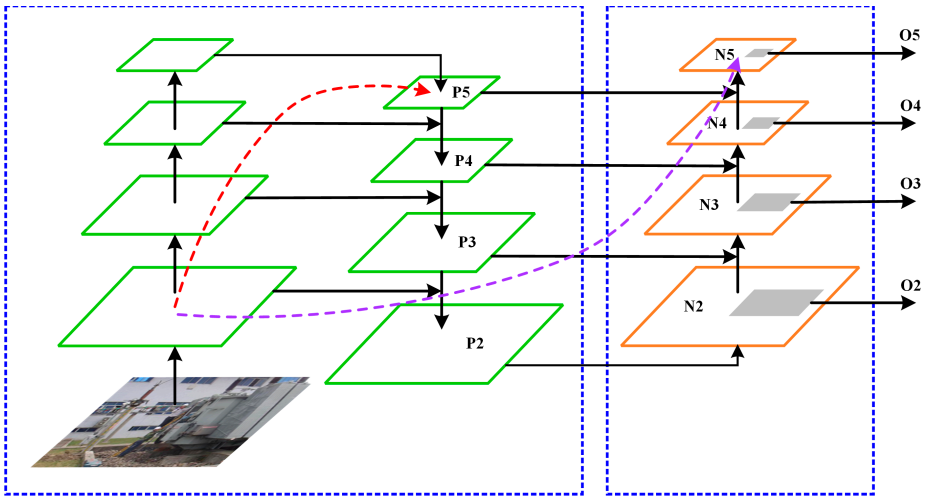


Figure 4. Structure diagram of the path aggregation network (PANet) [28].

We embed the CBAM on the key feature fusion path of neck. The CBAM module is introduced in the feature fusion stage, so that the model can optimize the selection of cross-scale features, focusing on the fine spatial details related to the helmet in the shallow features and the strong semantic information of the helmet state contained in the deep features. At the same time, it effectively suppresses the background or irrelevant object noise brought by features at different levels in the fusion process and highlights the features that are favorable to helmet detection. Through the spatial attention mechanism, it is ensured that the fused features get a significantly enhanced response in the area where the helmet is most likely to appear, such as the worker's head.

2.5. Prediction Module

The prediction module consists of three detection layers with feature maps of 80×80, 40×40, and 20×20, specifically optimized for small, medium, and large safety helmets in power scenarios. Each prediction branch contains a convolutional layer, which is used to predict bounding box coordinates (x, y, w, h), confidence score, and category probability. The prediction module adopts a decoupling head structure, which separates the object classification and bounding box regression tasks, improving the detection efficiency and accuracy. Crucially, the module directly processes feature maps enhanced by backbone's CBAM blocks (channel-optimized features) and neck's CBAM paths (spatially refined contexts), achieving high-precision fine-grained identification for helmets in complex power

environments. The model uses the high-quality feature map after backbone extraction, neck fusion, and attention mechanism enhancement to predict directly, which can more accurately complete the task of object classification and boundary box regression of safety helmets for power workers and perform the accurate identification of the wearing statuses of safety helmets.

2.6. Flow of Multi-State Safety Helmet Identification

YOLO-CBAM can realize the intelligent identification of safety helmets through multi-object detection, providing reliable technical support for intelligent safety monitoring. The specific process is shown in Figure 5, and the steps are as follows:

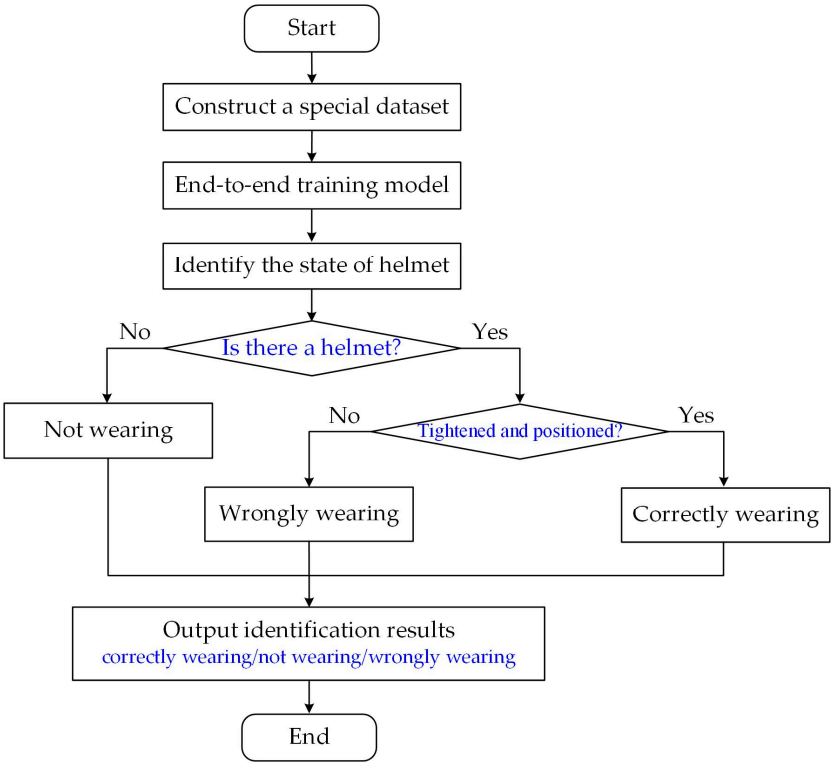


Figure 5. Flowchart of safety helmet identification based on YOLO-CBAM.

Step 1: Construct a special dataset for helmet status identification in complex power scenarios. Invite researchers to annotate images in the dataset and ask power industry experts to evaluate them.

Step 2: Use the special dataset constructed to train the YOLO-CBAM model end-to-end. When the loss function and precision tend to stabilize, the model training ends.

Step 3: Use the trained model for safety multi-state helmet identification. If there is no helmet, it is judged not wearing. Otherwise, it will be transferred to the next level of judgment.

Step 4: Continue to determine whether the helmet is correctly wearing. If the helmet's mandibular band is tightened and properly positioned, it is judged correctly wearing. Otherwise, it is judged wrongly wearing.

Step 5: Safety helmet identification results are output, including positioning box, category, and confidence level.

3. Experimental Setup and Result Analysis

3.1. Data Description

In order to verify the applicability and effectiveness of the proposed method, we constructed a special dataset for helmet status identification in complex power scenarios. We used videos of the electrical power production process collected from monitoring equipment of a power company in Xi'an China (official: http://www.sn.sgcc.com.cn/ (accessed on 9 August 2025)). These videos cover typical scenarios such as substations, distribution rooms, and user side. We extracted frames from the video and cleaned the images obtained to remove invalid samples such as high scene repetition, large-area occlusion of the lens, and no operation screens. Then the sample images are preprocessed, such as brightness and contrast enhancement, to form a special dataset for final helmet status identification. The dataset contains a total of 3000 images, of which 1000 are correctly wearing, 1000 are wrongly wearing, and 1000 are not wearing. To guarantee the generalization of the model, the images in the dataset include conventional, back-to-camera, complex outdoor, object occlusion, long-distance, fuzzy samples, and other situations, involving multiple workers and different operation types. Partial image samples of the dataset are shown in Figure 6.

We invited workers and researchers who have been engaged in electrical power production for a long time as annotators. Annotators used LabelImg 1.8.6 to add bounding boxes to all objects in the image and marked the mandibular band buckle point for the objects that are correctly wearing. When the angle between the helmet axis and the head axis is greater than 15°, the helmet is skewed and judged to be wrongly wearing. After the data annotation was completed, we invited safety experts from the power industry to review and evaluate it to ensure the high quality of the data. One image may contain multiple objects, and the total number of objects in 3000 images was 4870. The specific description of the number of images and objects for each category in the dataset is shown in Table 1. Partial image annotations in the dataset are shown in Figure 7. Finally, 70% of the dataset was selected as the training set, 10% as the validation set, and the remaining 20% as the test set. To ensure that the proportion of categories in each set was the same as the original dataset, we adopted a method of hierarchical sampling by category.

Table 1. Description of the number for each category in the dataset.

Object Types	Labels	Training Set		Validation Set		Test Set		Total	
		Images	Objects	Images	Objects	Images	Objects	Images	Objects
correctly wearing	aqm_zqpd	700	1357	100	188	200	375	1000	1920
wrongly wearing	aqm_wzqpd	700	763	100	112	200	225	1000	1100
not wearing	aqm_wpd	700	1302	100	181	200	367	1000	1850
Total	/	2100	3422	300	481	600	967	3000	4870



Figure 6. Partial image samples of the dataset for helmet status identification. (**a**) Conventional sample; (**b**) back-to-camera sample; (**c**) complex outdoor sample; (**d**) object occlusion sample; (**e**) long-distance sample; (**f**) fuzzy sample.



Figure 7. Partial image annotations in the dataset for helmet status identification. (**a**) Correctly wearing, label: aqm_zqpd; (**b**) wrongly wearing, label: aqm_wzqpd; (**c**) not wearing, label: aqm_wpd.

3.2. Experimental Environment

To test the performance of the proposed helmet state identification method, this paper conducts algorithm training and performance testing on the server. The specific configuration environment of the server is shown in Table 2. The experiment uses Linux to build the development platform and Ubuntu 20.04.6 LTS as the operating system. The model is trained and tested under the PyTorch 1.11.0 deep learning framework.

Table 2. Description of the experimental environment.

Configuration Name	Specific Information
Server type	DELL Precision T5820 GPU
CPU	i9-10980XE, 18 cores, 3.0 GHz
GPU	RTX3090, 24 GB
Memory	128 GB
Hard disk	10 T, solid-state drive (SSD)

3.3. Evaluate Metrics

This paper uses recall, precision, average precision (AP), and mean average precision (mAP) as the evaluate metric of the model [29,30]. Recall refers to the proportion of positive samples correctly detected by the model in all real positive samples, reflecting the ability of the model to find all objects. The specific calculation formula is as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

where TP (true positive) represents the number of positive cases correctly predicted by the model and FN (false negative) represents the number of negative cases incorrectly predicted by the model.

Accuracy is the proportion of true positive samples in all positive samples predicted by the model, reflecting the reliability of the prediction results. The specific calculation formula is as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

where FP (false positive) represents the number of positive cases incorrectly predicted by the model.

Average precision (AP) is the weighted average of accuracy under different recall for a single category. AP comprehensively reflects the detection performance of the category. AP can be obtained by calculating the area under the precision-recall curve for the category. When integrating the precision value on the interval of Recall$\in$[0,1], the actual calculation usually sums the sampling points. The specific calculation formula is as follows:

$$AP = \int_0^1 p(r)dr \approx \sum_{k=0}^{N} (r_{k+1} - r_k) p_{\text{interp}}(r_{k+1}) \tag{6}$$

where $p_{\text{interp}}(r) = \max_{\tilde{r} \geq r} p(\tilde{r})$ is the precision after interpolation.

Mean average precision (mAP) refers to the average AP of all detection categories, which is the core index for comprehensively evaluating the overall performance of multi-class models. The specific calculation formula is as follows:

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

where C is the total number of categories. And the value of C in this paper is 3, which are wearing, wrong wearing and not wearing. AP_i is the average precision of the i-th category.

Additionally, we also introduce mAPs with different intersection over union (IoU) thresholds as the evaluate metrics. mAP@0.5 refers to the mAP when the IoU threshold is fixed at 0.5, while mAP@[0.5:0.95] is the mAP with an IoU threshold from 0.5 to 0.95 (step size 0.05) [31].

3.4. Selection of YOLOv5 Models

To construct an efficient and accurate helmet status identification model for power workers, it is necessary to choose a detection architecture that takes into account both accuracy and speed. This section conducts benchmark testing on four versions of YOLOv5 based on the MS COCO dataset [32] and determines the optimal base model through quantitative indicator comparison. The YOLOv5 includes four different network structures: YOLOv5s (small), YOLOv5m (medium), YOLOv5l (large), and YOLOv5x (extra-large). The parameters and performance are shown in Table 3 and Figure 8.

Table 3. Comparison of different network structures of YOLOv5.

Model	Parameters (M)	Computation (GFLOPs)	mAP@0.5	mAP@[0.5:0.95]	Speed (FPS)
YOLOv5s	7.2	16.5	56.8%	37.4%	156
YOLOv5m	21.2	49.0	64.1%	45.4%	98
YOLOv5l	46.5	109.1	67.3%	49.0%	67
YOLOv5x	86.7	205.7	68.9%	50.7%	34

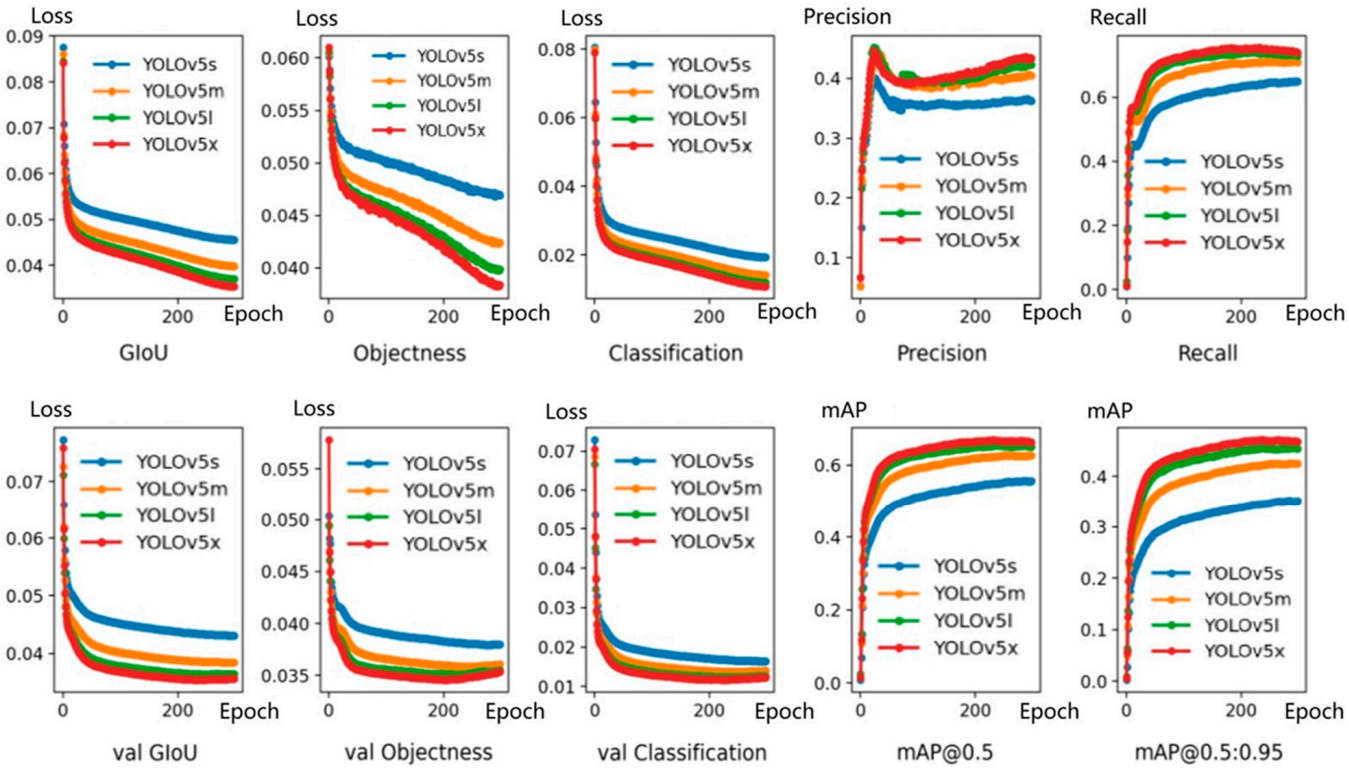


Figure 8. Performance comparison of different network structures of YOLOv5 on MS COCO. The figure shows the various loss and evaluation metric changes with epoch of the four models on the training set and the validation set.

As can be seen in Table 3, the four versions of the model exhibit significant precision-complexity trade-offs. YOLOv5s achieves 56.8% mAP@0.5 with 7.2 M parameters and 16.5 GFLOPs, and its inference speed reaches 156 FPS. In contrast, YOLOv5x achieves the highest accuracy (68.9% mAP@0.5), but the number of parameters (86.7 M) and computation (205.7 GFLOPs) increase to 12x and 12.5x, respectively, and the inference speed drops to 34 FPS. At the same time, as can be seen from Figure 6, YOLOv5s has better convergence performance compared to other models.

For the high real-time requirements of power operation scenarios and the constraints of equipment resources, YOLOv5s has become our best choice with 7.2 MB lightweight architecture and 156 FPS inference speed. Its fast convergence is suitable for the power helmet state identification dataset with limited samples, and the low computational load reserves room for the subsequent integration of attention mechanism to deal with complex background interference. Additionally, YOLOv5s has a computational efficiency ratio (mAP@0.5/GFLOPs) of 3.44, which is 4.3 times that of YOLOv5x. Therefore, based on the MS COCO benchmark and the requirements of power operation scenarios, YOLOv5s was selected as the basic model in this paper.

3.5. Model Parameter Setting and Training

In this subsection, we use the special dataset for helmet state identification in power operation scenarios constructed in Section 3.1 to train the model. In the training phase, Adam is used as the optimizer to update the parameters. The epoch is set to 300, and the batch size is set to 16. The learning rate is dynamically adjusted, the initial value is set to 1×10^{-4}, and the minimum value is set to 1×10^{-6}. To suppress over fitting and improve the generalization ability of the model in complex power scenarios, dropout is introduced into the training. Figure 9 shows the loss function change curve of the model on the training set and the validation set, as well as the dynamic change process of the key evaluation indicators on the test set, which directly reflects the optimization trajectory and stability of the model performance.

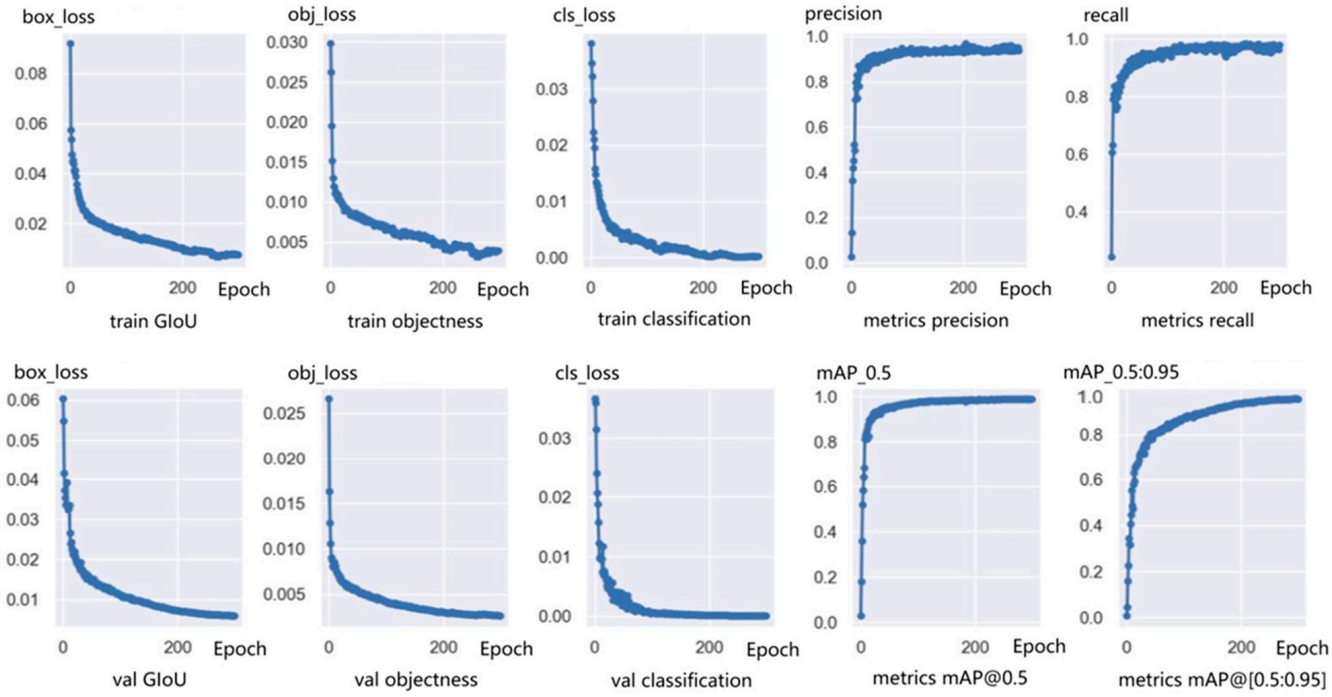


Figure 9. Results of model training and evaluation. The figure shows the various loss and evaluation metric changes with epoch of the YOLO-CBAM on the training set and the validation set.

The loss functions include bounding box loss (box_loss), confidence loss (obj_loss), and classification loss (cls_loss). The bounding box loss is used to measure the error between the prediction box and the label box and is calculated by the GIoU indicator. The confidence loss is used to calculate the confidence score of the network to determine the presence of an object. The classification loss is used to evaluate whether the anchor frame and the corresponding label are correctly classified. During model training, the validation set loss function and metrics are used to evaluate the generalization ability of the model and the optimization of the model. The core evaluation metrics include precision, recall, mAP@0.5, and mAP@[0.5:0.95]. As can be seen from Figure 9, after epoch reaches 100 generations, the model begins to converge. When epoch reaches about 200, the loss function tends to be stable. The recall of the model remains above 95%, the accuracy remains around 90%, and the mAP reaches 90%. The model shows good performance.

3.6. Ablation Experiments

In this paper, YOLOv5s is used as the main framework for helmet status identification, and the model is more focused on the detection of core objects by embedding the CBAM module, so as to improve the identification performance of the model. However, different

CBAM embedding positions also have differences in the effect on model performance. Therefore, this subsection compares the identification results of three different CBAM embedding strategies on the basis of YOLOv5s. The evaluate metrics are the AP and the FPS. The identification results of models under different CBAM embedding strategies are shown in Table 4. In the table, YOLOv5s is the baseline, and CBAM is not embedded. YOLOv5s-m1 only embeds CBAM in backbone, YOLOv5s-m2 only embeds CBAM in neck, YOLO-CBAM is the method proposed in this paper, and CBAM is embedded in both backbone and neck.

Table 4. Results of ablation experiments. The up arrow indicates an improvement in performance compared to the baseline model, and the down arrow indicates a decrease in performance compared to the baseline model.

Model	Location of CBAM	AP/% (↑Improvement)			Speed (FPS)
		Correctly Wearing	Wrongly Wearing	Not Wearing	
YOLOv5s	None	96.20	89.40	94.30	142.00
YOLOv5s-m1	Backbone only	97.63 (↑1.43)	93.85 (↑4.45)	96.82 (↑2.52)	131.50 (↓7.39%)
YOLOv5s-m2	Neck only	97.05 (↑0.85)	95.47 (↑6.07)	95.12 (↑0.82)	126.80 (↓10.70%)
YOLO-CBAM	Dual path	99.08 (↑2.88)	98.76 (↑9.36)	98.31 (↑4.01)	115.30 (↓18.80%)

As can be seen from the Table 4, the AP values for embedding CBAM only in backbone or only in neck are higher than the baseline but lower than dual-path embedded YOLO-CBAM. YOLOv5s-m1 (backbone only) is most effective for not wearing identification, because the detection of the head with or without a helmet relies more on overall feature extraction rather than local details. YOLOv5s-m2 (neck only) is the most effective for wrongly wearing identification, with an AP value increase by 6.07% compared to the baseline. This verifies the advantages of the spatial attention mechanisms in capturing subtle features such as helmet skew and mandibular bands. The AP value of wrongly wearing reaches 98.76% after dual-path embedded YOLO-CBAM, and the AP value increases by 9.36% compared with the baseline. This shows that the backbone's channel attention mechanism supplements global features such as color texture, which complements neck's spatial positioning. In terms of identification speed, compared with the baseline, the FPS of YOLOv5s-m1 decreases by 7.39%, and the FPS of YOLOv5s-m2 decreases by 10.70%. The FPS of the dual-path embedded YOLO-CBAM is 115.30, a decrease of 18.80% compared to the baseline. Although the FPS of the model after embedding CBAM has suffered a certain loss, it still meets the actual identification requirements.

3.7. Comparative Experiments

This subsection analyzes the performance of models under different confidence levels based on the precision-confidence threshold curve, the recall-confidence threshold curve, and the precision-recall curve and determines the optimal detection confidence threshold. Figure 10 is the precision-confidence threshold curve, showing the identification precision of the model at different confidence levels. As can be seen from Figure 10, the precision increases monotonically with the increase in the confidence threshold. With the increase in the confidence threshold, the quality of the prediction box retained by the model is improved, and the precision gradually increases. At this point, a large number of true objects are discarded because the confidence in the prediction is not high enough, resulting in a decrease in recall. Figure 11 shows the recall-confidence threshold curve, revealing the impact of different confidence levels on the recall of the model. As can be seen from Figure 11, the recall tends to decrease monotonically with the increase in the confidence threshold. As the confidence threshold decreases, the number of objects detected by the

model increases, and the recall increases. At this point, the precision of the model is low and contains a large number of false detections. Figure 12 shows the precision-recall curve, which visually illustrates the core trade-off relationship between precision and recall. The closer the value of the curve area is to 1, the better the performance of the model, which means that the model can maintain high precision despite high recall.

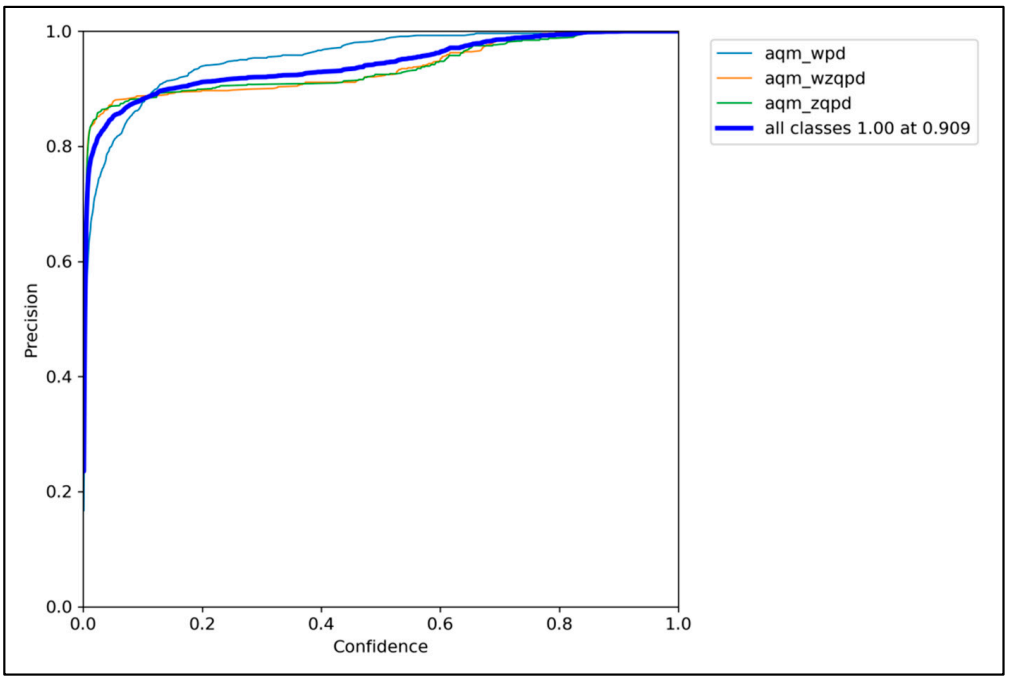


Figure 10. Precision-confidence threshold curve.

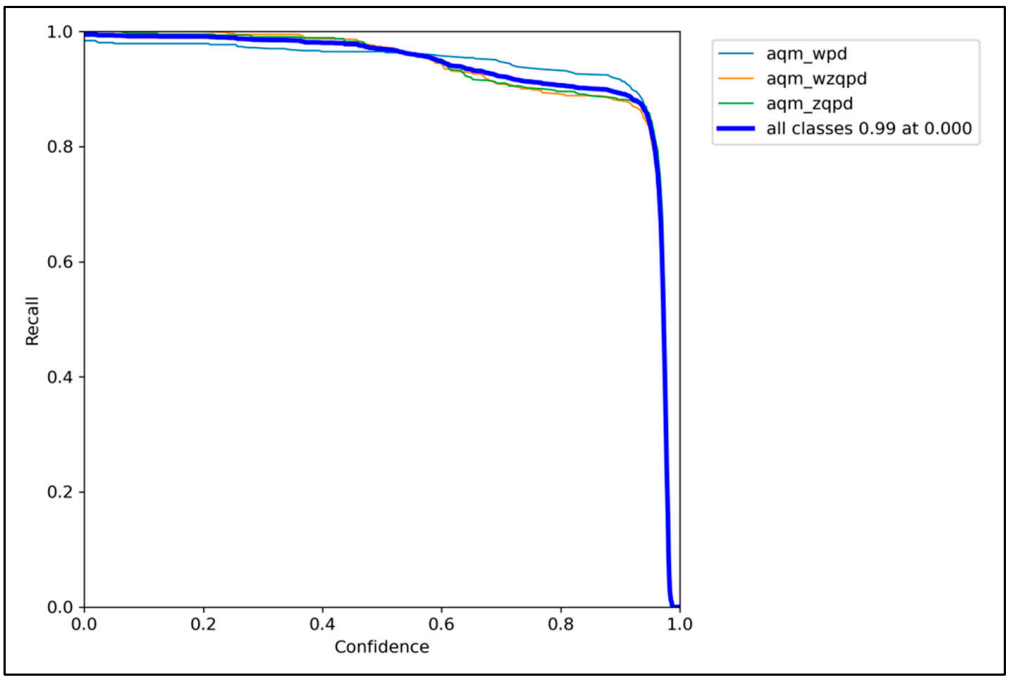


Figure 11. Recall-confidence threshold curve.

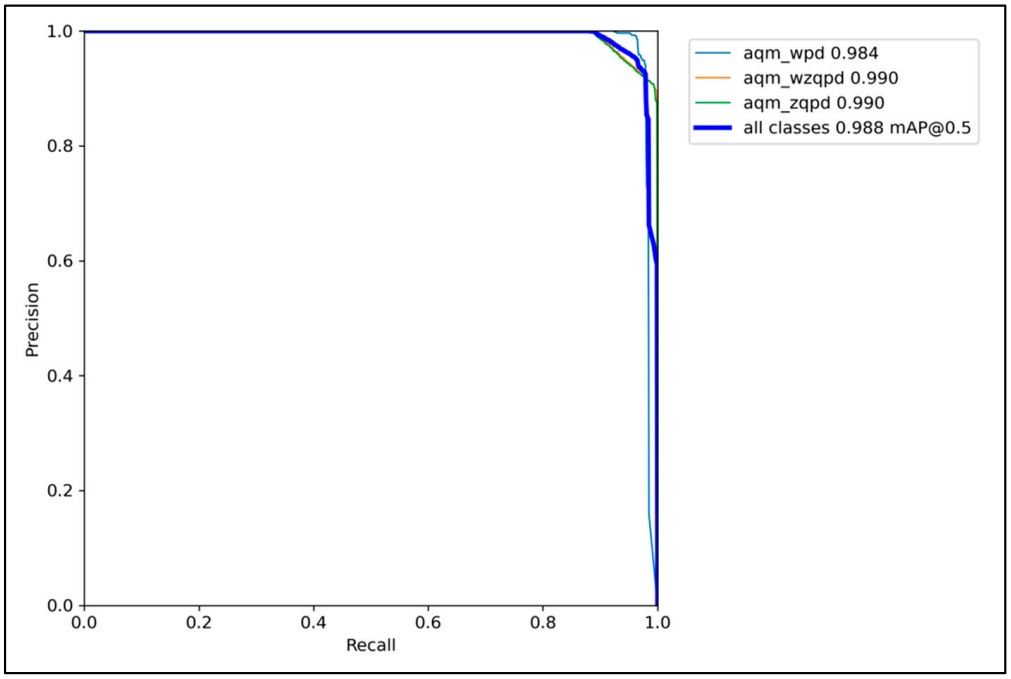


Figure 12. Precision-recall curve.

According to Figures 10 and 11, when the confidence is greater than 0.2, the change of precision gradually slows down, while the recall maintains a high value between the confidence of 0.2 and 0.3. Therefore, this paper sets the confidence threshold to 0.25, which can better balance the precision and recall. As can be seen from Figure 12, when the confidence threshold is 0.25, a curve area value close to 1 can be obtained, indicating that the model performs better.

To show the significant advantages of the YOLO-CBAM model in helmet status identification, we choose typical object detection models as comparison. The comparison models include SSD, Faster R-CNN, Mask R-CNN, and YOLOv5s, among which YOLOv5s is the baseline model in this paper. The results of the comparative experiments are shown in Table 5. It can be seen from Table 5 that, in the detection of wrongly wearing, the AP value of YOLO-CBAM at 98.76% is much higher than that of the contrast model (9.36% higher than that of the suboptimal YOLOv5s), which directly verifies the accurate capture ability of CBAM for fine-grained features. At the same time, the model achieves a 98.31% AP value (6.44% ahead of Mask R-CNN) in the state of not wearing, which proves that CBAM effectively suppressed the false detection interference with similar backgrounds. The rigorous index mAP@[0.5:0.95] reached 84.74% (8.64% higher than that of YOLOv5s), reflecting the strong adaptability of the model to complex scenarios such as occlusion and small objects. The improvement was mainly due to the optimization of dual path CBAM.

Table 5. Comparison results of helmet status identification of different models.

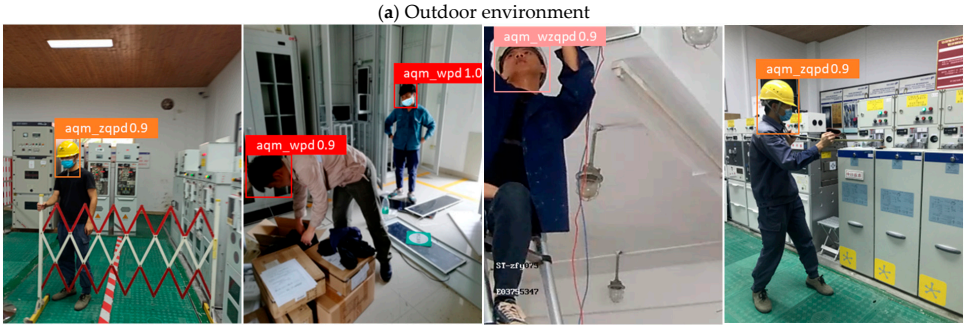
Model	AP/%			mAP@0.5/%	mAP@[0.5:0.95]/%
	Correctly Wearing	Wrongly Wearing	Not Wearing		
SSD	86.30	75.87	82.74	81.58	58.23
Faster R-CNN	92.13	84.72	90.49	89.12	67.78
Mask R-CNN	93.77	86.54	91.87	90.69	70.34
YOLOv5s	96.20	89.40	94.30	93.30	76.10
YOLO-CBAM	99.08	98.76	98.31	98.81	84.74

3.8. Identification Results

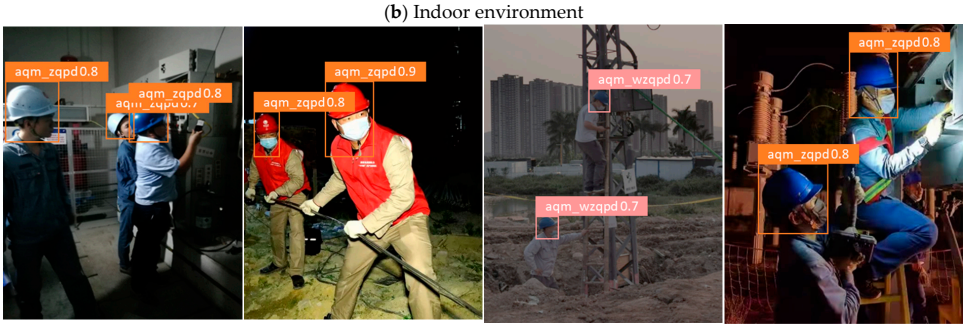
To visually demonstrate the effect of the method on the safety helmet identification state for power workers, part of the identification results in the test set are visualized in Figure 13. As can be seen from Figure 13, the proposed method can accurately identify the state of the helmet (wearing/wrong wearing/not wearing) in complex environments such as outdoor, indoor, poor light, and small objects with dense equipment. However, the YOLO-CBAM helmet identification method proposed in this study still has performance bottlenecks in the context of large-area occlusion scenarios and small-object in complex background. It can be seen from the figure that when the area of the helmet is occupied by the equipment is greater than 50%, the attention mechanism CBAM cannot extract effective local features, and there is a situation of missed detection. Additionally, in the complex background of dense equipment, the texture feature of the helmet with a small proportion of pixels is easily disturbed by similar backgrounds, making it difficult to identify the position of the mandibular band, which leads to incorrect identification. These limitations stem from the inherent contradiction of visual detection: occlusion leads to missing features and low signal-to-noise ratio for small targets.



(**a**) Outdoor environment



(**b**) Indoor environment



(**c**) Poor light at night



(**d**) Small objects with complex backgrounds

Figure 13. The identification results of the safety helmet status of electric power workers in different operating scenarios.

By combining the attention mechanism, the model can adaptively focus on key areas and enhance its robustness to distractions such as occlusion and blur in complex scenarios. At the same time, the introduction of the special dataset makes the model perform well in the adaptability of complex scenarios and can effectively deal with the challenges of diverse environments and the size difference of the detection object, which significantly improves the generalization ability of the model. The proposed method can ensure high identification accuracy under complex working conditions and fully verify its effectiveness and stability in practical engineering applications.

4. Conclusions

In order to solve the problem of helmet status identification in complex power operation scenarios, an intelligent method based on dual-path attention network is proposed. We embed the CBAM through dual path, in which the backbone layer suppresses background interference and the neck layer enhances fine-grained feature extraction. On this basis, we combine special datasets to significantly improve the robustness of the model in multiple scenarios. Experiments show that the proposed method can effectively identify various states of safety helmets (wearing/wrong wearing/not wearing) and reaches 98.81% of mAP@0.5 on the test dataset, which significantly improves the accuracy of workers' safety helmet states in complex power environments.

Future work will introduce a cross-modal behavior analysis framework to achieve high-precision detection of various personal protective equipment, such as safety helmets, insulated gloves, insulated shoes, etc., by integrating multi-source data such as infrared and visible light. At the same time, consider deploying the algorithm at the edge to further improve the intelligent level of safety prevention and control at the work site.

Author Contributions: Conceptualization, R.J. and G.C.; Methodology, W.L.; Formal analysis, X.C.; Investigation, Z.Z.; Data curation, R.J. and W.L.; Writing—original draft preparation, W.L.; Writing—review and editing, W.L., R.J., X.C., G.C., and Z.Z.; Visualization, X.C. and Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (52407143).

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: Author Xiangwu Chen was employed by the State Grid Shaanxi Marketing Service Center (Metrology Center), State Grid Shaanxi Electric Power Company Limited. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CBAM	Convolutional block attention module
YOLO	You only look once
SVM	Support vector machine
PCA	Principal component analysis
HOG	Histogram of oriented gradient
CNN	Convolutional neural network
SSD	Single shot multibox detector
RFEM	Receptive field enhancement module
FPN	Feature pyramid structure
PAN	Path aggregation network
CAM	Channel attention module

SAM	Spatial attention module
GAP	Global average pooling
GMP	Global max pooling
MLP	Multilayer perceptron
CSPNet	Cross stage partial network
PANet	Path aggregation network
AP	Average precision
mAP	Mean average precision
IoU	Intersection over union

References

1. Brenner, B.; Cawley, J.C.; Majano, D. Electrically hazardous jobs in the US. *IEEE Trans. Ind. Appl.* **2020**, *56*, 2190–2195. [CrossRef]
2. Lee, D.; Lim, D.; Lee, J. Safety Autonomous Platform for Data-Driven Risk Management Based on an On-Site AI Engine in the Electric Power Industry. *Appl. Sci.* **2025**, *15*, 630. [CrossRef]
3. Saleh, A.; Kanaan, D.A. An Experimental Study of a Novel System Used for Cooling the Protection Helmet. *Energies* **2023**, *16*, 4046. [CrossRef]
4. He, C.; Tan, S.; Zhao, J.; Ergu, D.; Liu, F.; Ma, B.; Li, J. Efficient and Lightweight Neural Network for Hard Hat Detection. *Electronics* **2024**, *13*, 2507. [CrossRef]
5. Ahmed, M.I.; 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]
6. Al-Ali, A.; Gupta, R.; Zualkernan, I.; Das, S.K. Role of IoT technologies in big data management systems: A review and Smart Grid case study. *Pervasive Mob. Comput.* **2024**, *100*, 101905. [CrossRef]
7. Delhi, V.S.K.; Sankarlal, R.; Thomas, A. Detection of personal protective equipment (PPE) compliance on construction site using computer vision based deep learning techniques. *Front. Built Environ.* **2020**, *6*, 136. [CrossRef]
8. Wang, B.Y.; Zhang, X.H.; Wu, H.B. Detection and tracking of safety helmet in construction site. *Integr. Ferroelectr.* **2021**, *218*, 139–146. [CrossRef]
9. Li, X.; Chen, W.; Yang, W.; Wang, W.; Fan, W.; Tian, Z. Segmentation method for personnel safety helmet based on super pixel features and SVM classification. *J. China Coal Soc.* **2021**, *46*, 2009–2022.
10. Sun, X.; Xu, K.; Wang, S.; Wu, C.; Zhang, W.; Wu, H. Detection and tracking of safety helmet in factory environment. *Meas. Sci. Technol.* **2021**, *32*, 105406. [CrossRef]
11. Yue, S.; Zhang, Q.; Shao, D.; Fan, Y.; Bai, J. Safety helmet wearing status detection based on improved boosted random ferns. *Multimed. Tools Appl.* **2022**, *81*, 16783–16796. [CrossRef]
12. Liang, H.; Seo, S. UAV low-altitude remote sensing inspection system using a small target detection network for helmet wear detection. *Remote Sens.* **2022**, *15*, 196. [CrossRef]
13. Song, H.; Zhang, X.; Song, J.; Zhao, J. Detection and tracking of safety helmet based on DeepSort and YOLOv5. *Multimed. Tools Appl.* **2023**, *82*, 10781–10794. [CrossRef]
14. Yang, B.; Wang, J. An improved helmet detection algorithm based on YOLO V4. *Int. J. Found. Comput. Sci.* **2022**, *33*, 887–902. [CrossRef]
15. Jayanthan, K.S.; Domnic, S. An attentive convolutional transformer-based network for road safety. *J. Supercomput.* **2023**, *79*, 16351–16377. [CrossRef]
16. Li, X.; Hao, T.; Li, F.; Zhao, L.; Wang, Z. Faster R-CNN-LSTM construction site unsafe behavior recognition model. *Appl. Sci.* **2023**, *13*, 10700. [CrossRef]
17. Yang, G.; Hong, X.; Sheng, Y.; Sun, L. YOLO-Helmet: A novel algorithm for detecting dense small safety helmets in construction scenes. *IEEE Access* **2024**, *12*, 107170–107180. [CrossRef]
18. 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]
19. Ji, C.; Hou, Z.; Dai, W. A Lightweight Safety Helmet Detection Algorithm Based on Receptive Field Enhancement. *Processes* **2024**, *12*, 1136. [CrossRef]
20. Chen, X.; Xie, Q. Safety Helmet-Wearing Detection System for Manufacturing Workshop Based on Improved YOLOv7. *J. Sens.* **2023**, *2023*, 7230463. [CrossRef]
21. Zhao, L.; Tohti, T.; Hamdulla, A. BDC-YOLOv5: A helmet detection model employs improved YOLOv5. *Signal Image Video Process.* **2023**, *17*, 4435–4445. [CrossRef]

22. Fan, Z.; Wu, Y.; Liu, W.; Chen, M.; Qiu, Z. Lg-yolov8: A lightweight safety helmet detection algorithm combined with feature enhancement. *Appl. Sci.* **2024**, *14*, 10141. [CrossRef]
23. Zhang, Y.; Huang, S.; Qin, J.; Li, X.; Zhang, Z.; Fan, Q.; Tan, Q. Detection of helmet use among construction workers via helmet-head region matching and state tracking. *Autom. Constr.* **2025**, *171*, 105987. [CrossRef]
24. Otgonbold, M.E.; Gochoo, M.; Alnajjar, F.; Ali, L.; Tan, T.H.; Hsieh, J.W.; Chen, P.Y. SHEL5K: An extended dataset and benchmarking for safety helmet detection. *Sensors* **2022**, *22*, 2315. [CrossRef]
25. Patil, K.; Jadhav, R.; Suryawanshi, Y.; Chumchu, P.; Khare, G.; Shinde, T. HelmetML: A dataset of helmet images for machine learning applications. *Data Brief* **2024**, *56*, 110790. [CrossRef] [PubMed]
26. Su, Q.; Hamed, H.N.A.; Zhou, D. Relation explore convolutional block attention module for skin lesion classification. *Int. J. Imaging Syst. Technol.* **2025**, *35*, e70002. [CrossRef]
27. Ju, C.; Guan, C. Tensor-cspnet: A novel geometric deep learning framework for motor imagery classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *34*, 10955–10969. [CrossRef]
28. Wu, Y.; Yao, Q.; Fan, X.; Gong, M.; Ma, W.; Miao, Q. PANet: A point-attention based multi-scale feature fusion network for point cloud registration. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 2512913. [CrossRef]
29. Warhade, K.K.; Merchant, S.N.; Desai, U.B. Performance evaluation of shot boundary detection metrics in the presence of object and camera motion. *IETE J. Res.* **2011**, *57*, 461–466. [CrossRef]
30. Putra, H.A.A.; Murni, A.; Chahyati, D. Enhancing Bounding Box Regression for Object Detection: Dimensional Angle Precision IoU-Loss. *IEEE Access* **2025**, *13*, 81029–81047. [CrossRef]
31. Fu, G.; Chu, H.; Tu, X. Enhancing object detection in low-light conditions with adaptive parallel networks. *J. Electron. Imaging* **2025**, *34*, 013007. [CrossRef]
32. Tong, K.; Wu, Y. Rethinking PASCAL-VOC and MS-COCO dataset for small object detection. *J. Vis. Commun. Image Represent.* **2023**, *93*, 103830. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.