An LCDR-YOLOv8 Model for Hazardous Behavior Detection on Construction Sites

Xiaolin Gu, Zhuo Xu, Chendi Cao

Abstract—**In hazardous environments, such as those found in industrial and construction sites, it is crucial that personnel are equipped with the appropriate protective gear, including reflective vests, helmets, safety harnesses, and goggles. This paper addresses the challenge of detecting reflective vests and helmets. Construction sites are densely populated and frequently obstructed, leading to a high incidence of missed inspections. Additionally, reflective vests and helmets are small targets available in various styles and colors, which negatively impacts detection accuracy. To address these challenges, this paper introduces an enhanced YOLOv8 model designed to improve the accuracy of detecting reflective vests and helmets. First, we utilize lightweight convolution in place of partial convolution to decrease the number of model parameters and enhance detection performance. Second, we introduce the CPCA attention mechanism to develop spatial attention through a multi-scale depth-separable convolutional module, which dynamically allocates attention weights and further enhances the model's detection accuracy. To mitigate the issue of semantic information loss for small targets, we propose adding a small target detection layer to improve the fusion of deep and shallow semantic information. Finally, to further improve the model's capability to comprehend the input data, we introduce the novel RepNCSPELAN4 module. This model effectively integrates contextual information through a series of convolutional operations and feature reorganization mechanisms, thereby significantly augmenting the model's feature extraction and representation capabilities. Following experimental validation, the LCDR-YOLOv8 algorithm achieves a detection accuracy of 91.6%, representing a 4.2% improvement in the mAP@0.5 metric over the original YOLOv8 algorithm. These improvements allow our algorithm to achieve superior detection performance in recognizing reflective vests and helmets.**

Index Terms—**YOLOv8, Reflective vests, Helmets, Lightweight Convolution, Assigning Attention Weights, Detection layer, Small targets**

I. Introduction

Safety accidents remain a significant concern in China. In 2023, there were 1,239 safety accidents resulting in 1,358 fatalities, averaging four deaths per day. The primary cause of these accidents is non-compliance with safety protocols by personnel. The use of reflective vests and helmets on construction sites, including building projects, is crucial for ensuring worker safety in complex environments. Helmets effectively reduce or prevent head injuries and enhance overall worker safety, while reflective vests mitigate accident risks by increasing visibility. Consequently, detecting the wearing of reflective vests and helmets in high-risk workplaces—such as coal mines, substations, and construction sites—holds significant practical importance and value as a critical technology for enhancing video surveillance systems for safety management.

Images of reflective vests and helmets typically feature small, densely clustered targets with significant size variations and diverse styles, often set against complex backgrounds, which increase the difficulty of the detection task. Effective detection systems must manage closely spaced targets, adapt to a wide range of sizes and styles of reflective vests and helmets, and accurately identify and localize them against complex backgrounds. These requirements place heightened demands on the robustness and accuracy of the algorithms.

Initially, some researchers concentrated on the interaction between the human body and helmets. Liu et al. [2] employed skin color detection to locate the facial region, extracted Hu moment feature vectors from the face, and utilized a Support Vector Machine (SVM) to recognize helmets. However, this method is significantly affected by the viewing angle and is applicable only to limited scenarios, primarily at the entrance and exit points of construction sites. Park et al. [3] employed the background subtraction method to extract foreground objects and combined Histogram of Oriented Gradients (HOG) features with SVM to detect both the human body and helmet. Following detection, these were matched based on spatial and geometrical relationships to verify helmet usage by workers. However, this approach has limitations in scenarios where workers are not standing, are obscured, or are motionless. Zhou et al. [4] employed statistical analysis to characterize the texture of the head from video footage recorded at a construction site. Subsequently, they applied a classifier and a backpropagation (BP) artificial neural network to perform classification tasks. However, the method's recognition rate needs improvement when addressing complex backgrounds.

With the rapid advancement of deep learning, researchers have increasingly applied these techniques to the detection of reflective vests and helmets. Sun et al. [5] employed the SwinTransformer as the backbone network to extract deeper semantic information and capture specific helmet features. They incorporated a self-attention mechanism into Faster R-CNN to extract multi-level global information, achieving better performance compared to smaller networks. However, false detections can occur when objects of the same color are present. Zhao et al. [6] developed a compact BiFPN structure with reduced parameters based on YOLOv7-tiny, which functions as a feature pyramid module for the original model's feature fusion. Incorporating this structure enhances the model's performance in multi-scale feature fusion, along with its efficiency and accuracy. However, it results in

Manuscript received April 26, 2024; revised November 16, 2024.

Xiaolin Gu is an associate professor at School of Railway Intelligent Engineering, Dalian Jiaotong University, Dalian, China(e-mail:guxiaolin60@126.com).

Zhuo Xu is a postgraduate student at School of Railway Intelligent Engineering, Dalian Jiaotong University, Dalian, China(corresponding author to provide phone: +086-15524805825; e-mail:xuzhuo99@outlook.com).

Chendi Cao is a postgraduate student at School of Railway Intelligent Engineering, Dalian Jiaotong University, Dalian, China(e-mail: ccd0712@hotmail.com).

increased missed detections in reflective scenarios. Liu et al. [7] proposed a two-channel architecture based on YOLOv5 to enhance the model's ability to detect helmets in complex environments. Additionally, they incorporated a focus layer to improve the model's processing speed. Chen et al. [8] introduced a lightweight PP-LCNet to optimize the YOLOv4 network. The representation of feature information was enhanced by incorporating a coordinate attention mechanism into the three output feature layers of the backbone network. Furthermore, SIOU was employed as the loss function, which accelerates the model's convergence and improves regression accuracy. These enhancements significantly reduced the model size while improving the detection rate. Han et al. [9] developed a super-resolution reconstruction module to accelerate helmet detection. They employed a multi-channel attention mechanism to improve feature extraction and proposed a cross-gradient (CSP) module to mitigate information loss and gradient confusion. Cheng et al. [10] introduced a Generalized Intersection over Union (GIoU) loss function and a YOLOX-based bidirectional weighted feature pyramid network (BiFPN) module for detecting reflective vests and helmets. However, they did not benchmark their approach against other advanced object detection algorithms. Han et al. [11] optimized the Single Shot MultiBox Detector (SSD) algorithm by incorporating ResNet50 and a deformable convolution module. These modifications enabled the model to better adapt to targets of varying sizes and improved detection accuracy for reflective vests and helmets. However, this optimization may result in significant overlapping of bounding boxes. Xie et al. [12] introduced the CAM and TBCA modules based on YOLOX to broaden the model's receptive field. They also incorporated the VarifocalLoss regression function to improve the detection of positive samples and enhance focus on foreground objects. However, detecting reflective vests and helmets in more complex scenarios necessitates a larger and more diverse dataset for training.

Although current algorithms for detecting reflective vests and helmets have achieved notable advancements, several limitations persist. In complex industrial environments, factors such as intricate backgrounds, lighting variations, shadows, and various types of occlusions can adversely affect the algorithm's performance. Additionally, reflective vests and helmets often present as relatively small targets that may be distant from the camera. Detecting such small targets remains a significant challenge in computer vision, as their features can be indistinct and prone to resolution limitations and noise.

To address these issues, an improved YOLOv8 [13] approach is proposed, beginning with the replacement of traditional convolution with LightConv in the backbone network. LightConv significantly reduces the number of parameters and enhances detection performance by adapting the network architecture to more effectively process helmet and reflective vests information. Subsequently, the CPCA attention mechanism is introduced. CPCA incorporates a channel attention mechanism that dynamically adjusts the feature map weights of various channels, aiding the model in focusing on the most relevant features. Additionally, a spatial attention mechanism is incorporated, enabling the model to dynamically allocate attention weights across the spatial dimension to more effectively capture spatial information within the feature map. Furthermore, due to the small size of reflective vests and helmets—whose features may be blurred or challenging to distinguish—a specially designed small target detection layer enhances the ability to capture and identify these targets, thereby improving detection accuracy and reducing missed detections. Finally, all instances of C2f are replaced with RepNCSPELAN4, a module that effectively integrates contextual information and enhances the network's ability to extract features from reflective vests and helmet images.

II. Related algorithms

Deep learning-based object detection techniques can be categorized into two primary types: regression-based and classification-based methods. The former are referred to as one-stage detection methods, whereas the latter are known as two-stage detection methods. Two-stage detection techniques, such as R-CNN [14], Fast R-CNN [15], and Faster R-CNN [16], typically involve extracting regions of interest prior to classification and regression. One-stage detection techniques, including SSD (Single Shot MultiBox Detector) [17], the YOLO (You Only Look Once) family of algorithms [18], and RetinaNet [19], utilize a single forward network to concurrently detect multiple targets. The YOLO series of algorithms has gained widespread popularity due to its superior performance and rapid detection speed. In 2023, Ultralytics released the latest YOLOv8 algorithm, which currently boasts the highest detection accuracy. YOLOv8 incorporates several enhancements over previous versions, including a redesigned backbone network, anchorless detection heads, and an optimized loss function. This algorithm family is widely utilized in practical applications such as pedestrian detection [20], traffic monitoring [21], and crop detection [22].

To address the requirements of diverse scenarios, the YOLOv8 series provides several versions, including YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. YOLOv8n is the fastest and smallest model in the series. In this study, we utilize the YOLOv8n architecture.

YOLOv8's network architecture comprises an input layer, a backbone network, a neck network, and a detection head network. The model's adaptability to diverse scenes is enhanced through techniques such as mosaic data augmentation, adaptive bounding box calculation, and adaptive grayscale filling applied at the input stage. The backbone and neck networks build upon the design principles of YOLOv7 ELAN, substituting YOLOv5's C3 structure with the C2f structure, which offers more robust gradient flow. Additionally, the number of channels is optimized across different scales to enhance model performance at each scale. The neck network aims to improve feature fusion and extraction from the backbone network's output, thereby enhancing overall network performance. YOLOv8 introduces two significant improvements in the detection head compared to YOLOv5: (1) a decoupled head structure that separates classification from detection tasks; (2) a shift from the Anchor-Based design to an Anchor-Free detection method, which directly predicts target centroids and width-to-height ratios. This change reduces the number of anchor frames and enhances both detection speed and accuracy.

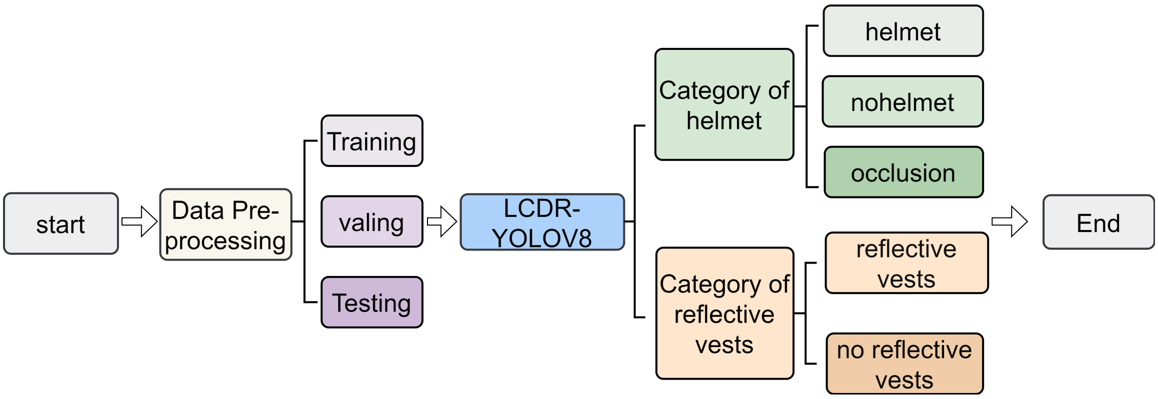


Fig. 1. LCDR-YOLOv8 detection process

III. IMPROVEMENTS

This research presents a deep learning system designed for recognizing reflective vests and helmets on construction sites. The overall process is illustrated in Fig. 1. Initially, the dataset is preprocessed and divided into three subsets: training, valing, and testing. Subsequently, the optimized YOLOv8 network is employed to detect reflective vests and helmets. In the detection process, helmets are classified into three categories: helmet, nohelmet, and occlusion. If the LCDR-YOLOv8 network identifies the helmet as fully exposed, it is classified as ”helmet”; if it is not worn, it is classified as ”nohelmet”; and if it is partially covered, it is classified as ”occlusion”. Reflective vests are detected based on two conditions: when worn and when not worn. If workers are detected wearing reflective vests, they are classified as ”reflective vests”; otherwise, they are classified as ”no reflective vests”. This concludes the testing process.

A. Lightweight Convolution

Lightweight convolution is typically designed to reduce a model’s computational complexity by decreasing both the computation load and the number of parameters. Models equipped with lightweight convolution generally demonstrate superior generalization capabilities compared to those using traditional convolutional methods. LightConv, as a lightweight convolutional method, enables the sharing of certain output channels and utilizes the softmax function to normalize weights across the time dimension. As illustrated in Fig. 2, unlike conventional convolution, LightConv employs a fixed context window, applying a set of weights with a constant time step to evaluate the significance of context elements.

LightConv performs the following computation for the ith element in the sequence and the output channel c:

$$\begin{aligned} LightConv(X, W_{\lceil \frac{cH}{d} \rceil}) =& DepthwiseConv(X, \\ & softmax(W_{\lceil \frac{cH}{d} \rceil}, :), i, c) \end{aligned} \quad (1)$$

LightConv utilizes weight sharing to connect parameters across H-bar channels, resulting in a significant reduction in the number of parameters. For example, a conventional convolution with d = 1024 and k = 7 has 7340032 weights ($d^2 \times k$), whereas a separable convolution has only 7,168 weights ($d \times k$), representing a relative reduction by a factor of d/H. This method of parameter sharing reduces both computational and memory overheads, making it ideal for resource-constrained applications. Additionally, we apply the softmax operation to normalize the weights $W \in R^{H \times k}$ along the temporal dimension k:

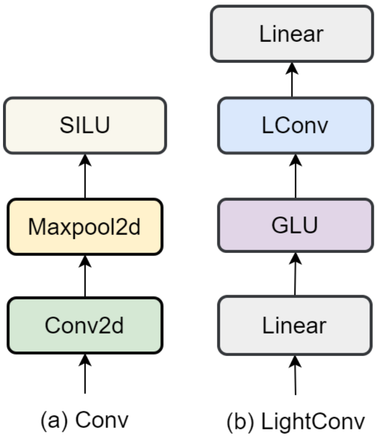


Fig. 2. LightConv structure

$$Softmax(W)_{h,j} = \frac{expW_{h,j}}{\sum_{j'\prime=1}^{k} expW_{h,j'}} \quad (2)$$

Fig. 2(b) illustrates the network architecture of the module incorporating LightConv. The design of the module involves the following steps:

1) Input Feature Map: Initially, the input is projected from dimension d to dimension 2d, effectively extending the input characteristics. This enhances the network’s representational capacity.
2) Gated Linear Unit (GLU): Subsequently, a Gated Linear Unit (GLU) is employed. The GLU processes the input by directing half of it to the sigmoid unit as a gate and performing an element-wise multiplication with the remaining half. This mechanism enables the network to selectively pass features, thereby enhancing the model’s expressive power.
3) Lightweight Convolution (LightConv): Following the GLU, LightConv is utilized. This lightweight convolution method performs convolution operations with fewer parameters, thereby maintaining performance while minimizing computational complexity.
4) Output Projection: Lastly, a projection of size $w \in R^{d*d}$ is applied to the output of LightConv to obtain

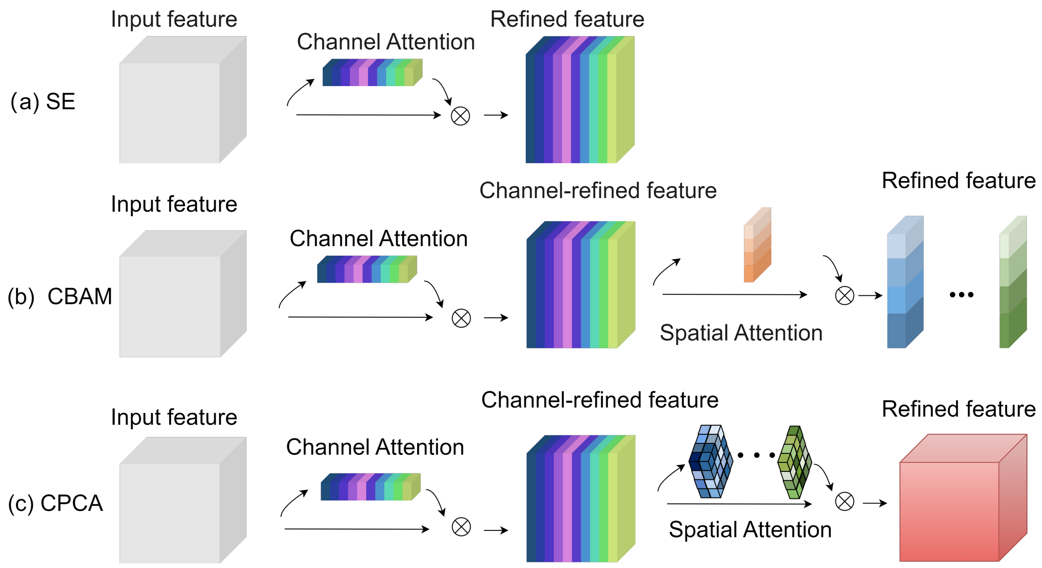


Fig. 3. Comparison of CPCA with SE, CBAM

the final result. This projection facilitates mapping the output to the required dimensions for the task.

B. Channel Prior Convolutional Attention

Both CBAM [23] and SE [24] are widely used attention mechanisms in deep learning. The SE mechanism (see Fig. 3(a)) integrates only channel attention, which limits its ability to select crucial regions. While CBAM integrates both channel and spatial attention, it applies spatial attention uniformly across all output channels. As illustrated in Fig. 3(c), we propose a novel Channel Prior Convolutional Attention (CPCA) mechanism [25] that supports dynamically distributed attention weights in both channel and spatial dimensions. By incorporating a multi-scale deep convolutional module, our method effectively extracts spatial relationships while preserving channel priors.

Fig. 4 illustrates the overall structure of CPCA, incorporating both channel and spatial attention mechanisms. Initially, channel attention aggregates spatial information from the feature map using methods such as average and maximum pooling. This aggregated spatial information is then processed by a MultiLayer Perceptron (MLP) and merged with the original feature map to generate the channel attention map. The channel prior is computed by element-wise multiplication of the input features with the channel attention map. Subsequently, the channel prior is input into a deep convolution module to produce the spatial attention map. The convolution module then integrates the spatial attention map with the channels. Finally, the result of channel integration is element-wise multiplied with the previous channel prior to produce the optimized features, which are then output.

where the calculation of channel attention is summarised as:

$$\begin{aligned}\mathrm{CA}(F) =&\sigma(\mathrm{MLP}(\mathrm{AvgPool}(F)) \\ &+\mathrm{MLP}(\mathrm{MaxPool}(F)))\end{aligned} \tag{3}$$

Where σ denotes the sigmoid function.

The calculation of spatial attention can be described as follows:

$$\mathrm{SA}(F) = \mathrm{Conv}_{1\times1}\left(\sum_{i=0}^{3}\mathrm{Branch}_i(\mathrm{DwConv}(F))\right) \tag{4}$$

Where DwConv represents deep convolution, and $Branch_i$, $i \in \{0,1,2,3\}$ denotes the i-th branch. $Branch_0$ is the identity connection.

C. Adding A Small Target Detection Layer

A significant challenge in YOLOv8 is detecting small targets, which often suffer from feature information loss in the original model. The original model processes input images of size 640×640, with a minimum detectable object size of 80×80 pixels. Consequently, if a target in the image has dimensions smaller than 8 pixels in height and width, the network may struggle to capture the essential feature information effectively.

To address this issue, this thesis proposes incorporating a small target detection layer into the network, with a size of $160{\times}160$, as illustrated in Fig. 5. The proposed layer includes a complementary fusion feature layer and an additional detection head, designed to enhance semantic information and feature representation for small targets. The implementation involves the following steps: First, the 80×80 feature layer from the fifth layer of the backbone network is combined with the up-sampled feature layer from the neck network. Following upsampling and the application of C2f (which refines features from coarse to fine-grained), a deep semantic feature layer is created, encapsulating critical information about the small target. This deep semantic feature layer is subsequently merged with the shallow positional feature layer from the third layer of the backbone network to produce a comprehensive 160×160 fusion feature layer, which more accurately represents the small target's semantic attributes and positional data. Finally, these features are directed to an additional detached head for further analysis.

The additional decoupled head is then processed through a convolutional layer and C2f to integrate the deep semantic feature layer with the positional feature layer from layer 15 in the neck network. This integration enables the crucial feature

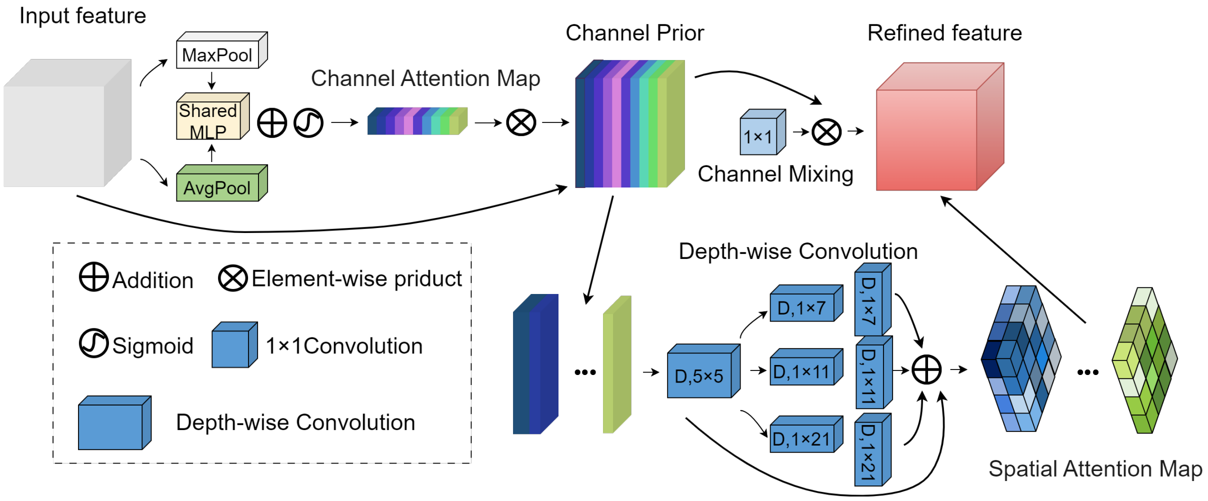


Fig. 4. CPCA Attention Mechanism

information of small targets to be communicated to the model's original three-scale feature layers, thereby enhancing the network's feature fusion capability and improving the accuracy of small target detection. The inclusion of this head extends the network's ability to detect reflective vests and helmets, thereby enabling more effective assessment of whether construction workers are wearing the necessary protective equipment.

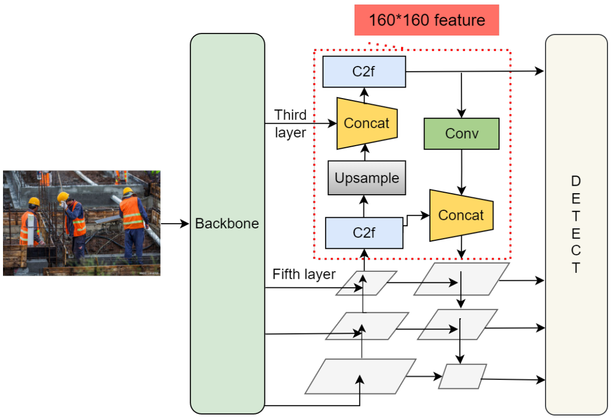


Fig. 5. Small target detection layer

D. Feature Fusion

In the previous subsection, a detection head for small targets was added to the original model, and the detection head serves as the network architecture responsible for processing the extracted features. Moreover, the capability for small target feature extraction is crucial for the model's detection performance. Although the C2f module in YOLOv8 enhances feature fusion, it may be hindered by feature map separation and re-fusion when detecting very small targets, resulting in the loss of detailed information and poor performance in detecting small objects. Therefore, this paper introduces the RepNCSPELAN4 module, which is capable of more effectively capturing detailed information and high-level semantic data through multi-scale feature extraction and fusion, thereby improving the accuracy of target detection, particularly for small targets and dense objects.

Fig. 6 illustrates the structure of RepNCSPELAN4.The RepNCSPELAN4 module primarily comprises Conv and RepNCSP components. RepNCSP is structurally analogous to the C3 and C2f modules but offers superior computational efficiency while maintaining similar performance to the C3 module. In comparison to the C2f module, RepNCSP enhances the feature hierarchy and improves the model's representation capability through a more sophisticated feature fusion strategy. The RepNCSP module consists of a Conv layer and a variable number of RepNBottleneck modules, with the number of RepNBottleneck modules determined by the model's width factor. RepNBottleneck is a foundational module with a residual structure.

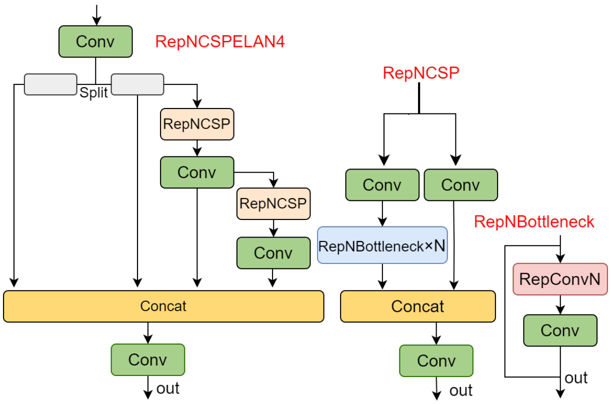


Fig. 6. RepNCSPELAN4 structure

The RepNCSPELAN4 module is initiated by an initial convolutional layer that extracts spatial features from the input data. Subsequently, the output is divided into multiple parallel branches, each independently processing distinct feature maps. In one branch, the input is processed through a RepNCSP block, which incorporates convolutional operations that enhance feature extraction. The resulting features are further processed by an additional convolutional layer. Simultaneously, a separate branch undergoes equivalent processing through its respective RepNCSP and convolutional operations. The outputs from both branches are concatenated, effectively merging the extracted features into a unified representation. Finally, the concatenated feature map is passed through a concluding convolutional layer, resulting in the module's output. This architecture, characterized by parallel

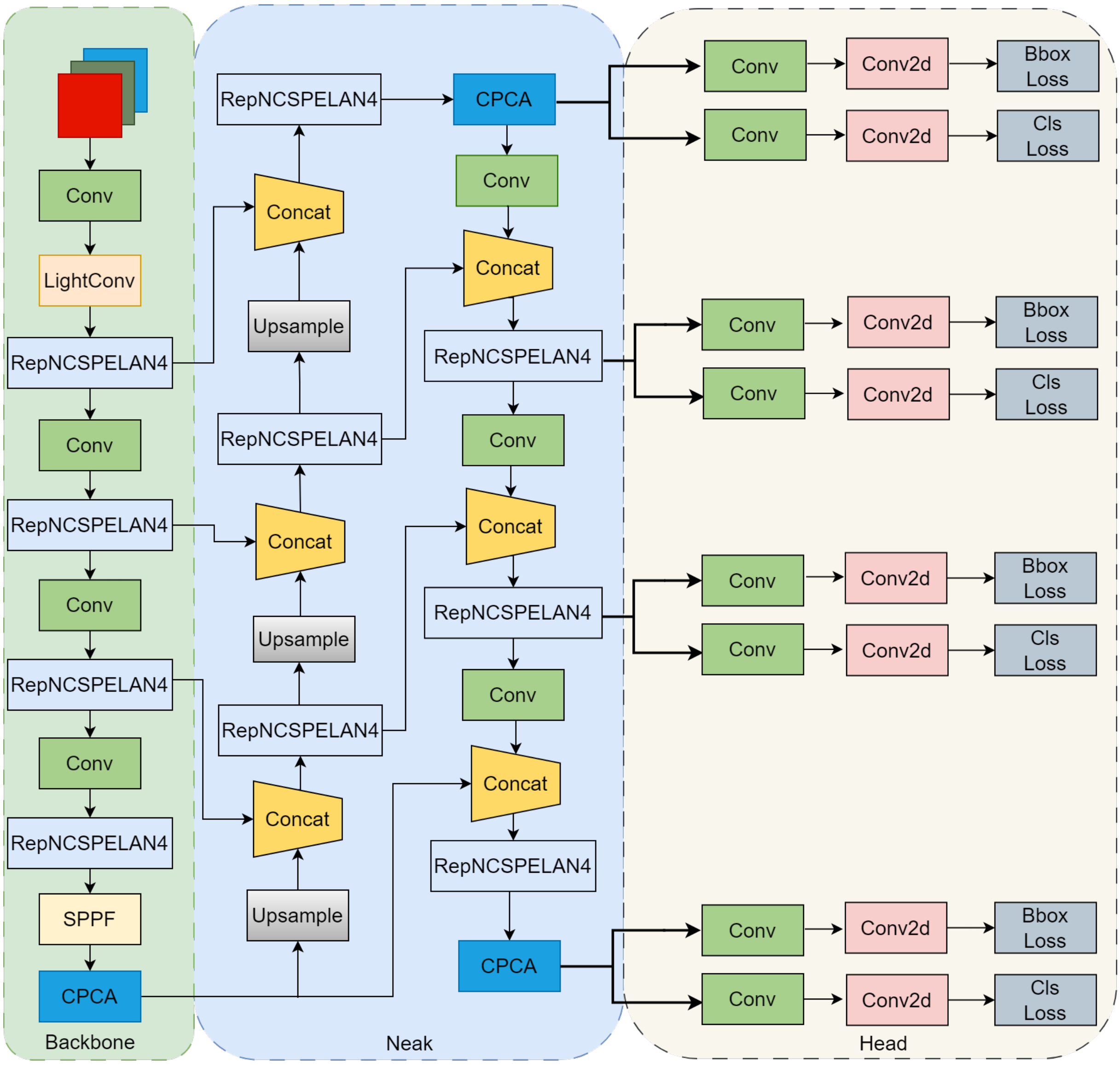


Fig. 7. LCDR-YOLOv8 model

feature extraction and subsequent feature fusion, enables the RepNCSPELAN4 module to efficiently capture complex spatial hierarchies, thereby improving model performance in downstream tasks.

To address the low detection accuracy of reflective vests and helmets in complex scenes, this paper employs YOLOv8 as the base model. Initially, the second Conv layer in the backbone is replaced with LightConv. LightConv utilizes a fixed context window and weight-based context determination, significantly reducing the number of parameters. Subsequently, three CPCA attention mechanism modules are introduced between the backbone and the neck network. These modules incorporate a multi-scale deep convolution component to enhance the localization and recognition of the feature map regions of interest, thereby mitigating the issue of missed detections due to target occlusion. Additionally, a 160×160 detection layer is integrated into the head section, utilizing four different scales of convolutional methods to improve the detection accuracy for small targets. Finally, all instances of C2f are replaced with RepNCSPELAN4, which enhances the network's feature extraction and understanding of the input data through multi-branch convolutional operations and feature reorganization mechanisms. The structure of the improved LCDR-YOLOv8 model is illustrated in Fig. 7.

IV. Experiment

A. Experimental Environment and Parameter Configuration

The experiments in this paper use the environment as in Table I, using the python framework, calling the GPU for the experiments, the model training weights are YOLOv8n.pt, the input image size is 640×640, the batch size is set to 8, the epochs are set to 200, the patency is set to 30, and the lr is set to 0.01.

B. Introduction to Database

Datasets are crucial for deep learning-based target detection. Currently, the Baidu Flying Paddle platform provides a VOC2021 dataset for reflective vests and helmets, comprising 1,083 images categorized into four classes: helmets, individuals wearing reflective vests, individuals without helmets, and individuals without reflective vests. However, this dataset's limited size poses challenges for achieving reliable detection in complex scenarios.

TABLE I
EXAMPLE CONFIGURATION

Name	Configuration
Operating System	Windows11
CPU	Intel(R) Core(TM) i5-13500HX 2.50 GHz
GPU	Nvidia Geforce RTX 4060
Memory	8GB
Cuda	11.8.0
Cudnn	8.5.0
Pytorch	2.1.1
Python	3.8.0

To overcome this limitation, we generated a new detection dataset for reflective vests and helmets to enhance both the size and diversity of the dataset. This effort involved collecting authentic building site scene data to ensure the dataset aligns more closely with practical application requirements. The dataset creation process was conducted in several stages: data collection, screening, and processing. We acquired photographs in JPG format through Internet searches, extranets, and web crawlers, ultimately gathering 6,800 images after rigorous screening and processing. Labeling was performed using the labelImg software, with annotations for five categories: helmets, nohelmets (including non-protective headgear), occlusion, reflective vests, and no reflective vests. The coordinate information for all annotated images was saved in a text file. Finally, the dataset was randomly divided into training, validation, and test sets in an 8:1:1 ratio, containing a total of 30,000 labeled objects.

C. Evaluation Indexs

The primary evaluation metrics for target identification algorithms encompass detection accuracy and model complexity. Detection accuracy is typically measured using Precision, Recall, and mean Average Precision (mAP).

Assuming that the number of true positive samples in the prediction result is TP, the number of false positive predicted as positive samples is FP, the number of positive samples predicted as negative samples is FN, the number of positive samples predicted as positive samples is TP+FP, and the number of true positive samples is TP+FN, the following formulas are used to calculate P (Precision), R (Recall), and mAP (mean Average Precision):

Precision(P):

$$P = \frac{TP}{TP+FP} \tag{5}$$

Recall(R):

$$R = \frac{TP}{TP+FN} \tag{6}$$

Average Precision(AP):

$$AP = \int_0^1 P(t)dt \tag{7}$$

where: t is the recall of the curve at different IOU, e.g., when $t = 0.7$, only $IOU \geq 0.7$ are considered positive samples.

mean Average Precision(mAP):

$$mAP = \equiv \frac{\Sigma_0^N AP_n}{N} \tag{8}$$

where:N denotes the number of categories in the dataset.AP stands for the average precision of a specific category in the dataset. This approach allows for a comprehensive evaluation of the algorithm's performance across all N categories, providing insights into its effectiveness in detecting and classifying objects within each specific category.

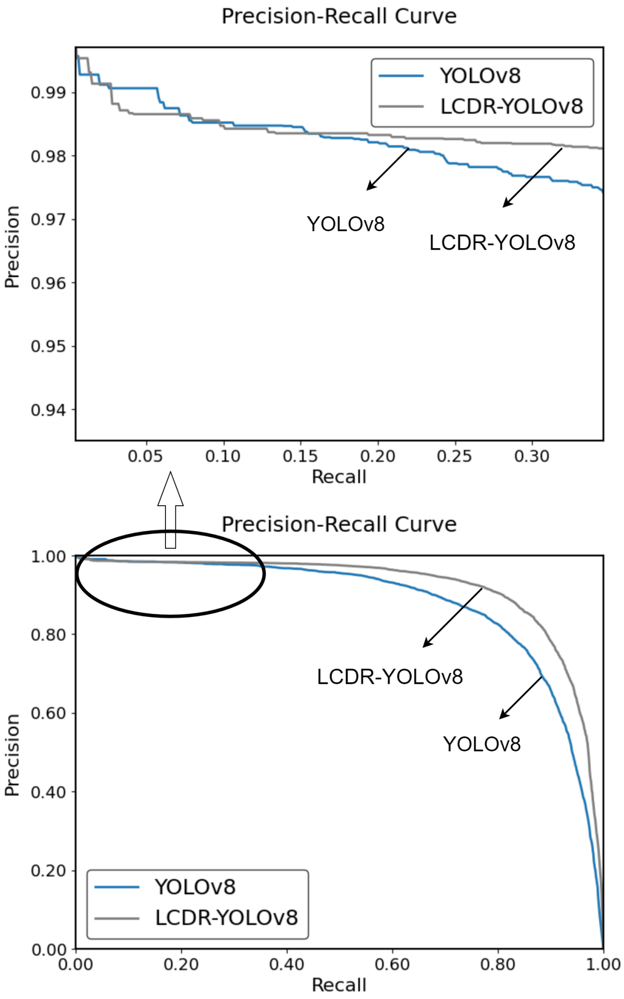


Fig. 8. P-R curves before and after improvement

Precision and recall significantly influence the model's target recognition capability. To comprehensively evaluate the network model's performance, the precision-recall (P-R) curves were compared before and after the improvements. The horizontal axis of the P-R curve represents "Recall," while the vertical axis represents "Precision." A larger area under the curve indicates better model performance. As shown in Fig. 8, the P-R curve of LCDR-YOLOv8 covers a larger area than that of YOLOv8, demonstrating the superior performance of LCDR-YOLOv8 in detecting reflective clothing and helmets.

D. Ablation Experiments

The algorithm presented in this paper improves YOLOv8 through the integration of LightConv , a 640×640 detection layer, the CPCA attention mechanism, and the RepNCSPELAN4 module. To evaluate the impact of each modification on model performance, ablation experiments were conducted under a consistent experimental environment and

TABLE II
RESULTS OF ABLATION COMPARISON EXPERIMENTS

Models	AP (%)					mAP@50/%
	Helmet	NoHelmet	Occlusion	Reflective	No Reflective	
YOLOv8	94.7	89.0	71.3	90.0	92.2	87.4
YOLOv8+LightConv	95.7	90.8	76.7	90.2	92.5	89.2
YOLOv8+detect	95.0	90.4	73.9	90.0	92.5	88.4
YOLOv8+CPCA	95.1	89.3	72.0	89.8	92.5 o	87.7
YOLOv8+RepNCSPELAN4	95.2	91.5	75.6	90.9	93.0	89.2
YOLOv8+LightConv+detect	95.9	92.0	78.9	90.6	92.8	90.0
YOLOv8+LightConv+CPCA+detect	95.9	92.9	78.6	91.7	93.5	90.5
All	95.8	93.7	81.4	92.9	94.3	91.6

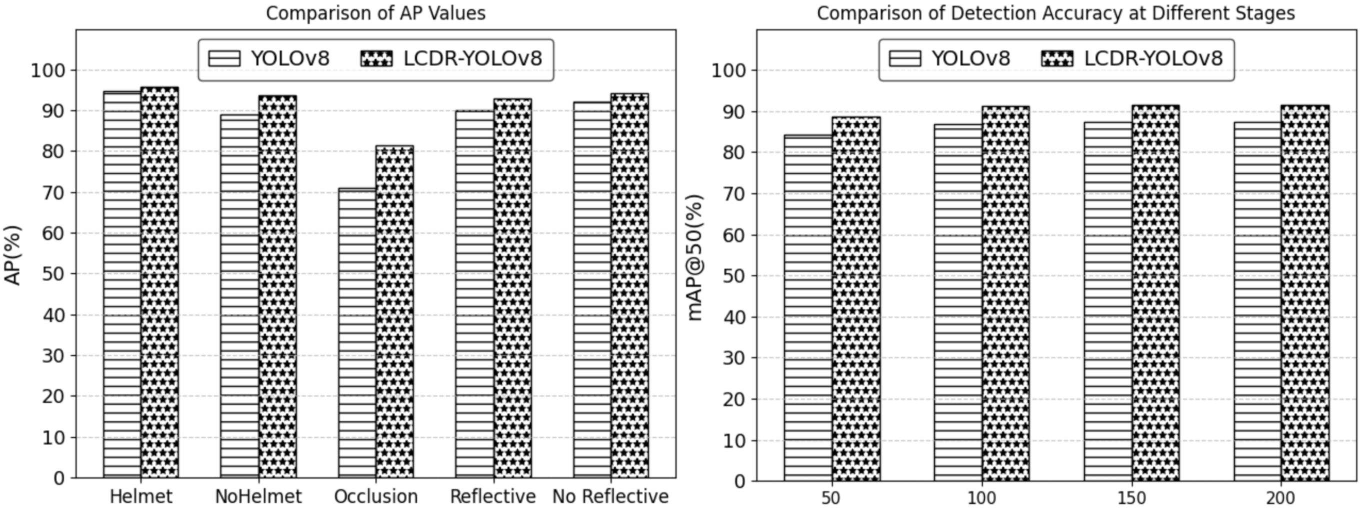


Fig. 9. Results of ablation experiment data

dataset. The results of these experiments are detailed in Table II.

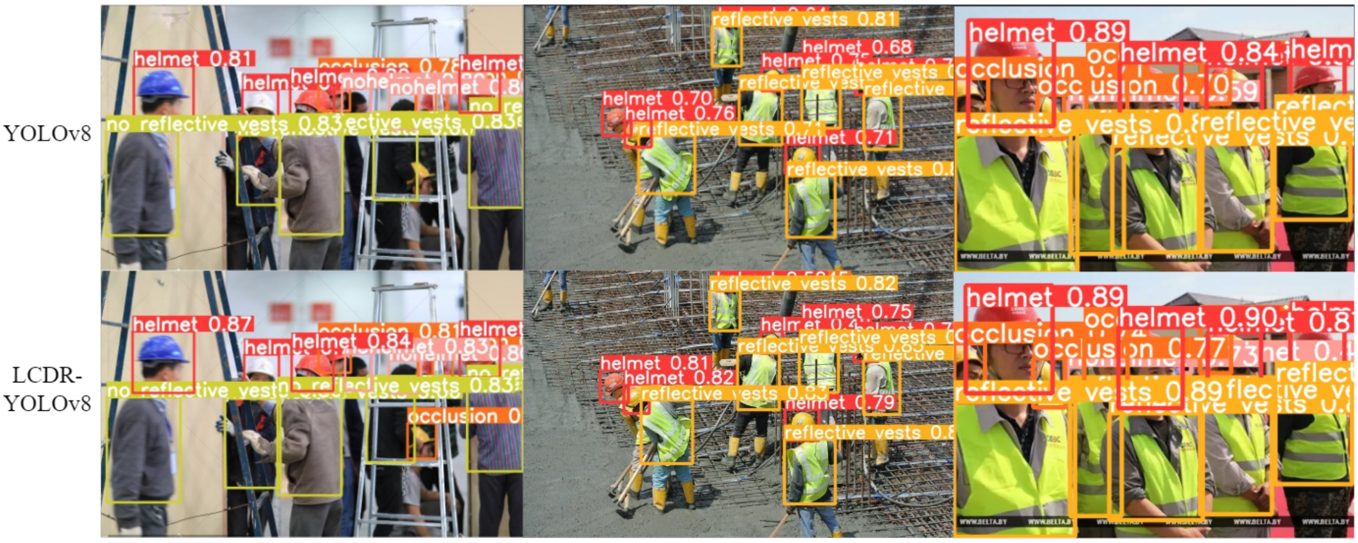
Table II demonstrates that replacing Conv with LightConv in the YOLOv8 network resulted in increased accuracy across all five categories, with the most notable improvement observed in the occlusion category, which saw a 5.4% increase. This substitution also led to an overall gain of 1.8% in the mean Average Precision (mAP). The integration of the 160x160 detection layer further enhanced the network's performance, elevating the accuracy of the occlusion category from 71.3% to 78.9%, representing a 7.6% improvement over the original model, along with a 2.6% increase in the mAP. These two modifications significantly improved the detection of occlusion categories. The introduction of the CPCA attention mechanism led to notable improvements in the no-helmet, reflective vests, and no-reflective vests categories, with an overall mAP increase of 3.1%. Replacing all instances of C2f with the RepNCSPELAN4 module caused a minor 0.1% decrease in accuracy for the helmet category but resulted in substantial increases in accuracy across the remaining four categories and a 4.2% rise in overall mAP.

To visualize the impact of these improvements on YOLOv8 algorithm performance, bar charts were created to display the experimental data, as shown in Fig. 9. The left graph displays the model's impact on category accuracy before and after improvements, while the right graph illustrates detection accuracy at various stages before and after the improvements. In conclusion, the proposed improvements have proven highly effective, significantly enhancing the recognition rate for reflective clothing and helmets in complex scenes.

E. Experimental Results and Analysis

Using the homemade dataset, we compared target recognition performance between YOLOv8 and LCDR-YOLOv8 detection algorithms across five categories: helmet, nohelmet, occlusion, reflective vests, and no reflective vests.

Fig. 10 illustrates the detection performance of YOLOv8 (up) and LCDR-YOLOv8 (down). Fig. 10(a) depict scenarios involving target occlusion. The original YOLOv8 model exhibits two omissions in detecting the occluded targets, whereas the LCDR-YOLOv8 algorithm successfully identifies all instances. This indicates a significant improvement in detection accuracy for the occlusion category with LCDR-YOLOv8. Fig. 10(b) focus on small targets, particularly those situated further from the camera. The revised method not only enhances the Average Precision (AP) values for these categories but also identifies targets missed by the previous model. Fig. 10(c) present images with dense targets, where the system demonstrates its capability to detect targets even when partially obscured by other objects.This validation underscores that LCDR-YOLOv8 significantly improves detection accuracy for small and medium-sized targets on construction sites and effectively mitigates the omission problem caused by occlusion, thereby enabling reliable automated detection of reflective vests and helmet usage.



(a) Targets occlusion (b) Small targets (c) Intensive targets

Fig. 10. Comparison of YOLOv8 and LCDR-YOLOv8 detection results

F. Comparison between Different Improvement Modules

In this section, the effectiveness of the improved convolutional layers, feature fusion, and the added attention mechanism will be validated through the presentation and analysis of experimental data.

The initial improvement in the convolutional layers is highlighted in Table III, demonstrating that LightConv exhibits superior performance in recall rate and mAP@0.5, signifying a clear advantage in detecting more positive samples and improving overall accuracy. Although Conv achieves marginally higher precision, its overall performance does not exceed that of LightConv. GhostConv shows poorer performance in both recall rate and mAP, indicating lower overall detection effectiveness compared to other convolutional layers. As a result, LightConv emerges as the optimal choice for this model based on comprehensive performance metrics. To assess the impact of the CPCA attention mechanism on the accuracy of the LCDR-YOLOv8 algorithm, comparative tests were conducted using various attention mechanisms within the same experimental environment and dataset. Table IV shows that incorporating NAM and Biformer caused a slight reduction in accuracy. The inclusion of EMA and Focused Linear did not change the accuracy compared to the original model. In contrast, incorporating SE, SK, MSCA, and CPCA resulted in a slight improvement in accuracy, with CPCA providing the most significant enhancement. As a result, the CPCA attention mechanism was chosen as the final optimization strategy for this experiment.

TABLE III
COMPARISON USING DIFFERENT CONVOLUTIONS

Model	Precision/%	Recall/%	mAP@0.5%
Conv	86.3	80.6	87.4
ODConv	86.0	81.0	87.7
DWConv	85.4	80.9	87.3
GhostConv	86.0	79.7	86.7
RepConv	85.8	80.2	87.3
GSConv	84.9	79.9	86.5
ConConv	84.0	80.0	87.5
LightConv	86.0	82.3	89.2

TABLE IV
COMPARISON USING DIFFERENT ATTENTION MECHANISMS

Models	Precision/%	Recall/%	mAP@0.5%
YOLOv8	86.3	80.6	87.4
YOLOv8+SE	86.8	80.4	87.5
YOLOv8+NAM	86.2	80.2	87.3
YOLOv8+EMA	85.6	81.8	87.4
YOLOv8+SK	86.1	81.0	87.6
YOLOv8+FL	86.6	80.8	87.4
YOLOv8+MSCA	85.4	81.2	87.5
YOLOv8+Biformer	84.0	74.2	81.8
YOLOv8+CPCA	86.2	80.0	87.7

Finally, the improvements in the feature fusion section are discussed. As indicated by the experimental data in Table V, RepNCSPELAN4 significantly outperforms other feature fusion models in terms of recall rate and mAP@0.5, demonstrating its superior accuracy and comprehensive detection capabilities. Although C2f also demonstrates commendable performance in precision, its overall metrics fall short compared to those of RepNCSPELAN4. The MobileOneBlock model exhibits relatively weaker performance across all metrics, indicating that its detection capabilities require further enhancement. Therefore, RepNCSPELAN4 emerges as the optimal choice when considering overall performance.

TABLE V
COMPARISON USING DIFFERENT FEATURE FUSION MODULES

Model	Precision/%	Recall/%	mAP@0.5%
C2f	86.3	80.6	87.4
C2f-uib	86.0	80.3	87.6
C2f-repghost	85.7	80.0	87.1
MobileOneBlock	84.8	79.2	86.3
C2f-repvit	85.0	80.2	86.6
C2f-DWR	85.8	80.7	87.1
RepNCSPELAN4	86.3	83.8	89.2

G. Compared with the Performance of Advanced Object Detection Algorithms

To further demonstrate the superiority of the proposed algorithm, we compared the improved YOLOv8 detection algorithm against several other algorithms, including SSD, Faster R-CNN, EfficientDet, and various classical YOLO series models such as YOLOv5, YOLOv7-tiny, YOLOX, and YOLOv10. The comparison results are presented in Table VI. The average detection accuracy of SSD, Faster R-CNN, YOLOv5, YOLOX, and EfficientDet is notably lower than that of the improved YOLOv8. The YOLOv7-tiny and YOLOv10 models demonstrated good performance, achieving mAP scores of 89.1% and 89.5%, respectively; however, these results were surpassed by the improved YOLOv8, which achieved a mAP of 91.6%. Thus, the improved YOLOv8 demonstrates superior performance in terms of detection accuracy.

TABLE VI
COMPARISON OF RESULTS WITH OTHER EXPERIMENTS

Model	Precision/%	Recall/%	mAP@0.5/%
SSD	72.2	64.9	78.1
Faster-Rcnn	64.3	60.6	67.4
EfficientDet	89.5	69.2	79.2
YOLOv5	85.6	80.4	86.8
YOLOv7-tiny	86.7	82.1	89.1
YOLOX	73.5	65.2	79.8
YOLOv8	86.3	80.6	87.4
YOLOv10	87.0	82.9	89.5
Ours	88.2	86.1	91.6

H. Detection on Random Images

To better illustrate these improvements, a visualization file containing 1,385 images was generated to demonstrate the detection results of true positives, missed detections, and false detections. In the visualization file, True Positives (TP) are highlighted in green, False Negatives (FN) in red, and False Positives (FP) in the regions where green and red overlap. Fig.11 presents part of the visualized detection results, comparing the original model's detection on the left with the improved LCDR-YOLOV8 model on the right. The figure clearly shows that the improved model has fewer missed and false detections compared to the original model, indicating superior performance of the LCDR-YOLOV8 model in complex backgrounds with severe occlusions, particularly in detecting reflective vests and helmets.

V. SUMMARY

This paper proposes an enhanced YOLOv8 detection algorithm specifically for reflective vests and helmets. The algorithm first introduces the LightConv module to replace a portion of the convolutional layers in YOLOv8, thereby reducing the model's parameter count and improving detection accuracy. Next, the CPCA attention mechanism is integrated into the network backbone to dynamically allocate attention weights across channel and spatial dimensions, mitigating the issue of detection leakage. Additionally, a 160×160 detection layer is added to enhance the accuracy for small targets. Finally, the C2f modules are completely replaced with RepNCSPELAN4, which utilizes RepNCSP and RepNBottleneck to enable the network to learn hierarchical features more effectively while maintaining computational efficiency. The effectiveness of these improvements is validated through ablation and comparative experiments. Experimental results show that the enhanced LCDR-YOLOv8 model achieves a 4.2% improvement in mAP compared to the original YOLOv8 network and performs robustly in complex backgrounds, making it more suitable for detecting reflective vests and helmets. Future work will focus on the automatic recognition of goggles, safety harnesses, and worker postures in hazardous environments, such as construction sites, to further enhance the detection system. This will contribute to improved workplace safety, reduced accident rates, and better protection for workers' health and safety.

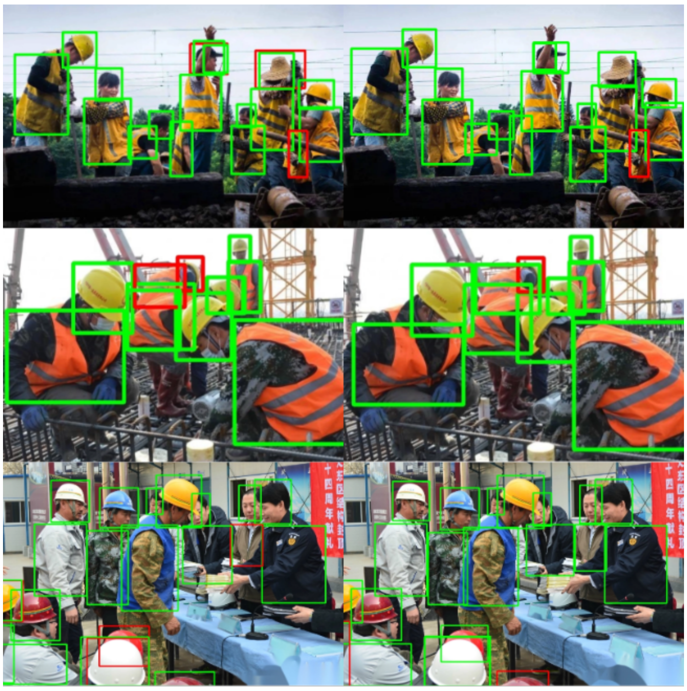

(a) YOLOv8 model (b) LCDR-YOLOv8 model

Fig. 11. Visual inspection map

REFERENCES

[1] China Work Safety Big Data Platform Work Safety Type Accident Briefing (2023 Report)[2024.1.4] https://www.safetybd.cn/aqscsjpt

[2] Xiaohui Liu, Xining Ye. ”Application of skin color detection and Hu moments in helmet recognition,” Journal of East China University of Science and Technology (Natural Science Edition), Vol.40,No.3,pp.365-370,2014.

[3] Park, Man-Woo, Nehad Elsafty, and Zhenhua Zhu. ”Hardhat-wearing detection for enhancing on-site safety of construction workers.” Journal of Construction Engineering and Management 141, no. 9 (2015): 04015024.

[4] Yanqing Zhou, Heru Xue, Xinhua Jiang, et al. ”Low resolution helmet recognition based on LBP statistical features,” Computer System Applications, Vol.24,No.7,pp.211-215,2015.

[5] Guodong Sun, Chao Li, Hang Zhang. ”A helmet wearing detection method incorporating self-attention mechanism,” Computer Engineering and Applications,Vol.58,No.20,pp.300-304,2022.

[6] Min Zhao, Guoliang Yang, Jixiang Wang. ”Improved real-time helmet detection algorithm for YOLOv7-tiny,” Radio Engineering, Vol.53,No8,pp.1741-1749,2023.

[7] Yulu Liu, and Ying Tian, ”DCMS-YOLOv5: A Dual-Channel and Multi-Scale Vertical Expansion Helmet Detection Model Based on YOLOv5,” Engineering Letters, vol. 31, no.1, pp373-379, 2023

[8] Chen, Junhua, et al. ”Lightweight helmet detection algorithm using an improved YOLOv4,” Sensors 23.3 (2023): 1256.

[9] Han, Ju, et al. ”Safety helmet detection based on YOLOv5 driven by super-resolution reconstruction,” Sensors 23.4 (2023): 1822.

[10] Changxin CHENG, Zeqin JIANG, Li CHENG et al. ”A helmet reflective clothing detection algorithm based on improved YOLOX-S[J],” Electronic Measurement Technology,Vol.45,No.6.pp:130-135,2022.

[11] Zejia Han, Qinkun Xiao, Liqi Zhang. ”Improved SSD-based reflective clothing detection algorithm for safety helmets,” Automation and Instrumentation ,Vol.36,No.9,pp:63-68,2021.

[12] Guobo Xie, Jianhui Xie, Zhiyi Lin et al. “Reflective clothing and helmet detection algorithm based on CT-YOLOX,” Foreign Electronic Measurement Technology,Vol.42,No.10,2023.

[13] ultralytics, Available: https://github.com/ultralytics/ultralytics

[14] Agrawal, Pulkit, Ross Girshick, and Jitendra Malik. ”Analyzing the performance of multilayer neural networks for object recognition.” In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13, pp. 329-344. Springer International Publishing, 2014.

[15] GIRSHICK. ”R. Fast R-CNN.” 2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2016: pp. 1440-1448.

[16] S. Ren, K. He, R. Girshick and J. Sun, ”Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 2017.

[17] LIU W, ANGUELOV D, ERHAN D, et al. ”SSD: Single shot multibox detector.” European Conference on Computer Vision(ECCV). Cham: Springer, 2016: pp. 21-37.

[18] REDMON J, DIVVALA S, GIRSHICK R, et al. ”You only look once: unified, real-time object detection.” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: pp. 779-788.

[19] Lin, Tsung-Yi, et al. ”Focal loss for dense object detection.” Proceedings of the IEEE International Conference on Computer Vision. 2017

[20] LIU, Yao, et al. YOLOv4 Pedestrian Target Detection Based on Embedded Platform. In: 2023 2nd International Joint Conference on Information and Communication Engineering (JCICE). IEEE, pp131-136, 2023

[21] Xiaoming Zhang, and Ying Tian, ”Traffic Sign Detection Algorithm Based on Improved YOLOv8s,” Engineering Letters, vol. 32, no. 1, pp168-178, 2024

[22] Weike Zhang, Yanna Zhao, Yujie Guan, Ting Zhang, Qiaolian Liu, and Weikuan Jia, ”Green Apple Detection Method Based on Optimized YOLOv5 Under Orchard Environment,” Engineering Letters, vol. 31, no.3, pp1104-1113, 2023

[23] Woo, Sanghyun, Jongchan Park, Joon-Young Lee, and In So Kweon. ”Cbam: Convolutional block attention module.” In Proceedings of the European Conference on Computer Vision (ECCV), pp3-19, 2018.

[24] Jie Hu, Li Shen, Gang Sun; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7132-7141

[25] Huang, Hejun, et al. ”Channel prior convolutional attention for medical image segmentation.” arxiv preprint arxiv:2306.05196 (2023).