Safety Helmet Wear Detection Algorithm Based on ASG-YOLOv8s

Li-Zhen He, Zhi-Sheng Wang, Yi-Wei Duan, Jin-Hai Sa
School of Software Engineering, Xinjiang University, Xinjiang, China

***Abstract*—In the field of industrial safety, the standardised wearing of safety helmets by workers constitutes a core protective measure against head injuries. However, in industrial settings, multi-scale background interference arising from variations in monitoring distance renders traditional detection models ineffective at capturing the contour features of small-sized helmets. This study, therefore, proposes the ASG-YOLOv8s safety helmet detection network, based on YOLOv8s, to address the challenge of complex scene background interference. First, the AKC-SCAM unit is introduced within the YOLOv8 backbone network to replace certain standard convolutions. This module dynamically adjusts the sampling shape of convolutional kernels, enhancing the extraction of multi-scale defect features. Secondly, a cross-scale interaction architecture (Slim-neck) is constructed in the Neck section, employing GSConv instead of conventional convolutions. This combines with a cross-level feature pyramid to achieve cross-scale interaction between deep semantic features and shallow details. Finally, GAM attention is embedded before the multi-scale output for head detection, establishing a dual-stream attention mechanism that synergistically optimises feature response intensity for low-quality candidate boxes, while suppressing background noise interference. Experimental results demonstrate that the enhanced ASG-YOLOv8s achieves improvements of 2.54%, 2.94%, and 3.16% over the original model in Precision (P), Recall (R), and mean average precision (mAP), respectively, on the SHWD dataset.**

***Keywords*—YOLOv8; safety helmet wearing detection; slim-neck; attention mechanism**

I. Introduction

In industrial production environments, complex and hazardous work types such as elevated operations and concurrent tasks frequently occur. Among these, incidents involving falling objects, impact collisions, and collapses often result in severe property damage and casualties. Statistics indicate that within the construction industry, accidents caused by falls, collisions, and collapses account for over 50% of all annual incidents [1]. The correct wearing of approved protective equipment represents the most direct and effective method to prevent such losses. However, in practical working environments, personnel frequently neglect wearing protective gear for various reasons. Consequently, monitoring the use of protective equipment by on-site workers is an effective means of ensuring construction safety. Currently, methods for monitoring the wearing of safety helmets within industrial sites primarily include the following: manual inspections, sensor-based detection, traditional image processing recognition, and deep learning-based target recognition methods. Early approaches primarily relied on sensors to detect helmet wear. For instance, Kelm et al. [2] developed an RFID-based system to monitor personal protective equipment compliance as personnel passed through security gates at construction site entrances. Dong et al. [3] employed pressure sensors embedded within protective gear to determine whether workers were wearing it. Traditional detection methods have gradually been phased out due to their complexity in both wearability and detection.

With the advancement of image processing technology, traditional image processing algorithms have gradually been applied to the detection of workers' safety equipment usage. Zhu et al. [4] proposed a rapid and stable algorithm for identifying safety helmets within substations. This approach first detects pedestrians within the substation; then, based on human body geometric features, locates the head position among pedestrians; finally, utilising the characteristic of minimal deformation in safety helmets, employs template matching to identify helmets, thereby determining whether the worker is wearing one. Jia et al. [5] proposed an algorithm for safety helmet detection. This approach utilises a deformable part model to generate block-based local binary pattern histograms, thereby recovering information lost when solely employing gradient direction histograms as features. However, these methods demand high image clarity and involve complex computations, resulting in low efficiency and recognition accuracy. Consequently, they prove unsuitable for complex and diverse industrial environments.

In recent years, deep learning technology has advanced rapidly. Object recognition networks based on deep learning are increasingly applied in practical scenarios due to their excellent real-time performance. Deep learning-based object detection primarily falls into two categories: two-stage networks and one-stage networks. The former primarily consists of two layers: the first layer extracts candidate regions, while the second layer classifies and precisely localises these regions. Representative models include Faster R-CNN [6]. The latter comprises a single layer, integrating classification and localisation into a unified implementation. Representative models include the SSD [7] and YOLO [8] series. Due to their superior detection efficiency and real-time performance, single-stage networks are extensively employed in practical research. Zhang et al. [9] proposed a novel helmet detection algorithm based on an enhanced YOLOv5 framework. This approach redesigns bounding box sizes using the K-Means++ algorithm, matching them to corresponding feature layers, and incorporates a multispectral channel attention module. This enhances the network's ability to distinguish foreground from background and improves overall detection performance. Su et al. [10] proposed a helmet detection algorithm based on an enhanced YOLOX-s framework. This approach strengthened the model's feature representation capabilities by incorporating a small object detection head and integrating methods such as the SIoU loss function and CA attention mechanism. Han et al. [11] introduced an attention-based helmet-wearing status detection model, optimising it through a bidirectional multi-layer connection fusion feature pyramid network to enhance accuracy. Most recently, Yang et al. [12] designed an epistemic feature enhancement module to construct the backbone feature extraction network. They proposed a dual-scale feature fusion module to acquire and amplify local epistemic feature information of occluded targets, thereby improving detection accuracy for partially obscured objects. In the same year, Ding et al. [13] employed deeper convolutions to replace the small residual network in DeepSORT, enhancing appearance information extraction capabilities. Simultaneously, they added a query head to the decoupled head section to bolster detection performance. Although these approaches optimised algorithms to achieve certain improvements in model accuracy, they still suffer from issues such as high parameter counts and slow detection speeds, failing to meet edge computing requirements. Therefore, further refinements are necessary to enable models to deliver superior detection performance in complex small-object scenarios.

To address the aforementioned issues, this study proposes an enhanced safety helmet detection network, ASG-YOLOv8s, based on YOLOv8s to achieve efficient and precise safety management at the edge. The primary research contributions include:

- Introducing the AKC-SCAM module into the YOLOv8 backbone network. By dynamically adjusting the sampling shape of convolutional kernels, this enhances the extraction capability of multi-scale helmet features.
- Constructing the cross-scale interaction architecture Slim-neck within the feature fusion module. This replaces standard convolutions with GSConv and combines cross-level feature pyramids to facilitate effective interaction between deep semantic information and shallow detail features.
- Embedding the GAM attention mechanism before the multi-scale prediction layer for head detection. This establishes a dual-stream attention structure coordinating both channel and spatial attention to enhance effective feature responses while suppressing background noise.

II. Model Introduction

A. ASG-YOLOv8s Model

In the task of safety helmet detection, the YOLOv8s [14] model emerges as an ideal choice due to its lightweight design and high-efficiency performance. Compared to smaller-scale models such as YOLOv8n, YOLOv8s enhances feature reuse capabilities through its C2f [15] module. Combined with a decoupled head architecture that separates classification and regression tasks, it significantly improves detection accuracy for small-scale safety helmets. Compared to larger models like YOLOv8 l/x, it achieves a compact model size of 22-25MB by streamlining parameters and computational complexity. This strikes an efficient balance between accuracy and efficiency in industrial settings, making it well-suited for resource-constrained real-time detection requirements.

To enhance the detection performance of the YOLOv8 model in complex industrial settings, we first propose the AKC-SCAM module within the YOLOv8 backbone network to replace certain standard convolutional layers. This module integrates Adaptive Kernel Convolution (AKConv) with a Spatial Context-Aware Module (SCAM), dynamically adjusting the sampling shape of convolutional kernels to strengthen multi-scale defect feature extraction. This enhances the ability to capture edge textures of small-sized safety helmets worn by workers during bending operations. Secondly, a Slim-neck module is constructed within the Neck section, employing GSConv to replace conventional convolutions. This combines with a cross-hierarchical feature pyramid to achieve cross-scale interaction between deep semantic features and shallow-level details. Finally, a GAM module is embedded prior to the multi-scale outputs of the detection head, establishing a dual-stream attention mechanism for channels and spatial domains. This dynamically optimises the feature response intensity of low-quality candidate boxes, effectively mitigating missed detections in scenarios with dense occlusions. The overall architecture of the enhanced YOLOv8 network model is illustrated in Fig. 1.

B. AKC-SCAM Module

Whilst YOLOv8's backbone architecture enhances detection efficiency through multi-scale feature fusion and lightweight design, its core components still exhibit two pressing limitations requiring resolution: Firstly, the fixed kernel size of traditional 3×3 convolutions results in a static receptive field, rendering it ill-suited to dynamically adapt to the geometric deformation characteristics of irregular objects encountered in industrial inspection scenarios. Secondly, the consecutive stacking of Bottleneck structures within the C2f module induces parameter redundancy and a surge in computational complexity. To mitigate this phenomenon, this study proposes embedding the AKC-SCAM module—comprising AKConv (Adaptive Kernel Convolution) and SCAM (Spatial Context-Aware Module)—into the Bottleneck residual path of the C2f module, replacing the original 3×3 convolutional layer. This approach reduces model parameters while adaptively adjusting kernel sizes to extract features at different scales, thereby enhancing multi-scale feature extraction capabilities.

The core concept of the AKConv [16] module lies in efficiently extracting features through a convolutional kernel that dynamically adjusts both the number of parameters and the sampling shape, thereby enhancing network performance. The structural diagram of the AKConv module is shown in Fig. 2.

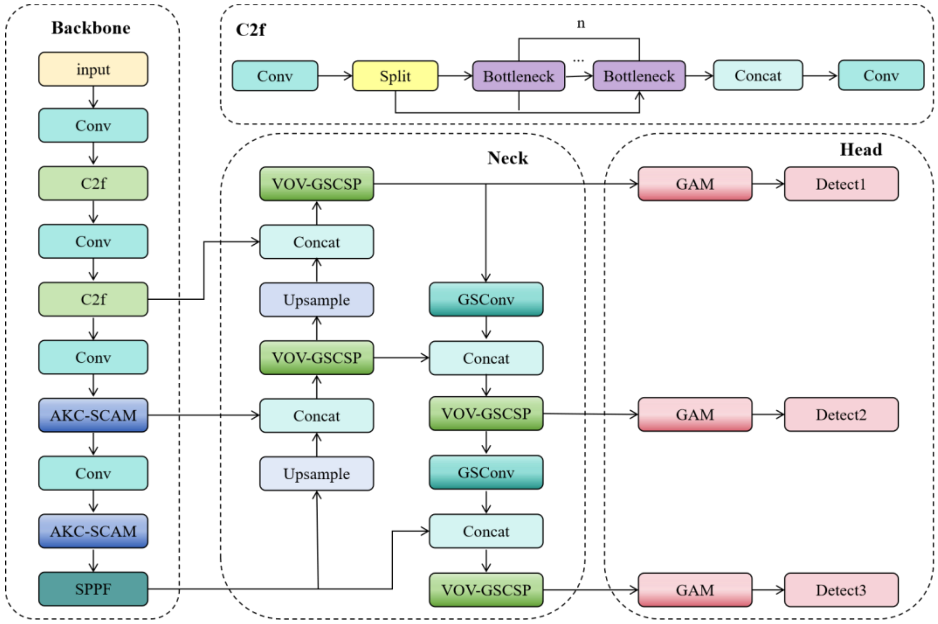

Fig. 1. The architecture of the ASG-YOLOv8s.

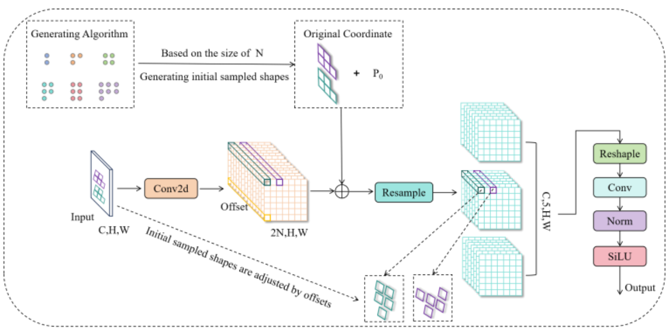

Fig. 2. Structure of the AKConv module.

In the diagram, the input to the AKConv module is a feature map with dimensions (C, H, W), where C denotes the number of channels, and H and W represent the height and width, respectively. The module first generates an initial sampling structure for the convolutional kernel using an initial sampling shape. It then performs a standard convolution operation on the feature map via Conv2d (two-dimensional convolution). Learned offsets dynamically adjust the initial sampling shape; this adaptive property enables the convolutional kernel to flexibly accommodate the spatial distribution of the feature map. The adjusted sampling shape further drives a resampling operation on the feature map to obtain a more optimal local feature representation.

SCAM [17] comprises a channel attention mechanism and spatial-channel reconstruction units. It first decomposes global pooling into horizontal pooling (X Avg Pool) and vertical pooling (Y Avg Pool), thereby effectively capturing long-range spatial interactions with precise positional information. The pooled feature maps undergo concatenation (Concat), convolution (Conv), and splitting (Split) operations to enhance the accuracy of feature representation. Furthermore, the Spatial Reconstruction Unit (SRU) and Channel Reconstruction Unit (CRU) introduced in SCConv perform sequential spatial and channel-wise restructuring on input feature maps. The SRU converts spatial features into spatial weight information, applying thresholds to differentiate weight values. Subsequently, two sets of weighted information are applied to the original feature map to obtain non-redundant and redundant feature maps. These are split and interleaved, producing a spatially reorganised feature map of identical scale to the original. This process refines the spatial information.

Following spatial reorganisation, the feature map employs CRU to restructure channel information, partitioning the input feature map channels into two segments. One portion of the channel features serves as a carrier for rich information, while the other functions as a carrier for secondary information. Distinct convolutional operations are applied to each segment, yielding a feature map refined from the channel information. At this stage, the original feature map undergoes spatial and channel information reorganisation through the SRU and CRU, reducing redundancy within the features. Subsequently, the feature maps restructured by the SRU and CRU are concatenated with the reweighted feature map. This approach enhances the accuracy of feature representation while maintaining lightweight processing, minimises redundant information within the features, and effectively improves model performance. The SCAM architecture is illustrated in Fig. 3.

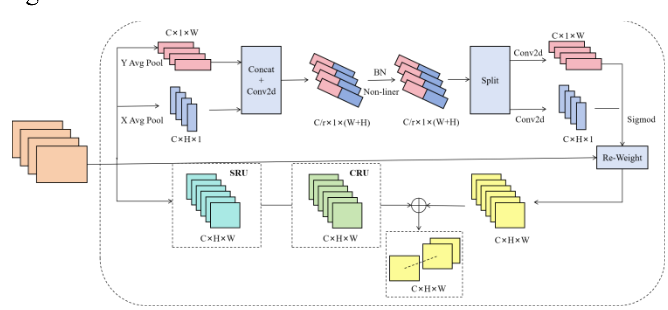

Fig. 3. Structure of the SCAM module.

C. Slim-neck Module

The neck network of YOLOv8 relies on dense convolutional operations during feature fusion, resulting in high parameter counts and computational complexity. To address this issue, this study proposes introducing the Slim-neck [18] module into the neck section. The Slim-neck module incorporates lightweight structures derived from GS bottleneck and VOV-GSCSP upon the foundation of GSConv [19] (Global Sparse Convolution), reducing model complexity while maintaining detection accuracy. GSConv represents a novel, lightweight convolution technique that enhances detection accuracy while reducing computational burden, differing from traditional approaches that increase accuracy by deepening network architecture and amplifying computational load. The GSConv architecture, illustrated in Fig. 4, processes input features through standard convolutional layers followed by depth-wise separable convolutions applied independently to each channel. Output features are then concatenated and rearranged via a shuffle operation to enhance inter-feature information flow. The GS bottleneck architecture further optimises GSConv, enhancing the network's feature processing capabilities. VOV-GSCSP [20] employs a one-time aggregation strategy, optimising certain module designs to improve feature utilisation efficiency and overall coherence.

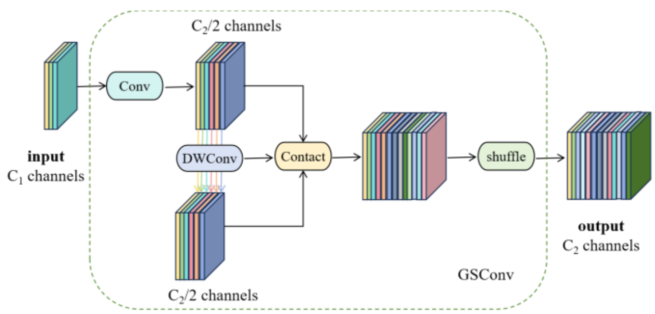

Fig. 4. Structure of the GSConv module.

D. GAM Attention Mechanism

To enable the model to adaptively focus on key regions within the input data while enhancing robustness and reducing computational complexity, this study incorporates a Global Attention Mechanism (GAM) module [21] before each detection head in YOLOv8. By integrating dual mechanisms of channel attention and spatial attention, GAM collaboratively captures key features across channel, spatial, and feature dimensions, achieving global context modelling of input data. Its primary structure is illustrated in Fig. 5. Specifically, the channel attention module extracts channel importance weights through Global Average Pooling (GAP) and Global Max Pooling (GMP), then employs a shared Multi-Layer Perceptron (MLP) to generate channel weight vectors, thereby reinforcing key semantic features, as depicted in Fig. 5(a). The spatial attention module generates spatial saliency maps through parallel global pooling and 7×7 convolutions, focusing on the spatial distribution information of target regions, as depicted in Fig. 5(b). Subsequently, the two weight types interact through element-wise multiplication to achieve multidimensional feature interaction, ultimately outputting an optimised feature representation. Compared to the conventional Convolutional Attention Module (CBAM) [22], GAM further incorporates a dynamic weight allocation strategy. This enhances the perception of fine-grained features in complex scenes while reducing redundant computational overhead through lightweight design, thereby improving model performance.

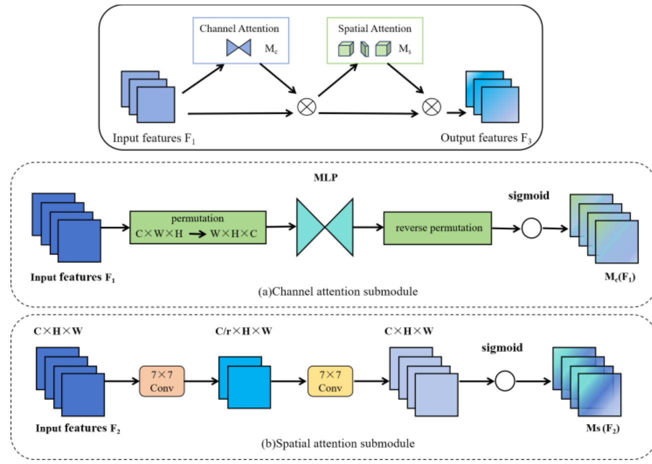

Fig. 5. Structure of the GAM module.

III. Experimental Data and Analysis

A. Datasets and Evaluation Metrics

The dataset employed herein should encompass construction sites across different seasons and complex weather conditions to fulfil the experimental requirements of this study. The dataset utilised in this experiment was created by refining the publicly available Safety-Helmet-Wearing-Dataset (SHWD). This involved removing unsuitable images and substantially augmenting the dataset with additional construction site scenarios, resulting in the SHWD-T construction site helmet dataset comprising 21,480 images. The dataset was randomly partitioned into training, validation, and test sets in an 8:1:1 ratio to facilitate subsequent data processing and analysis. All experiments were conducted without loading pre-trained weights.

To evaluate the model's performance and the accuracy of helmet detection, this study employs three metrics: Precision (P), Recall (R), and Mean Average Precision (mAP). The corresponding equations are presented in Eq. (1) to Eq. (3):

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

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

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{3}$$

In the equation: TP (True Positive) denotes the number of predicted boxes matching actual boxes; FP (False Positive) denotes the number of predicted boxes differing from actual boxes; FN (False Negative) denotes the number of objects contained within images that were not correctly detected during prediction; N represents the number of categories delineated in this study.

B. Ablation Experiment

To validate the effectiveness of the algorithmic improvement strategy proposed herein, eight sets of ablation experiments were conducted on the SHWD-T dataset based on the YOLOv8 baseline. The experimental results obtained from various experimental models are presented in Table I.

TABLE I. ABLATION EXPERIMENT RESULTS

AKC-SCAM	*Slim-neck*	*GAM*	*P/%*	*R/%*	*mAP/%*
-	-	-	87.26	80.52	83.62
√	-	-	88.07	81.22	84.30
-	√	-	88.92	81.88	85.16
-	-	√	88.44	82.73	84.41
√	√	-	89.07	81.65	84.83
√	-	√	88.49	82.07	85.29
-	√	√	89.09	82.50	85.63
√	√	√	89.80	83.46	86.78

Experiments reveal that incorporating the AKC-SCAM backbone network yields an 0.86% improvement in mAP compared to YOLOv8, demonstrating that the AKC-SCAM module effectively enhances the model's object recognition accuracy and strengthens its feature representation capabilities. Following the integration of the Slim-neck module, the model's mAP value increased by 1.84%, demonstrating that this lightweight neck design optimises feature propagation and enhances detection performance without significantly increasing computational overhead. The combined approach of YOLOv8 + AKC-SCAM + Slim-neck + GAM elevated the P value to 89.80%, the R value to 83.46%, and the mAP to 86.78%. This demonstrates that the synergistic interaction among these three components can substantially enhance the model's comprehensive detection capabilities.

Overall, the refined algorithm demonstrably enhances the model's comprehensive detection capabilities compared to YOLOv8, particularly by achieving a comprehensive improvement in detection accuracy while maintaining model lightness. This validates the effectiveness and superiority of the proposed methodology.

C. Contrast Experiment

To comprehensively evaluate the performance of the improved YOLOv8 network in helmet detection tasks, comparative experiments were conducted on the SHWD-T dataset against several classical object detection networks and state-of-the-art algorithms. As shown in Table II, the proposed ASG-YOLOv8s model demonstrates significantly superior detection accuracy compared to other methods.

TABLE II. CONTRAST EXPERIMENT RESULTS

Model	**Experimental parameters**		
	P/%	*R/%*	*mAP/%*
ASG-YOLOv8s	89.80	83.46	86.78
RFBNet	82.6	73.4	75.60
YOLOv5	86.40	78.60	82.70
YOLOv8	87.26	80.52	83.62
SSD	86.30	76.70	82.20
DAAM-YOLOv5	86.70	78.60	83.30
YOLOv7	87.98	84.21	85.76
YOLOv3-tiny	79.87	75.89	78.65
SSD	79.91	70.34	77.65
Faster RCNN	80.35	71.26	78.14
MEM-YOLOv8	87.50	76.60	83.20

D. Experimental Results Analysis

To validate the performance of the ASG-YOLOv8s model in practical detection scenarios, this study tests images captured across various complex construction site environments. Fig. 6 presents detection results from both the ASG-YOLOv8s and YOLOv8s models across distinct scenarios: Fig. 6(a) depicts a scene with background interference, Fig. 6(b) shows a densely populated target scenario, and Fig. 6(c) illustrates an occlusion scenario. The left column presents detection results from the YOLOv8s model, while the right column shows results from the ASG-YOLOv8s model. Comparing these outcomes enables an assessment of performance differences between the two models in diverse environments. It is evident that the YOLOv8s model is prone to both missed detections and false positives in complex environments, exhibiting lower detection accuracy. Conversely, the ASG-YOLOv8s model demonstrates superior performance, accurately detecting small objects within intricate settings.

(a) Scenes with background interference.

(b) Dense Target Scenario.

(c) Occlusion of Target Scenes.

Fig. 6. Comparison of model detection results in different job scenarios.

IV. Conclusion

To address the issues of delayed response times in manual industrial safety monitoring and high false detection rates for small targets such as safety helmets due to dynamic changes in monitoring distance, this study proposes the ASG-YOLOv8s enhanced algorithm. Firstly, the AKC-SCAM module is introduced into the backbone network to replace some standard convolutional layers in the conventional backbone. Secondly, a Slim-neck module is constructed in the Neck section, employing GSConv to substitute traditional convolutional kernels. Finally, a GAM module is embedded prior to the multi-scale outputs of the detection head. Experimental results demonstrate that the proposed ASG-YOLOv8s model achieves improvements of 2.54%, 2.94%, and 3.16% in Precision, Recall, and mAP, respectively, over the original model on the SHWD-T dataset. This indicates that the ASG-YOLOv8s algorithm exhibits superior detection accuracy, with significant enhancements in detecting occluded and small targets, rendering it well-suited for safety helmet detection applications.

References

[1] TANG K, CHEN L, ZHANG Z J, et al.Statistical analysis and countermeasures of production safety accidents i n my country's construction industry[J]. Construction Saf ety, 2020, 35(09): 40-43.

[2] Kelm A ,Laußat L ,Meins-Becker A , et al.Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment (PPE) on construction sites[J].Automation in Construction,2013,36:38-52.

[3] DONG S , HE Q , LI H ,et al. Automated PPE Misuse Identification and Assessment for Safety Performance Enhancement[M]//ICCREM 2015. 2015: 204-214.

[4] ZHU F Y, DAI X L, FU J, et al. Research on safety helmet detection algorithm in substation based on video surveillance[J]. Power System and Clean Energy, 2018, 34(12): 71-76.

[5] JIA J S, BAO Q J, TANG H M. Safety helmet wearing detection based on deformable component model[J]. Journal of Computer Application Research, 2016, 33(03): 953-956.

[6] SHAOQING R ,KAIMING H ,ROSS G , et al.Faster R- CNN: Towards Real-Time Object Detection with Region Proposal Networks.[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137-1149.

[7] LIU W ,ANGUELOV D ,ERHAN D, et al. Ssd: Singles hot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference,

Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[8] REDMON J ,DIVVALA K S ,GIRSHICK B R , et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vi sion and pattern recognition. 2016: 779-788.

[9] ZHANG J, QU P Q, SUN C, et al. Helmet wearing detection algorithm based on improved YOLOv5[J]. Journal of Computer Applications, 2022, 42(04): 1292-1300.

[10] SU P, LIU M, MA S Q. Safety helmet detection based on improved YOLOX-s[J]. Computer Systems and Applications, 2023, 32(07): 145-154.

[11] HAN F T,LIU Y Q,FANG Y D ,et al.Detection model for wearing status of safety helmet based on attention mechanism[J].Journal of Safety Science and Technology, 2024,20(08):196-202.

[12] YANG G L, HONG X F, SHENG Y Y ,et al.A helmet detection algorithm based on appearance feature enhance ment and fusion[J/OL].Computer Engineering and Applications,1-12[2024-09-29].

[13] Ding L, Miao X R, Hu J F, et al.Improved Miner Chin Strap Detection and PersonnelTracking with YOLOv8s and DeepSORT[J].Computer Engineering and Applications,2024,60(5):328-335.

[14] Wang Y ,Yang H,Gao L, et al.Research on Target Detection Algorithm for Sea Surface Oil Spill Recognition in SAR Image on YOLOS-Taking Improved YOLOv8 as an Example[J].IAENG International Journal of Computer Science,2025,52(12):

[15] Ai S ,Li J ,Li Y .MDC-YOLO: an enhanced steel plate surface defect detection algorithm based on YOLOv8[J].Engineering Research Express,2025,7(4):0452a5-0452a5.DOI:10.1088/2631-8695/AE2237.

[16] Feng S ,Niu X .Vehicle Detection in Aerial Drone Overhead Images based on YOLOv8[J].International Core Journal of Engineering,2025,11(5):147-156.

[17] Jenipher N V ,Radhika S .Lung tumor cell classification with lightweight mobileNetV2 and attention-based SCAM enhanced faster R-CNN[J].Evolving Systems,2024,15(4):1381-1398.DOI:10.1007/S12530-023-09564-3.

[18] Chen Z ,Fang Y ,Yin J , et al.A novel lightweight YOLOv8-PSS model for obstacle detection on the path of unmanned agricultural vehicles[J].Frontiers in Plant Science,2024,151509746-1509746.

[19] Ancha K V ,Gonuguntla V ,Vaddi R .GSS-YOLO: an improved YOLOV5 prediction head with slim-neck for defect detection in printed circuit board assembly[J].Signal, Image and Video Processing,2025,19(11):915-915.DOI:10.1007/S11760-025-04511-Y.

[20] Song L ,Lu S ,Tong Y , et al.YOLOv8s-GSW: a real-time detection model for hexagonal barbed wire breakpoints[J].The Journal of Supercomputing,2024,81(1):222-222.DOI:10.1007/S11227-024-06738-X.

[21] Yu Q ,Ouyang X,Su B, et al.Vehicle Detection Algorithm in Complex Scenes Based on Improved YOLOv8[J].IAENG International Journal of Computer Science,2025,52(4).

[22] Sayem S ,Sanny I S S ,Nabil R H .Hybrid CNN Model with CBAM for Malaria Disease Classification: Enhanced Feature Extraction[J].Biomedical Materials & Devices,2025,(prepublish):1-15.DOI:10.1007/S44174-025-00607-8.