IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Safety Helmet Detection Based on Improved YOLOv8

Bingyan Lin
Fujian Polytechnic of Information Technology, Fuzhou 350003, China

Corresponding author: Bingyan Lin (e-mail: bingyanlin123456@gmail.com)

This work was supported in part by the Research Project of Fujian Polytechnic of Information Technology (grant number Y22104) and the Key Research Project of Fujian Polytechnic of Information Technology (grant number ZK2023-07).

ABSTRACT Wearing safety helmets can effectively reduce the risk of head injuries for construction workers in high-altitude falls. In order to address the low detection accuracy of existing safety helmet detection algorithms for small targets and complex environments in various scenes, this study proposes an improved safety helmet detection algorithm based on YOLOv8, named YOLOv8n-SLIM-CA. For data augmentation, the mosaic data augmentation method is employed, which generates many tiny targets. In the backbone network, a coordinate attention (CA) mechanism is added to enhance the focus on safety helmet regions in complex backgrounds, suppress irrelevant feature interference, and improve detection accuracy. In the neck network, a slim-neck structure fuses features of different sizes extracted by the backbone network, reducing model complexity while maintaining accuracy. In the detection layer, a small target detection layer is added to enhance the algorithm's learning ability for crowded small targets. Experimental results indicate that, through these algorithm improvements, the detection performance of the algorithm has been enhanced not only in general scenarios of real-world applicability but also in complex backgrounds and for small targets at long distances. Compared to the YOLOv8n algorithm, YOLOv8n-SLIM-CA shows improvements of 1.462%, 2.969%, 2.151%, and 3.549% in precision, recall, mAP50, and mAP50-95 metrics, respectively. Additionally, YOLOv8n-SLIM-CA reduces the model parameters by 6.98% and the computational load by 9.76%. It is capable of real-time and accurate detection of safety helmet wear. Comparison with other mainstream object detection algorithms validates the effectiveness and superiority of this method.

INDEX TERMS Safety helmet detection, YOLOv8 algorithm, YOLOv8n-SLIM-CA, coordinate attention mechanism, Slim-Neck.

I. INTRODUCTION

In work environments such as construction sites, tunnels, and coal mines, wearing a safety helmet is one of the fundamental requirements to ensure personnel safety. It effectively reduces the risk of head injuries when construction workers fall from heights, providing crucial protection [1–3]. Monitoring whether individuals are wearing safety helmets as per regulations relies on video data collected by cameras and assessed through manual supervision. However, traditional monitoring methods face increased labor costs, surveillance fatigue, and subjective judgments. Therefore, the development of high-performance safety helmet detection algorithms holds significant importance.

Safety helmet detection methods have been enhanced with the continuous advancement of algorithms in computer vision and improvements in computational capabilities. As a highly regarded technology, deep learning has found widespread application in safety helmet recognition. Compared to traditional methods, deep learning algorithms, especially the YOLO series, have achieved a remarkable balance between accuracy and speed [4]. However, Yolo-based safety helmet detection methods still encounter challenges in achieving high accuracy for small targets in complex backgrounds. Complex environments may feature numerous interfering objects, such as buildings and trees, making it difficult for the algorithm to locate and identify safety helmets accurately. Additionally, safety helmets' small and monochromatic features make them susceptible

to interference from other objects in complex backgrounds, leading to misjudgments. Complex backgrounds may also involve occlusion phenomena, such as overlapping crowds and passing vehicles, causing the safety helmet's shape to be incomplete or partially obscured, making it challenging for the algorithm to identify.

Since the introduction of the YOLO single-stage object detection algorithm, it has garnered widespread attention from industry scholars. In recent years, the YOLO algorithm has undergone continuous optimization. In 2023, the Ultralytics team proposed the YOLOv8 version, which, while meeting real-time requirements, exhibits high detection accuracy and a lightweight network structure suitable for object detection.

The progress in object detection has inspired the development of safety helmet detection methods using deep learning. Numerous researchers assert that deep learning technology represents a crucial avenue for tackling construction security management challenges. YOLOv8 employs deep learning technology, enabling it to learn and comprehend complex visual features. This capability ensures robust execution of helmet detection tasks under varying lighting, angles, and background conditions. However, the current usage of YOLOv8 in safety helmet detection is limited, and its effectiveness in detecting small targets in complex backgrounds could be better. Therefore, based on the YOLOv8 algorithm, we optimized the model to enhance its accuracy, introducing the novel YOLOv8n-SLIM-CA safety helmet detection algorithm. The main contributions of this paper are as follows:

(1) To enhance the small-scale safety helmet dataset and substantially boost the algorithm's detection accuracy, this approach utilizes the mosaic data augmentation method. By employing random augmentation and diverse scaling techniques on the dataset, numerous small targets are generated. This deliberate augmentation and unpredictable scaling contribute to strengthening the network's overall robustness.

(2) This paper introduces the YOLOv8n-SLIM-CA model. The model incorporates three key enhancements: the integration of a Coordinate Attention (CA) mechanism, the adoption of a Slim-Neck structure, and the addition of a small target detection layer. These improvements collectively strengthen the model's detection capabilities for complex background objects and small targets.

(3) The model proposed in this paper is tested on the SHWD dataset in comparison with various algorithms. The results indicate that the algorithm surpasses other algorithms in terms of detection performance. The algorithm exhibits higher robustness in real-world scenarios and across different working environments.

The remainder of the paper is organized as follows: Section Ⅱ describes the literature review. In Section Ⅲ, we discuss the model's architecture. Section Ⅳ presents the experimental analysis results and discussion. Finally, we summarize and conclude our work in Section Ⅴ.

II. LITERATURE REVIEW

A. RELATED RESEARCH INTO THE SAFETY HELMETS DETECTION

Traditional methods heavily rely on manually extracting image features for detection algorithms. For instance, Dahiya et al. [5] utilized local binary patterns, gradient histogram features, and scale-invariant feature transforms to extract safety helmet features. They classified the wearing status using a support vector machine (SVM). However, the reliance on gradient histogram operators, primarily intended for describing edge features, leads to relatively high error detection rates when similar objects to helmet edge features appear in images. To address this issue, Rubaiyat et al. [6] combined color features with CHT (Circular Hough Transform) features, achieving an 81% detection accuracy. Park et al. [7] employed HOG (Histogram of Oriented Gradients) features and then utilized SVM for safety helmet detection. Mneymneh et al. [8] determined helmet-wearing status through spatial information matching. Nonetheless, these traditional methods exhibit poor robustness and low real-time capabilities, limiting their use to specific scenarios and failing to meet the dynamic demands for real-time and versatile safety helmet detection.

B. DEEP LEARNING-BASED OBJECT DETECTION

In recent years, deep learning has emerged as a prominent technology in machine learning, finding extensive applications in object detection. Compared to traditional methods, deep learning holds significant advantages for safety helmet recognition. It leverages convolutional neural networks (CNNs) to extract higher-level features, improving the accuracy and speed of safety helmet recognition.

Deep learning-based safety helmet detection techniques fall into two primary categories: two-stage detection algorithms based on candidate regions and one-stage detection algorithms based on regression.

Two-stage detection algorithms generate a series of candidate boxes, extract features from each, and subsequently use a region classifier for prediction. Girshick's region-based convolutional neural network (R-CNN) [9] is used for extracting image information. However, R-CNN needs help generating candidate boxes with complex backgrounds, potentially resulting in the loss of image information during the feature extraction process. To address this, Ross et al. [10] proposed Fast R-CNN, which replaced SPP-Net's spatial pooling layer, simplifying the network model and saving computational resources. Nonetheless, region pruning relies on selective search methods to generate interested areas. In the same vein, Ren et al. [11] introduced Faster R-CNN, employing a Region Proposal Network (RPN) instead of traditional region prediction algorithms and enhancing image robustness using fully connected layers.

However, faster R-CNN cannot share parameters among multiple related regions in the second stage, adding computational burden. Furthermore, fully connected layers might lead to information loss [12]. Due to their generation of numerous candidate boxes, two-stage detection algorithms have slower detection speeds, failing to meet the real-time demands of safety helmet detection on construction sites.

On the other hand, one-stage algorithms accomplish object classification and position prediction in a single feature extraction. The progress in one-stage detection has motivated the development of a safety helmet detection technique. The YOLO (You Only Look Once) algorithm, a popular one-stage algorithm, has undergone improvements by various scholars for safety helmet wear detection [13]. Modifications to YOLOv3 involved enhancing the feature fusion steps, using upsampling to blend high-level features with low-level ones. Cheng et al. [14] replaced the original convolutional layers in YOLOv3-tiny with depthwise separable convolutions and residual blocks, reducing parameter and computational load while enhancing spatial pyramid pooling modules for more feature extraction. Improvements in YOLOv4 [15] employed a lightweight network to increase detection speed, using the PP-LCNet lightweight network as the backbone and employing depthwise separable convolutions to reduce model parameters. Li et al. [16] reevaluated sample selection methods in the YOLO series and introduced a Hierarchical Positive Sample Selection (HPSS) mechanism during training, improving YOLOv5's fitting ability.

Additionally, inspired by target detection in continuous frame videos, a post-processing algorithm based on box density effectively suppressed false detections. YOLO-M[17] was introduced to tackle issues in helmet wearing detection algorithms, such as excessive parameters, high detection interference, and low accuracy. It utilized MobileNetv3 for feature extraction in YOLOv5s, reducing model parameters and size. Bao et al. [18] integrated the C2F (a faster version of the CSP Bottleneck with two convolutions) module and the FE (FasterNet with EMA) module into the YOLOV8 network architecture, creating a new attention mechanism module named C2F-FE. This module enhances the model's perception of safety helmet targets by fusing features from different levels and incorporating attention mechanisms, simultaneously reducing computational expenses.

The YOLOv8 algorithm introduces new improvements over YOLOv5, exhibiting outstanding performance in object detection and achieving an unprecedented balance in accuracy and speed [19,20]. Despite numerous scholars enhancing YOLO algorithms for safety helmet detection, the accuracy of single-stage algorithms could be higher when detecting small objects or encountering complex background interference. Therefore, urgent improvements are needed in algorithms to achieve better performance for small objects in complex environments.

However, in some current helmet detection algorithms, two-stage algorithms have a large number of parameters and slower detection speed, making it challenging to meet real-time demands. Although one-stage algorithms are faster, their accuracy is lower compared to two-stage algorithms, especially in identifying small and dense targets. To address these issues and achieve high accuracy and fast detection speed, this paper selects YOLOv8 [18] as the base network. The improved YOLOv8 algorithm discards the original feature fusion Neck structure and adopts the lightweight Slim-Neck [30] as the feature fusion network, which is significantly superior to other lightweight networks such as Xception [28] and ShuffleNet [29] in terms of inference latency and accuracy balance. Through these lightweight improvements, the parameter count of YOLOv8 is significantly reduced. However, the enhancement from lightweight improvements may lead to a decrease in accuracy and poor performance in detecting small targets. Therefore, to mitigate the impact of lightweight improvements without increasing the parameter count, this paper adopts three methods. First, the coordinate attention mechanism (CA) [27] is introduced and added to the outputs of the backbone network, allowing the network to acquire inter-channel information and direction-related location information, aiding in better target localization. Second, to address YOLOv8's suboptimal detection of small and dense targets, the detection head is increased from three to four [31] to enhance the detection capabilities for these targets. Finally, Mosaic [21] is employed to increase the number of small targets during the training process.

III. YOLOv8n-SLIM-CA MODEL

The architecture of YOLOv8 is illustrated in Figure 1. YOLOv8 replaces the C3 structure of YOLOv5 with the more gradient rich C2f structure. Different channel numbers are adjusted for models of different scales, showcasing meticulous fine-tuning of the model structure. No longer a set of parameters applied universally to all models, this alteration significantly enhances model performance. The employed anchor-free detection method directly predicts the target's center point and aspect ratio, as opposed to predicting the position and size of anchor boxes. This approach reduces the number of anchor boxes, enhancing detection speed and accuracy.

While YOLOv8 is a versatile object detection algorithm, it exhibits shortcomings in detecting small targets in complex backgrounds. To address this issue, we propose an improved algorithm based on YOLOv8, named YOLOv8n-SLIM-CA. YOLOv8n-SLIM-CA makes the following improvements: the adoption of Mosaic data augmentation, incorporation of an attention mechanism in the backbone network, utilization of Slim-Neck in the neck network, and the addition of a small target detection layer.

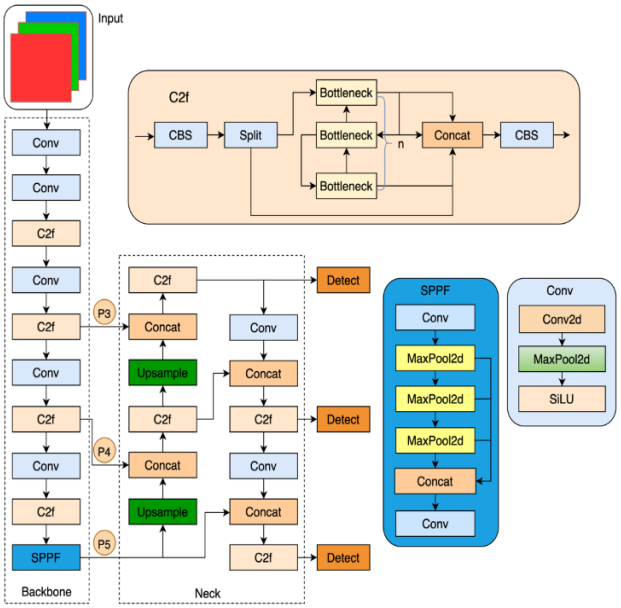


FIGURE 1. **YOLOv8n network architecture.**

A. MOSAIC DATA AUGMENTATION

Mosaic data augmentation constitutes a pivotal technique within the YOLOv8s algorithm [21]. This approach randomly selects four images, sequentially crops, concatenates them clockwise, and ultimately scales them to the designated input dimensions, generating novel sample input data. This strategy enriches the target background, augments the quantity of small targets, and balances the distribution among targets of varying scales. Given the limited categories in the dataset of this study and a training dataset comprising approximately 5000 images, the relatively modest dataset size necessitates data augmentation to enhance the algorithm's generalization capability. As depicted in Figure 2, the mosaic data augmentation approach has been expanded in this study from concatenating four images to nine. This augmentation method yields a considerable number of small targets, enriches the dataset of safety helmet samples, and significantly enhances the algorithm's performance in detecting small scale safety helmets [22,23].

FIGURE 2. **Mosaic data augmentation. (a) Mosaic data augmentation was applied to four images; (b) Mosaic data augmentation was applied to nine images.**

B. COORDINATE ATTENTION MODULE

This study introduces an attention mechanism to enhance the accuracy of safety helmet detection in complex background environments. Inspired by the human visual perception system, the attention mechanism allows neural networks to selectively focus on relevant parts of input data, thereby improving the model's performance in recognizing crucial features [24].

In object detection tasks, attention mechanisms have proven to enhance model performance by directing the model's focus towards important features and diminishing the weight of irrelevant information, thereby improving recognition accuracy. In this field, attention modules such as the channel attention module (SENet) proposed by Hu et al. [25] and the convolutional attention module (CBAM) introduced by Woo et al. [26] have been widely applied. However, to capture long distance dependencies more effectively and retain precise positional information, this paper introduces the CA module [27]. The CA module divides the attention mechanism into two parallel one dimensional feature encoding processes (in the x and y directions), effectively consolidating spatial coordinate information into the generated attention map, thereby enhancing the model's performance. The schematic diagram of the CA attention module is illustrated in Figure 3. Coordinated attention encodes channel relationships and long-term dependencies through precise positional information, involving two steps: coordinate information embedding and coordinate attention generation.

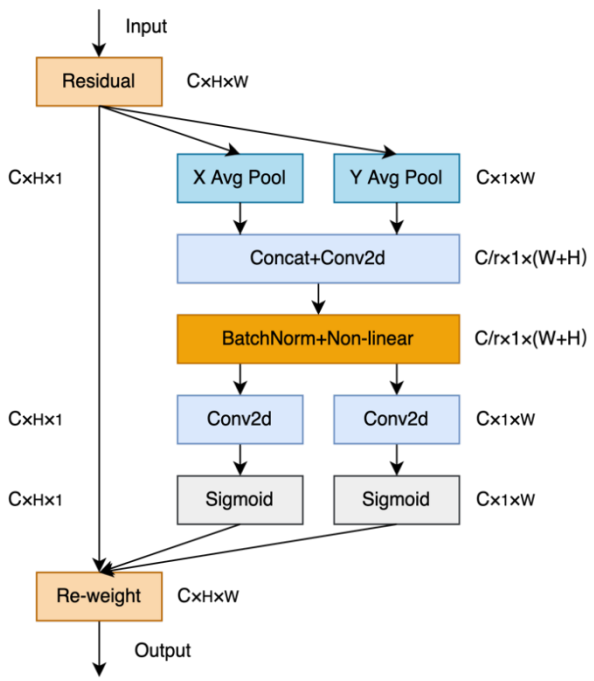


FIGURE 3. **CA module.**

1) EMBEDDING OF COORDINATE INFORMATION

Global pooling methods are commonly employed to channel attention to encode global spatial information. However, this approach compresses global spatial information into channel descriptors, making it challenging to preserve positional information. In order to facilitate the attention module in capturing distant spatial interactions with precise positional information, the CA module decomposes global pooling into a pair of one-dimensional feature encoding operations.

Specifically, for a given input feature map X with dimensions C×H×W, we utilize pooling kernels of sizes (H,1) and (1, W), coding along the x and y directions of each channel in the input feature map to acquire positional information in the horizontal and vertical directions. The computation is expressed as formulas 1–2:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \le i < W} x_c(h, i) \tag{1}$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \le j < H} x_c(j, w) \tag{2}$$

In the equations, $z_c^h(h)$ represents the horizontal projection of the (c)-th channel in the matrix. It involves summing the elements of each row in the (c)-th channel and dividing by the width (W) of the matrix to obtain the average. Similarly, $z_c^w(w)$ signifies the vertical projection of the (c)-th channel in the matrix. This entails summing the elements of each column in the (c)-th channel and dividing by the height (H) of the matrix to obtain the average.

These two transformations enable the attention module to capture long-term dependencies along one spatial direction while preserving precise positional information along another spatial direction. This capability aids the network in more accurately locating the regions of interest within the target, enhancing its overall performance.

2) THE GENERATION OF COORDINATE ATTENTION

After embedding transformation of the information, the concatenated results of formulas (1) and (2) undergo convolutional transformation to derive the attention map, computed as depicted in formula (3):

$$\boldsymbol{f} = \delta(F_1([\boldsymbol{z}^h, \boldsymbol{z}^w])) \tag{3}$$

In this expression, f signifies the feature map of spatial information across horizontal and vertical dimensions. δ represents a non-linear activation function. F_1 denotes the concatenation of pooled results. The CA attention mechanism significantly enhances the neural network's precision with minimal additional computational overhead. To bolster the extraction capability of safety helmet features, we integrated the CA attention mechanism just before the SPPF layer in the backbone network of YOLOv8. The structural modifications, pre- and post-improvement, are illustrated in Figure 4.

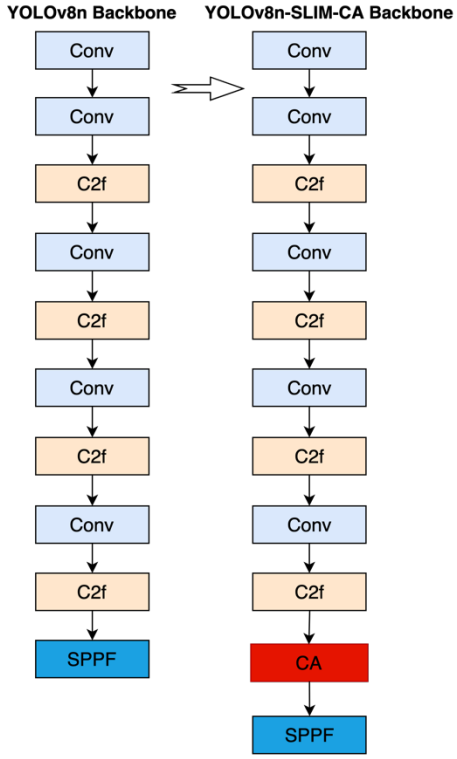


FIGURE 4. Addition of the CA attention mechanism to the backbone network.

As shown in Table 1, we integrated the SE, CBAM, and CA modules into the YOLOv8n algorithm, comparing their respective accuracies. We observed a significant improvement in detection accuracy with the CA module. It achieved the highest mAP@0.5 of 93.143% and concurrently the highest mAP@0.5:0.95 of 59.625% (mAP calculation detailed in formula 8). Figure 5 illustrates the results of different attention mechanisms' heatmaps. The outcomes indicate that the heatmap with the CA module exhibits broader coverage, a stronger focus on targets, and mitigates interference from complex backgrounds.

TABLE 1. Comparison of YOLOv8n with three different attention mechanisms.

Algorithm	P(%)	R(%)	mAP @0.5(%)	mAP @0.5:0.95(%)
YOLOv8n	0.92411	0.8587	0.9221	0.58215
YOLOv8n+SE	0.91023	0.84991	0.91372	0.56497
YOLOv8n+CBAM	0.9227	0.86261	0.92526	0.58444
YOLOv8n+CA	0.92902	0.86654	0.93143	0.59625

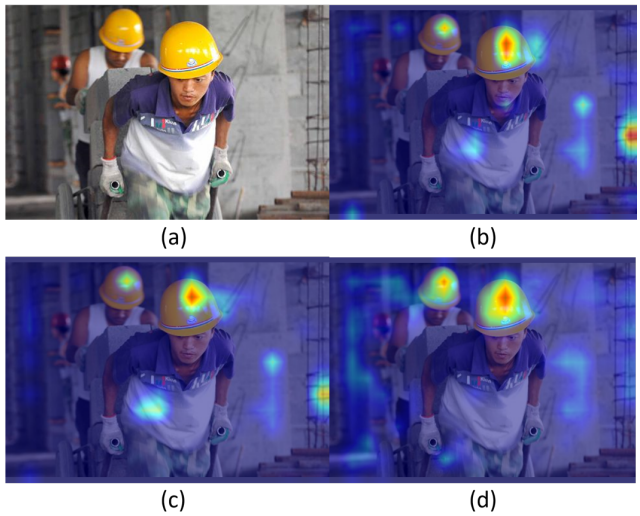


FIGURE 5. Compared the heatmaps after incorporating three different attention mechanisms. (a) Original image; (b) YOLOv8n + SE heatmap; (c) YOLOv8n + CBAM heatmap; (d) YOLOv8n + CA heatmap.

C. SLIM-NECK STRUCTURE

In the YOLOv8n algorithm, many standard convolutions and C2f modules are utilized to enhance accuracy. However, this comes at the cost of reduced speed and increased model parameters. The Slim-Neck structure is employed to fuse features extracted from different-sized feature maps in the backbone network to mitigate model complexity while maintaining accuracy.

To lightweight the network, depthwise separable convolutions (e.g., Xception [28] and ShuffleNet [29]) have been proposed to address the computational cost of standard convolutions effectively. However, these lightweight methods often sacrifice detection accuracy. GSConv [30] combines spatial convolutions (SC), depthwise separable convolutions (DSC), and Shuffle operations, achieving a computational cost of only 60% to 70% compared to standard convolutions while maintaining competitive performance. To enhance the model, this paper opts to replace the standard convolutions in the neck layer with GSConv and introduces the VoV-GSCSP module [30] based on GSConv. The detailed structures of GSConv and the VoV-GSCSP module are depicted in Figure 6 and Figure 7.

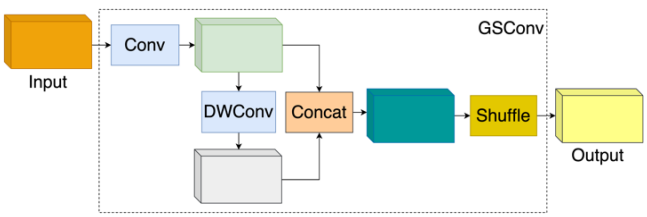


FIGURE 6. GSConv convolution operation.

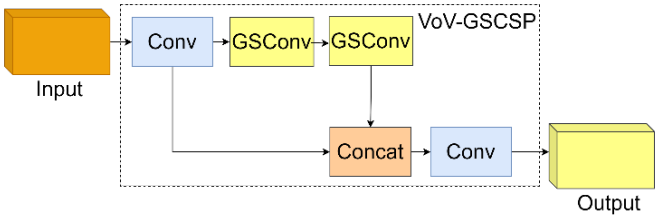


FIGURE 7. VoV-GSCSP structure.

The Slim-Neck structure replaces traditional convolutional networks with the lightweight GSConv. While GSConv's computational cost is approximately 60% to 70% of standard convolutions, its contribution to the model's learning capacity is comparable. A one-shot aggregation method is also introduced to design the cross-level part network (GSCSP) module, VoV-GSCSP. The VoV-GSCSP module effectively reduces computational and structural complexity while maintaining adequate accuracy.

As shown in Figure 8, in the neck design, VoV-GSCSP is employed instead of the traditional CSP, embedding the Slim-Neck module into the YOLOv8 Neck network. The Slim-Neck module is explicitly constructed for object detection tasks, serving as a feature fusion module. Its design aims to enhance model speed and efficiency by reducing network parameters and computational load. The module operates by adding low dimensional feature maps to input feature maps, followed by convolutional operations to extract richer semantic information, effectively boosting model speed and efficiency. Table 2 illustrates that the YOLOv8n algorithm with Slim-Neck reduces FLOPs by 9.76%, parameters by 6.98%, and increases speed by 9.52% compared to the original YOLOv8n algorithm.

TABLE 2. YOLOv8n performance comparison with the Slim-Neck structure.

Algorithm	FLOPs (GB)	Parameters (MB)	Speed (ms)
YOLOv8n	8.2	3.01	2.1
YOLOv8n+Slim-Neck	7.4	2.80	1.9

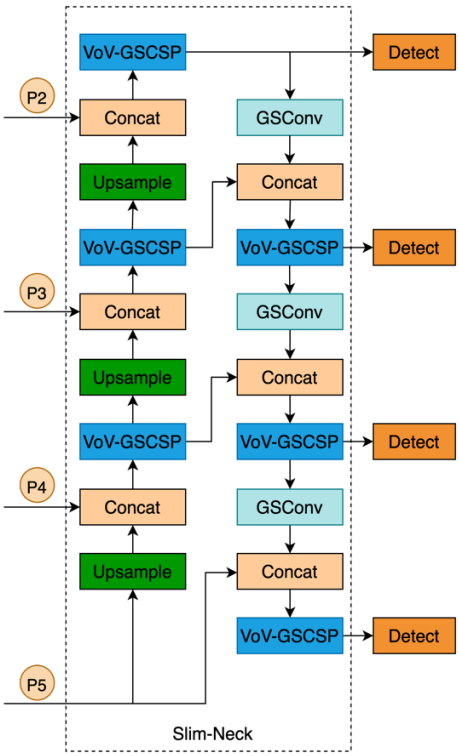


FIGURE 8. **Slim-Neck embedded in YOLOv8n structure.**

D. ADDING A SMALL TARGET DETECTION LAYER

In practical construction scenarios, personnel are distributed across various locations on the worksite. The safety helmets in an image may exhibit various scales, particularly in images containing densely packed targets, encompassing large, medium, and small scales [31]. To address the detection of small objects in complex scenes, we augmented the original YOLOv8n algorithm by introducing a small object detection layer, thereby increasing the number of detection layers to four. Adding these layers facilitates the extraction of more scale specific feature information [32], enhancing the model's ability for multiscale learning in intricate environments. This, in turn, enables improved learning of multilevel feature information, ultimately enhancing the model's detection performance.

The original YOLOv8 detection layer outputs feature maps of three sizes: 20×20, 40×40, and 80×80. However, effective detection becomes challenging when the targets wearing safety helmets are too small. Therefore, based on the original YOLOv8n model in this study, we introduced a small object detection layer with a size of 160×160 to enhance sensitivity to small objects. As illustrated in Figure 8, we extracted features from the output of the leading network's P2, P3, P4, and P5 layers in the YOLOv8n-SLIM-CA model and achieved feature fusion through the Slim-Neck network at the neck level. Finally, we added the small object detection layer at the output layer [32]. Despite a slight increase in computational load due to this enhancement, it significantly elevated the YOLOv8n-SLIM-CA model's performance in small target detection, effectively reducing false positives and negatives across different scales.

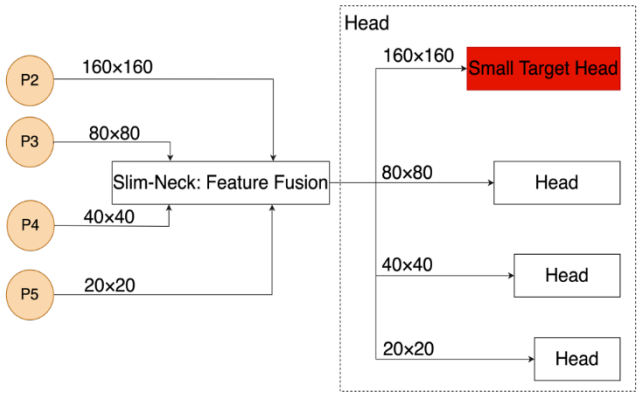


FIGURE 9. **Adding a small target detection layer (indicated by a red background box).**

E. THE ARCHITECTURE OF YOLOV8N-SLIM-CA

The overall structure of our improved model is depicted in Figure 10. We introduced a CA attention mechanism and a small object detection layer into YOLOv8n to enhance the interaction and expressive capability of different level features. Employing GSConv convolution and Slim-Neck paradigm design allowed us to reduce the model's computational and parameter load, thereby enhancing the model's operational efficiency.

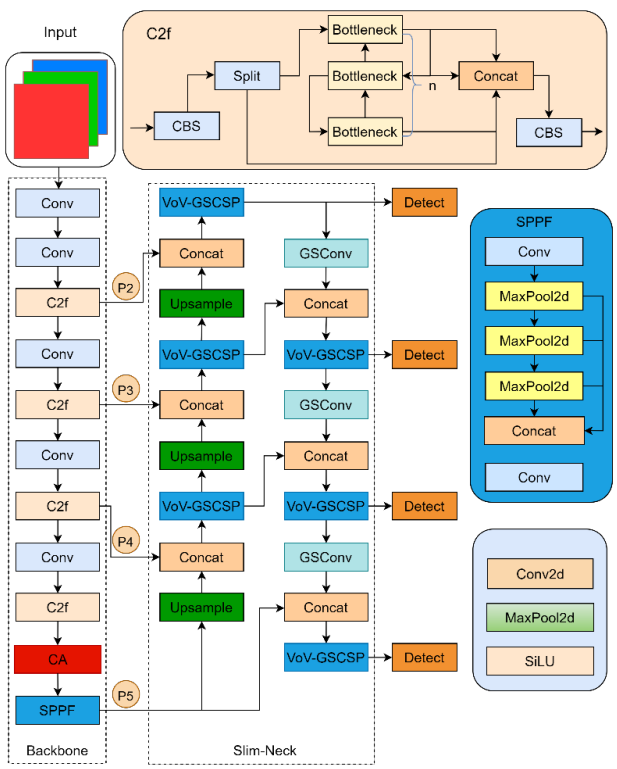


FIGURE 10. **YOLOv8n-SLIM-CA network structure.**

IV. EXPERIMENTAL ANALYSIS AND DISCUSSION

A. DATASET

In deep learning research, the caliber of the dataset profoundly influences the quality of the network model. This study employs the Safety Helmet Wearing Dataset (SHWD) [33–35], comprising 7,581 images sourced from diverse application scenarios, accessible online at https://github.com/njvisionpower/Safety-Helmet-Wearing-Dataset.

The dataset is randomly partitioned into training, validation, and test sets in a 7:2:1 ratio to assess the model. During training, all images undergo mosaic data augmentation. The dataset encompasses targets of various scales within safety helmet images, including large, medium, and small-scale objectives across diverse scenarios. The data is formatted into YOLO standards and categorized into two classes (head without safety helmet and helmet with safety helmet), providing bounding box coordinates for target positions. As depicted in Figure 11, dataset analysis and visualizations reveal a slight imbalance in inter-category sample distribution, addressed during the mosaic data augmentation phase. In Figure 11 (c), where x and y represent the center coordinates of the target box, darker colors indicate denser distributions of target box centers. Figure 11(d) illustrates the width and height, denoting the dimensions of the targets in the images. The dataset exhibits a relatively uniform object distribution, with a significant proportion of medium to small-sized objects and instances of object occlusion, aligning with real world scenarios.

B. EXPERIMENTAL SETUP AND EVALUATION METRICS

For experimentation, the PyTorch 2.10 deep learning framework was employed. The experimental environment featured an Intel Core i5@2.90 GHz processor, 32 GB of memory, and the Ubuntu 18.04 operating system. Training acceleration utilized an NVIDIA GeForce RTX 3080 Ti. The Adam optimizer was employed with an initial learning rate of 0.01, a momentum parameter of 0.937, a weight decay of 0.0005, and a warmup learning rate for the first three epochs to mitigate early-stage overfitting. Following the warm-up phase, a cosine annealing schedule was applied to update the learning rate. Training occurred over 100 epochs, with image pixel dimensions set at 640×640 for training and testing.

Model performance evaluation utilized precision (P), recall (R), and mean average precision (mAP) as critical metrics. Precision gauges the accuracy of model detections (i.e., positive predictive value), while recall assesses the comprehensiveness of model detections (i.e., sensitivity). Single class precision is calculated using integral methods, considering precision-recall curves and the area enclosed by the axes. The mAP value, obtained by summing individual class precision and dividing by the number of classes, is generally computed at an intersection over union (IOU) of 0.5, denoted as mAP@0.5 [36]. The formulas for parameter metrics are presented in Equations 4-8:

$$I_{IOU} = \frac{A \cap B}{A \cup B} \tag{4}$$

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

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

$$A_P = \int_0^1 P\,(r)dr \tag{7}$$

$$mAP = \frac{\sum_{i=1}^{N} A_{P,i}}{N} \tag{8}$$

In the given expressions, A and B denote the sets of predicted and actual bounding boxes in the given expression, respectively. *TP* (true positive), *FP* (false positive), and *FN* (false negative) represent the quantities of correctly predicted, incorrectly predicted, and missed safety helmet targets, respectively. *P(r)* signifies the smoothed precision-recall curve, and integrating it yields the area under the smoothed curve. N denotes the number of detection classes, set to 2 in this context, corresponding to the classes of detected and undetected safety helmets. $A_{P,i}$ denotes the precision of the i-th class, where i is the index. The proximity of predicted and actual bounding boxes is determined by *IOU*, with *IOU* set to 0.5. Average precision (*AP*) is computed for each class at IOU 0.5, and the mean AP (*mAP*) is obtained by averaging across all classes. Detection results with an *IOU* greater than 0.5 are correct, and the corresponding *mAP* value is labeled mAP@0.5. mAP@0.5:0.95 represents the average *mAP* across different *IOU* thresholds (from 0.5 to 0.95, with a step size 0.05).

Additionally, model complexity is measured through parameters and floating-point operations (FLOPs), while speed detection time is used to compare and assess the model's detection speed.

C. MODEL TRAINING

To validate the improvement of the proposed algorithm, YOLOv8n-SLIM-CA, we conducted model training using the same training set and hyperparameters as the original YOLOv8n algorithm. Figure 14 illustrates the final mAP@50 curve and loss curve for comparison. Figure 12(a) shows that the mAP@0.5 of the YOLOv8n algorithm stabilizes around 0.89 after approximately 30 iterations and reaches 0.922 after 100 iterations. In contrast, the proposed improved algorithm, YOLOv8n-SLIM-CA, achieves a mAP of 0.91 or higher after around 18 iterations and ultimately reaches 0.944.

Figure 12(b) shows the overall loss comparison, with the YOLOv8n model dropping to 3.31 after about 40 iterations and converging to 2.97. The improved algorithm reduces the loss to 3.12 after approximately 30 iterations and converges to 2.78.

Compared to YOLOv8n, the improved algorithm YOLOv8n-SLIM-CA not only performs better in terms of accuracy but also exhibits superior convergence speed.

D. ABLATION EXPERIMENT

As shown in Table 3, eight data sets were trained and compared. The original YOLOv8n algorithm is the baseline, and "+" indicates mixed module improvements. The results indicate that the original YOLOv8n algorithm achieves precision, recall, mAP50, and mAP50-95 of 92.411%, 85.87%, 92.21%, and 58.215%, respectively. With each improvement, there is a noticeable enhancement in metrics. The simultaneous application of all four improvement modules yields the best results. Compared to the baseline algorithm, precision, recall, mAP50, and mAP50-95 increase by 1.462%, 2.969%, 2.151%, and 3.549%, respectively, further validating the feasibility of the improved algorithm. Ablation experiment results demonstrate that the combined use of improvement modules can enhance precision, recall, and mAP values, further optimizing algorithm performance.

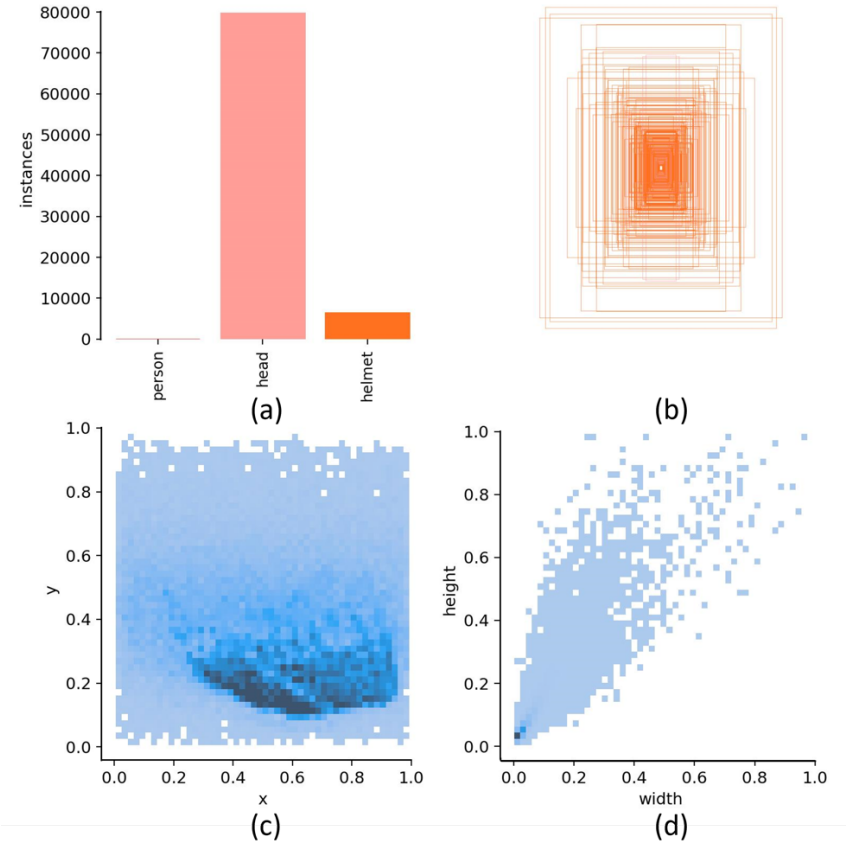


FIGURE 11. Visualization results of the training data. (a) Histogram illustrating the number of instances for each category; (b) Distribution of bounding boxes for all data; (c) Histogram of x and y variables, displaying the spatial distribution of the dataset; (d) Histogram of width and height variables illustrating the dataset's distribution.

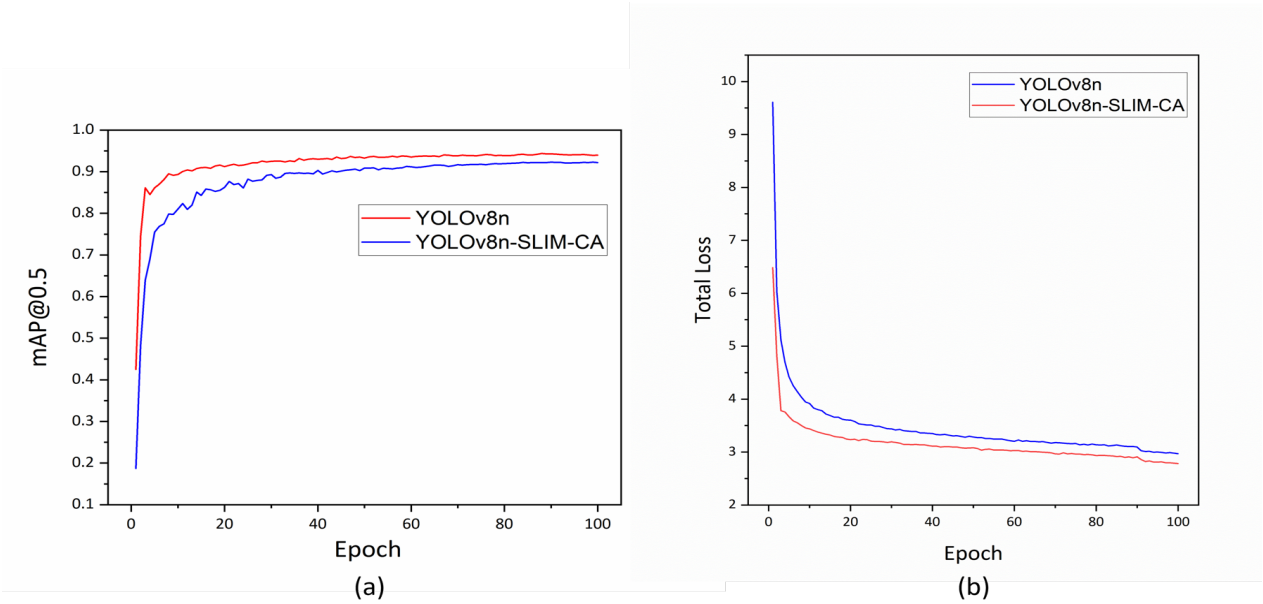


FIGURE 12. Comparison of mAP and loss between YOLOv8n and YOLOv8n-SLIM-CA. (a) mAP(%). (b) loss.

TABLE 3. Comparative analysis of ablation experiments.

Algorithm	P(%)	R(%)	mAP @0.5(%)	mAP @0.5:0.95(%)
YOLOv8n	0.92411	0.8587	0.9221	0.58215
YOLOv8n + Mosaic	0.92715	0.85862	0.92372	0.59011
YOLOv8n + Adding small target layer	0.92907	0.86574	0.93265	0.60054
YOLOv8n + CA	0.92902	0.86654	0.93143	0.59625
YOLOv8n + Slim-Neck	0.92153	0.85819	0.92227	0.58685
YOLOv8n + Mosaic +CA	0.93451	0.87081	0.93631	0.6096
YOLOv8n + Mosaic + Adding small target layer + Slim-Neck	0.93113	0.88367	0.93974	0.61215
YOLOv8n + Mosaic + Adding small target layer + Slim-Neck + CA (YOLOv8n-SLIM-CA)	0.93873	0.88839	0.94361	0.61764

E. COMPARISON OF DETECTION ALGORITHMS

In order to comprehensively evaluate the performance of the enhanced algorithm in the task of safety helmet detection, we employed a consistent experimental platform on the same dataset. The proposed YOLOv8n-SLIM-CA model was trained alongside existing object detection algorithms, including SSD (VGG), YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5m, YOLOv8m, YOLOv8s, YOLOv8m, YOLOv8l, YOLO-M [17], PG-YOLO [37] and YOLOv5s-Improved [38]. The corresponding performance statistics are presented in Table 4. The corresponding performance statistics are presented in Table 4. The YOLOv8n-SLIM-CA algorithm performs better than the original YOLOv8n algorithm in terms of both parameters and mAP when inference speed is comparable. The improved YOLOv8n-SLIM-CA has 2.12% and 4.32% increases in the mAP@0.5 and mAP@0.5:0.95, respectively, compared to YOLOv5s-Improved [38]. It also has a 2.32M and 4.7ms decrease in the model parameters and inference speed, respectively. Upon comparison, it is evident that the YOLOv8n-SLIM-CA model, an improvement upon YOLOv8n, exhibits an inevitable increase in FLOPs computational complexity. However, there is a notable decrease in the parameter count. When compared to YOLOv8l, YOLOv8n-SLIM-CA demonstrates comparable performance in terms of mAP @0.5 and mAP @0.5:0.95 while showcasing a significant reduction in both FLOPs computational complexity and parameter count. This translates to a noticeable improvement in computational speed. These findings indicate that the YOLOv8n-SLIM-CA algorithm has achieved a commendable balance between model light weighting and algorithmic performance, surpassing some standard algorithms.

TABLE 4. Performance comparison of improved algorithms and existing object detection algorithms.

Algorithm	FLOPs (GB)	Parameters (MB)	mAP @0.5(%)	mAP @0.5:0.95(%)	Speed (ms)
SSD(VGG)	28.7	12.3	0.90186	0.55685	2.9
YOLOv5n	4.1	1.9	0.91674	0.58047	1.9
YOLOv5s	15.9	7.1	0.92015	0.58088	2.5
YOLOv5m	48.1	21.0	0.93098	0.5948	4.2
YOLOv5l	107.9	46.0	0.93947	0.6158	7.4
YOLOv8n	8.2	3.01	0.9221	0.58215	2.1
YOLOv8s	28.3	11.1	0.93165	0.5976	2.9
YOLOv8m	78.5	25.7	0.93723	0.60955	4.5
YOLOv8l	165.1	43.2	0.94118	0.61407	5.3
YOLO-M [17]	-	5.5	0.9387	-	-
PG-YOLO [37]	-	0.63	0.933	-	32.9
YOLOv5s-Improved [38]	-	5.06	0.924	0.592	7.0
YOLOv8n-SLIM-CA	11.3	2.74	0.94361	0.61764	2.3

F. CASE ANALYSIS

Figure 13 visually compares the algorithm's detection results more intuitively. The first row of images represents the detection results of the original YOLOv8n algorithm, while the second row depicts the results of the YOLOv8n-SLIM-CA algorithm. As illustrated in Figure 13(a), in complex backgrounds, the YOLOv8n-SLIM-CA algorithm accurately identifies all objects, whereas YOLOv8n exhibits three missed detections, high-lighted by yellow boxes. The original algorithm suffers from significant missed detections in complex backgrounds, yet YOLOv8n-SLIM-CA demonstrates precise identification of small targets in such scenarios, accompanied by an enhancement in confidence scores. In Figure 13(b), for densely packed and overlapping small targets, the YOLOv8n-SLIM-CA algorithm correctly identifies them, while YOLOv8n shows suboptimal performance, failing to detect a person with a partially obscured safety helmet. Figure 13(c) showcases the YOLOv8n-SLIM-CA algorithm's ability to identify small targets with partial occlusion, rectifying the missed detections of the original YOLOv8n algorithm on the left side without safety helmets. Moreover, it assigns higher confidence scores to each target category. In Figure 13(d), for distant small targets, the YOLOv8n-SLIM-CA algorithm successfully identifies all, while YOLOv8n erroneously identifies the black camera on the right as an uncovered safety helmet (head).

The shortcomings of the YOLOv8n algorithm in the detection results of Figure 13 can be attributed to various factors. In Figure 13(a), the missed detections of the two individuals are due to their posture changes, crouching, and bending, leading to significant morphological variations. The missed detection on the right is mainly caused by partial occlusion from the person in front, a complex background, and a small target size. In Figure 13(b), the missed detections are primarily a result of severe target occlusion. Figure 13(c) features targets at the left edge of the image, resulting in missed detections due to incomplete target visibility. The error in Figure 13(d) misclassifying the camera as a target is mainly due to the camera's morphology sharing some resemblance with the target. Especially in scenarios involving small targets, the likelihood of misjudgment increases. From the results, it is evident that YOLOv8n lacks robustness in complex backgrounds and with small targets. The YOLOv8n-SLIM-CA model exhibits enhanced generalization capabilities in complex backgrounds and scenarios involving small targets, effectively reducing missed and false detections in dense and occluded small targets.

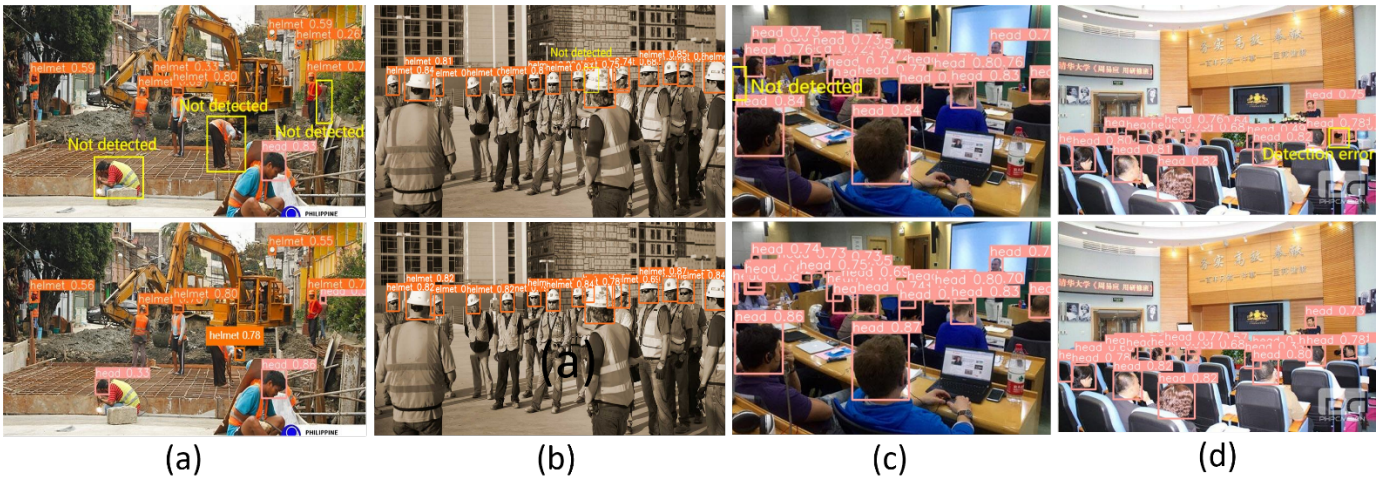


FIGURE 13. Comparison of the detection performance for small objects in various complex scenarios between YOLOv8n and YOLOv8n-SLIM-CA (the first row represents the detection results of YOLOv8, and the second row represents the detection results of YOLOv8n-SLIM-CA). (a) Small targets with complex backgrounds; (b) Densely packed small targets; (c) Small targets at the edge; (d) Distant and small targets.

As illustrated in Figure 14, we substantiated the detection efficacy of the proposed algorithm in diverse environments and workplaces, including open outdoor areas, construction sites, the power industry, mines, engineering operations and traffic safety. In Figure 14 (a) and Figure 14 (b), instances of missed or erroneous detections were observed in both YOLOv8n algorithms. In Figure 14 (c) to Figure 14 (f), both YOLOv8n and the proposed YOLOv8n-SLIM-CA algorithm accurately identified the targets. Notably, the proposed algorithm consistently demonstrated higher confidence in its detections compared to YOLOv8n.

In summary, compared to YOLOv8n, the proposed algorithm's detection performance has significantly improved through strategic enhancements. Notably, it now excels not only in common real-world scenarios but also demonstrates heightened effectiveness in complex backgrounds and for detecting small targets at extended distances. These refinements represent a substantial leap in the algorithm's capabilities, ensuring superior performance across diverse settings and addressing challenges posed by intricate environments and remote targets.

V. CONCLUSIONS

With the continuous advancement of deep learning technology, its positive impact on helmet wearing detection for enhanced workplace safety is evident. However, existing helmet detection models face challenges in recognizing small targets and complex backgrounds. This study proposes and implements an improved algorithm named YOLOv8n-SLIM-CA to address these issues. Through a series of comparative and ablation experiments, the following conclusions are drawn:

Adopting the Slim-Neck structure for feature fusion in the backbone network significantly reduces the model's size and computational load. Specifically, FLOPs decreased by 9.76%, parameters decreased by 6.98%, and speed improved by 9.52%, with minimal compromise on accuracy. Hence, the Slim-Neck structure proves to be an excellent lightweight module.

Secondly, introducing Mosaic data augmentation, a small target detection layer, and the CA module effectively improves accuracy. Mosaic data augmentation enriches the dataset with small scale helmet samples; the small target detection layer aids the model in focusing on multiscale features, especially for small sized targets, thereby enhancing the accuracy of small target helmet detection. The CA attention module outperforms SE and CBAM attention mechanisms, allowing more focused attention on crucial regions and reducing interference from complex backgrounds.

In summary, the proposed YOLOv8n-SLIM-CA algorithm, compared to the YOLOv8n algorithm, achieves a 2.151% improvement in mAP@0.5, reaching 94.361%. Its detection performance surpasses other algorithms in scenarios involving small targets, dense targets, and complex environments. This algorithm meets real-time and accuracy requirements for helmet detection and has low computational demands, with 11.3GB FLOPs, 2.74MB parameters, and 2.3 ms inference speed. It is suitable for deployment on mobile and edge devices, making it applicable for monitoring construction site videos and having broad applications in the industrial sector.

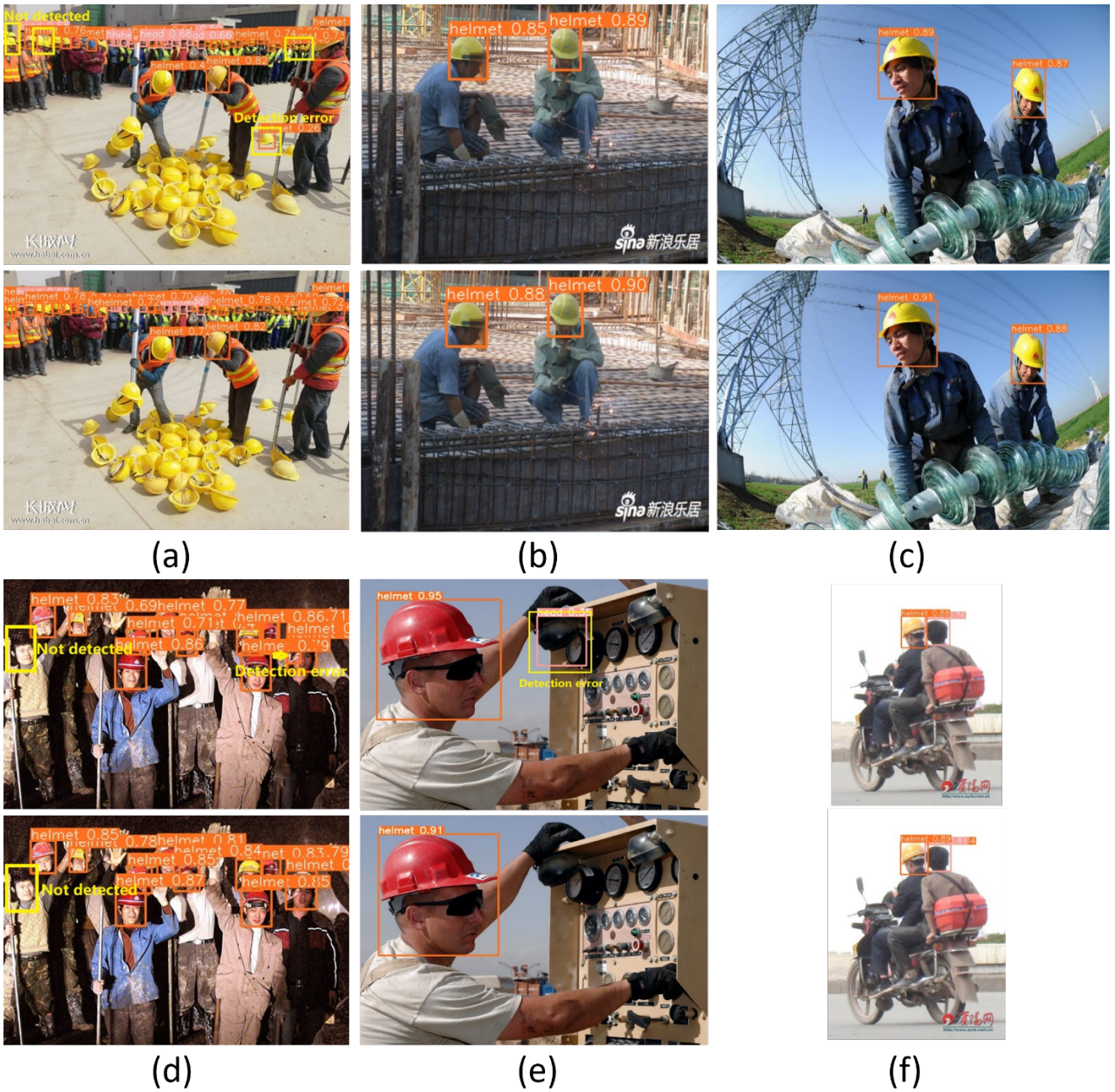


FIGURE 14. Comparison of the detection performance in various real-world application scenarios between YOLOv8n and YOLOv8n-SLIM-CA (the first row represents the detection results of YOLOv8, and the second row represents the detection results of YOLOv8n-SLIM-CA). (a) Open outdoor spaces; (b) Construction sites; (c) The power industry; (d) Mines; (e) Engineering operations; (f) Traffic safety.

REFERENCES

[1] Zhou, Fangbo, Huailin Zhao, and Zhen Nie. "Safety helmet detection based on YOLOv5." 2021 IEEE International conference on power electronics, computer applications (ICPECA). IEEE, 2021.

[2] Huang, Li, et al. "Detection algorithm of safety helmet wearing based on deep learning." Concurrency and Computation: Practice and Experience 33.13 (2021): e6234.

[3] Sanjana, S., et al. "A review on various methodologies used for vehicle classification, helmet detection and number plate recognition." Evolutionary Intelligence 14.2 (2021): 979-987.

[4] Farooq, Muhammad Umer, Muhammad Aslam Bhutto, and Abdul Karim Kazi. "Real-Time Safety Helmet Detection Using Yolov5 at Construction Sites." Intelligent Automation & Soft Computing 36.1 (2023).

[5] Dahiya, Kunal, Dinesh Singh, and C. Krishna Mohan. "Automatic detection of bike-riders without helmet using surveillance videos in real-time." 2016 International Joint Conference on Neural Networks (IJCNN). IEEE, 2016.

[6] Rubaiyat, Abu HM, et al. "Automatic detection of helmet uses for construction safety." 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW). IEEE, 2016.

[7] 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 141.9 (2015): 04015024.

[8] Mneymneh, Bahaa Eddine, Mohamad Abbas, and Hiam Khoury. "Automated hardhat detection for construction safety applications." Procedia engineering 196 (2017): 895-902.

[9] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[10] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision. 2015.

[11] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015).

[12] Liu, Hai, et al. "Infrared head pose estimation with multi-scales feature fusion on the IRHP database for human attention recognition." Neurocomputing 411 (2020): 510-520.

[13] Yang, Wei, et al. "Safety Helmet Wearing Detection Based on an Improved Yolov3 Scheme." International Journal of Innovative Computing, Information and Control 18.3 (2022): 973-988.

[14] Cheng, Rao, et al. "Multi-scale safety helmet detection based on SAS-YOLOv3-tiny." Applied Sciences 11.8 (2021): 3652.

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

[16] Li, Zhishan, et al. "Toward efficient safety helmet detection based on YoloV5 with hierarchical positive sample selection and box density filtering." IEEE Transactions on Instrumentation and Measurement 71 (2022): 1-14.

[17] Wang, Lili, Xinjie Zhang, and Hailu Yang. "Safety Helmet Wearing Detection Model Based on Improved YOLO-M." IEEE Access 11 (2023): 26247-26257.

[18] Bao, Junjie, et al. "Improved YOLOV8 Network and Application in Safety Helmet Detection." Journal of Physics: Conference Series. Vol. 2632. No. 1. IOP Publishing, 2023.

[19] Talaat, Fatma M., and Hanaa ZainEldin. "An improved fire detection approach based on YOLO-v8 for smart cities." Neural Computing and Applications 35.28 (2023): 20939-20954.

[20] Li, Yiting, et al. "A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition." Drones 7.5 (2023): 304.

[21] Wu, Delin, et al. "Detection of Camellia oleifera fruit in complex scenes by using YOLOv7 and data augmentation." Applied Sciences 12.22 (2022): 11318.

[22] Cao, Zhihao, et al. "MaskHunter: real‐time object detection of face masks during the COVID‐19 pandemic." IET Image Processing 14.16 (2020): 4359-4367.

[23] Jiang, Xinbei, et al. "Real-time face mask detection method based on YOLOv3." Electronics 10.7 (2021): 837.

[24] Zhang, Guoqing, et al. "Hybrid-attention guided network with multiple resolution features for person re-identification." Information Sciences 578 (2021): 525-538.

[25] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[26] Woo, Sanghyun, et al. "Cbam: Convolutional block attention module." Proceedings of the European conference on computer vision (ECCV). 2018.

[27] Hou, Qibin, Daquan Zhou, and Jiashi Feng. "Coordinate attention for efficient mobile network design." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.

[28] Chollet, François. "Xception: Deep learning with depthwise separable convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[29] Ma, Ningning, et al. "Shufflenet v2: Practical guidelines for efficient cnn architecture design." Proceedings of the European conference on computer vision (ECCV). 2018.

[30] Li, Hulin, et al. "Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles." arXiv preprint arXiv:2206.02424 (2022).

[31] Zhang, Ke, et al. "Semantic context-aware network for multiscale object detection in remote sensing images." IEEE Geoscience and Remote Sensing Letters 19 (2021): 1-5.

[32] Xiaoguang, Li, Fu Chenpin, and Wang Zhanghui. "Improved Faster R-CNN for Multi-Scale Object Detection [J]." Journal of Computer-Aided Design & Computer Graphics 31.07 (2019): 1095-1101.

[33] Wang, Jie, et al. "Worker's helmet recognition and identity recognition based on deep learning." Open Journal of Modelling and Simulation 9.2 (2021): 135-145.

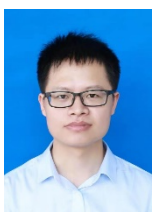
[34] Yue, Shiqin, et al. "Safety helmet wearing status detection based on improved boosted random ferns." Multimedia Tools and Applications 81.12 (2022): 16783-16796.

[35] Sun, Haotian, and Ping Gong. "A Safety-Helmet Detection Algorithm Based on Attention Mechanism." 2021 7th IEEE International Conference on Network Intelligence and Digital Content (IC-NIDC). IEEE, 2021.

[36] Liu, Yifan, et al. "Research on the use of YOLOv5 object detection algorithm in mask wearing recognition." World Scientific Research Journal 6.11 (2020): 276-284.

[37] Dong, Chenhang, et al. "PG-YOLO: A Novel Lightweight Object Detection Method for Edge Devices in Industrial Internet of Things." IEEE Access 10 (2022): 123736-123745.

[38] An, Qing, et al. "Research on safety helmet detection algorithm based on improved YOLOv5s." Sensors 23.13 (2023): 5824.



Bingyan Lin received the B.S. in electronic science and technology and M.S. degree in circuits and systems from Fuzhou University, Fuzhou, China in 2014 and 2017, respectively. Then, He is currently a Lecturer at Fujian Polytechnic of Information Technology, Fuzhou, China. Her research interests include license plate recognition, machine learning, deep learning, computer vision, and image processing.