Indonesian Journal of Electrical Engineering and Computer Science
Vol. 38, No. 3, June 2025, pp. 1765~1781
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v38.i3.pp1765-1781

CGDE-YOLOv5n: a real-time safety helmet-wearing detection algorithm

Wanbo Luo[1,2], Ahmad Ihsan Mohd Yassin[3], Khairul Khaizi Mohd Shariff[3], Rajeswari Raju[4]
[1]School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia
[2]Department of Artificial Intelligence, Leshan Vocational and Technical College, Leshan, China
[3]Microwave Research Institute, Universiti Teknologi MARA, Shah Alam, Malaysia
[4]School of Computing and Mathematics, College of Computing, Informatics and Mathematics,
Universiti Teknologi MARA, UiTM Terengganu Kuala Terengganu Campus, Kuala Terengganu, Malaysia

Article Info

Article history:

Received Aug 6, 2024
Revised Nov 14, 2024
Accepted Nov 30, 2024

Keywords:

CBAM
DCNv2
Efficient-GFPN
Focal-EIoU
Safety helmet-wearing
YOLOv5n

ABSTRACT

Due to numerous parameters and calculations, existing safety helmet-wearing detection models are challenging to deploy on embedded devices. Therefore, this paper proposed a you only look once (YOLO) v5n-based lightweight detection algorithm called CGDE-YOLOv5n to address the shortcomings in the following areas: (i) the YOLOv5n algorithm was selected to minimize the model's parameters and calculations, reducing the hardware cost. (ii) The convolutional block attention module (CBAM) was integrated into the backbone to enhance the network's feature extraction capability. (iii) The neck was improved using the efficient re-parameterized generalized feature pyramid network (efficient RepGFPN) to enhance the multi-scale object detection capability. (iv) The C3 module was improved using the deformable ConvNets v2 (DCNv2) module to enhance the network's adaptability to geometric changes of objects. (v) The complete intersection over union (CIoU) loss was replaced with focal-efficient IoU (focal-EIoU) loss to reduce the missed detection rate. Experimental results demonstrated that the customized gradient descent estimation (CGDE)-YOLOv5n achieved a mean average precision (mAP) 50 of 89.5% and recall of 84%, which is 1% and 0.8% higher than the YOLOv5n. In particular, the recall of workers not wearing safety helmets increased by 1.7%. Furthermore, the improved model achieved a detection speed of 68.5 frames per second (FPS), meeting the real-time requirements.

Corresponding Author:

Ahmad Ihsan Mohd Yassin
Microwave Research Institute, Universiti Teknologi MARA
40450 Shah Alam, Selangor, Malaysia
Email: ihsan_yassin@uitm.edu.my

1. INTRODUCTION

Wearing safety helmets represents an effective method of protecting workers from casualties. Nevertheless, accidents have occurred frequently in recent years due to some workers' lack of safety awareness. The accident rate in the construction industry is markedly higher than that in others [1]. In particular, the incidence of disability resulting from head injuries is the highest [2].

In 2022, 549 production safety incidents happened in housing and municipal engineering projects across China, resulting in 622 fatalities, including 12 major accidents causing 52 deaths [3]. In 2023, these numbers were 584 incidents, 635 fatalities, 8 major accidents, and 28 deaths [4]. From the beginning of 2024 to now, 188 production safety accidents have happened nationwide, resulting in 204 deaths. Among these, two major accidents have happened, resulting in seven deaths [5]. Consequently, the current state of

production safety remains severe, necessitating reinforced safety management to ensure the security of construction operations.

Analysis revealed six main types of accidents: falling from heights, collapses, object strikes, injuries from construction machinery, vehicle injuries, and electric shocks. A case study in Hunan Province revealed that in 2023, 40 production safety accidents occurred, resulting in 42 fatalities. Among these incidents, 26 were due to falls from heights, representing 65%. Six incidents were collapses: five trench pipe network collapses and one wall collapse, accounting for 15%. Three incidents were object strikes and vehicle injuries, representing 7.5%. One incident was a construction machinery injury, and one was an electric shock, each representing 2.5% [6]. Most safety incidents are due to falls from heights, collapses, and object strikes. Therefore, the safety helmet can play a significant protective role in these three types of accidents [7].

The use of safety helmets on construction sites is an effective way to reduce damage from accidents [8]. Consequently, enhancing the administration of safety helmets for construction workers can more effectively guarantee their safety [9]. Currently, the predominant method of monitoring workers' safety helmet-wearing combines manual observation and video surveillance. However, this approach necessitates a significant investment of human resources and is inherently inefficient [10].

With the development of computer vision, one-stage object detection algorithms such as the you only look once (YOLO) series and the single shot multibox detector (SSD) series have been proposed [11], [12]. The one-stage object detection algorithm receives the image at the input end and then outputs the location and class of the object at the output end. This end-to-end technology markedly enhances the detection speed, rendering it well-suited to scenarios where real-time performance is critical. Recently, many scholars have proposed deep learning-based helmet-wearing detection methods, mainly including SSD-based, YOLOv4-based, and YOLOv5-based categories. These methods have proven effective in detecting workers' compliance with wearing safety helmets.

In 2022, Feng and Hu [13] proposed a helmet detection method based on the improved SSD algorithm. The algorithm used four feature fusion modules to combine high-level with low-level features. This enhanced the semantic information of low-level features and improved the algorithm's detection ability for small and medium-sized objects. The algorithm improved mean average precision (mAP) by 2.2 compared to the original SSD algorithm.

Zhan and Pei [14] proposed an improved helmet-wearing detection algorithm based on SSD in 2023. Firstly, the residual network (ResNet)-50 backbone replaced the visual geometry group-16 (VGG-16) backbone of the SSD algorithm to enhance the network's feature extraction capability. Furthermore, the coordinate attention (CA) module was integrated into the backbone to improve the capture of object localization information. The improved algorithm achieved a mAP of 94.5%.

Although the SSD algorithm was once a research hotspot, with the continued evolution of the YOLO series, especially the commercial success of YOLOv5, only a handful of researchers currently are studying the SSD algorithm. Calle Quispe *et al.* [15] analyzed the performance of scaled-YOLOv4 and showed that Scaled-YOLOv4 outperformed previous studies on two public datasets in terms of mAP and F1-score. The model achieved a mAP50 of 96.7 and F1-score of 95.0%.

Chen *et al.* [16] proposed an improved YOLOv4 model for helmet-wearing detection in aerial photography in 2022. The model first increased the channel dimension of the convolutional feature layer in the backbone to improve the utilization of fine-grained features. Secondly, the cross-stage partial (CSP) structure was introduced into the neck to enhance the aggregation efficiency of effective features at different scales. The improved model achieved a mAP of 91.03%.

In 2023, Huang *et al.* [17] designed an adaptive recalibrated multi-scale feature fusion module (ARMFFM) integrated into the original YOLOv4 network to improve the detection accuracy of small targets. Second, a soft complete intersection over union-non-maximum suppression (CIoU-NMS) post-processing algorithm was developed for overlapping object detection. The improved YOLOv4 algorithm achieved an accuracy of 95.1% in indoor helmet-wearing detection.

Xie *et al.* [18] proposed an improved helmet detection algorithm small-medium detection (SMD)-YOLOv4 based on YOLOv4 in 2023. First, the squeeze-and-excitation network (SE-Net) attention module was used to improve the ability of the model backbone network to extract effective features. Second, dense atrous spatial pyramid pooling (DenseASPP) replaced spatial pyramid pooling (SPP) to optimize the extraction of global context information. The mAP of the SMD-YOLOv4 algorithm on the customized dataset reached 97.34%.

In 2024, Li *et al.* [19] proposed a lightweight helmet detection algorithm YOLO-PL (personalized lightweight) based on YOLOv4. First, the enhanced path aggregation network (E-PAN) structure improved the detection accuracy. Second, the diluted convolutional cross stage partial (DCSPX) module with X res unit was proposed to improve the efficiency. This detector outperformed the current object detector in the detection of helmet wearers.

Although the performance of the YOLOv4 algorithm exceeds previous versions of the YOLO series, its model has numerous parameters and calculations. To illustrate, the YOLOv5s model exhibits a similar degree of accuracy to that observed in the YOLOv4 model, yet its size is only one-tenth of that of the YOLOv4 [20]. Consequently, deploying the YOLOv4-based safety helmet-wearing detection model on embedded devices with limited resources is challenging.

In 2023, An *et al.* [21] proposed an improved version of the YOLOv5s network. The network initially integrated the global attention mechanism (GAM) and CBAM into the backbone and neck to enhance the network's ability to extract features. Furthermore, the SCYLLA-IoU loss function was employed instead of the CIoU loss function to accelerate the convergence speed and accuracy of the prediction boxes. The improved network achieved a mAP50 of 92.4%.

Deng *et al.* [22] proposed an enhanced helmet-wearing detection network, YOLOv5-SN in 2023. The network initially employed the ShuffleNet backbone instead of the YOLOv5s backbone, thereby reducing parameters. Secondly, the model was optimized through quantization and layer fusion operations, which resulted in a reduction in computing power and an acceleration of reasoning. The enhanced network exhibited a notable superiority in terms of reasoning speed in comparison to the existing YOLOv5 model.

In 2024, Dong *et al.* [23] employed an enhanced object detection algorithm based on YOLOv5 to identify helmet usage. The algorithm initially incorporated a smaller detection head, which enhanced the algorithm's capability to detect small objects. Secondly, the coordinate attention (CA) module was incorporated to improve the object localization capability. Finally, the normalized wasserstein distance (NWD) was employed instead of the IoU method to quantify the similarity between bounding boxes. The enhanced model achieved a mAP50 of 95.09% in the context of helmet-wearing detection.

In the same year, Hou *et al.* [24] introduced an enhanced object detection algorithm, YOLOv5-GBCW. The algorithm initially employed ghost convolution to transform the backbone network, thereby significantly reducing the complexity of the model. Secondly, a bidirectional feature pyramid network (BiFPN) was utilized to enhance feature fusion, thereby improving the ability to detect small objects. Finally, the Beta wise-IoU loss function is proposed to improve the model's generalizability. The improved algorithm achieved a mAP50 of 94.5%.

Iparraguirre-Villanueva *et al.* [25] implemented a personal protective equipment (PPE) detection system based on the YOLOv5x object detection algorithm in 2024. The system initially converted the video into frames and performed resolution adjustments during the data collection phase. Next, the dataset was subjected to labelling and cleansing, and the labels and bounding boxes were revised. Finally, the detection model was trained on a customized dataset. The model achieved an accuracy of 91% and a recall of 74% for helmets. The accuracy of goggles reached 85% and a recall of 87%. The accuracy of not wearing a mask reached 92% and a recall of 89%.

In 2023, Kisaezehra *et al.* [26] described a system strategy based on a deep learning model of the YOLOV5 architecture for real-time monitoring of workers' helmets. The proposed system employed five distinct models of YOLOV5, namely YOLOV5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, to train the models on a bespoke dataset comprising 7,063 images. The YOLOV5x model exhibited the highest performance, achieving a mAP50 of 95.8%. In contrast, the YOLOV5n model demonstrated the fastest detection speed of FPS 70.4.

Kwak and Kim [27] proposed a method for the automated identification of personal protective equipment, namely helmets and masks, worn by workers in indoor settings. Firstly, the detection algorithm was generated by transfer learning the YOLOv5s and YOLOv5m models. Secondly, the two models were trained by adjusting the learning rate, batch size, and epoch. Ultimately, the model with the optimal performance was selected as the model for detecting masks and helmets. The YOLOv5s model exhibited the most optimal performance, with a mAP50 of 0.954.

Although YOLOv5-based methodologies demonstrated high accuracy in safety helmet-wearing detection, the trained models had numerous parameters and calculations, leading to high hardware costs. Some methods used lightweight techniques, like the ghost convolution module or ShuffleNet backbone, to reduce parameters and computations. However, these models still had significant parameters and calculations. Considering cost-effectiveness, a lightweight model is a viable solution, despite a slight compromise in accuracy. The YOLOv5n model with 1.9 million parameters meets this requirement but has insufficient feature fusion capability, leading to suboptimal detection accuracy.

Most datasets in the above studies are labeled into two classes: 'helmet' and 'person', representing workers wearing and not wearing helmets. However, the label "helmet" is inappropriate for workers wearing safety helmets. Many workers' bodies are obscured in the images, and these should be labeled as "head_with_helmet" instead of "helmet". Consequently, models trained with these datasets lack robustness. Additionally, samples of workers not wearing helmets are insufficient, which leads to missed detections for these objects. Furthermore, since images are captured at construction sites, workers occluded by steel bars and bricks are prone to missed detection.

Balancing cost, speed, and accuracy is crucial for deploying a safety helmet-wearing detector in real-world applications. Therefore, the lightweight nature of the detection model, with minimal parameters and computations, is a significant feature of this paper. The detection model must achieve an optimal balance between accuracy and speed to facilitate the implementation of the safety helmet-wearing detection system. The main contributions of this paper are:

i) Integrating CBAMs into the YOLOv5n network enhances feature extraction capability, compensating for the decrease in detection accuracy of a lightweight network.
ii) Refining and fusing high-level semantic and low-level spatial features using efficient RepGFPN to enhance the YOLOv5n neck, thereby improving detection accuracy.
iii) Introducing the DCNv2 into the C3 module of YOLOv5n enhances the network's adaptability to geometric changes of objects and improves its focus on relevant image areas, reducing the missed detection rate.
iv) Replacing the CIoU loss of YOLOv5n with focal-EIoU loss further reduces the missed detection rate.

Compared with previous studies, this paper adjusts the dataset annotation strategy to improve the model's robustness, which is beneficial for real-world detection. Moreover, previous studies utilize models with numerous parameters to achieve high accuracy. However, this study improves the lightweight YOLOv5n model to reduce parameters significantly and achieve the same high-level accuracy.

2. THE PROPOSED ALGORITHMS

2.1. YOLOv5n

The YOLO series models are classified into five categories: YOLOv5x, YOLOv5l, YOLOv5m, YOLOv5s, and YOLOv5n, based on their sizes. YOLOv5n has the lowest latency on the Tesla V100 b1, at only 0.6 milliseconds. The overall architecture of the YOLO series is similar, differing in network depth and width. The YOLOv5n architecture comprises the backbone, neck, and head. The backbone extracts features from the input image. The neck includes the FPN and path aggregation network (PANet). The FPN fuses feature maps from the backbone and the PANet further extracts the feature. The head predicts objects. Figure 1 displays the YOLOv5n architecture.

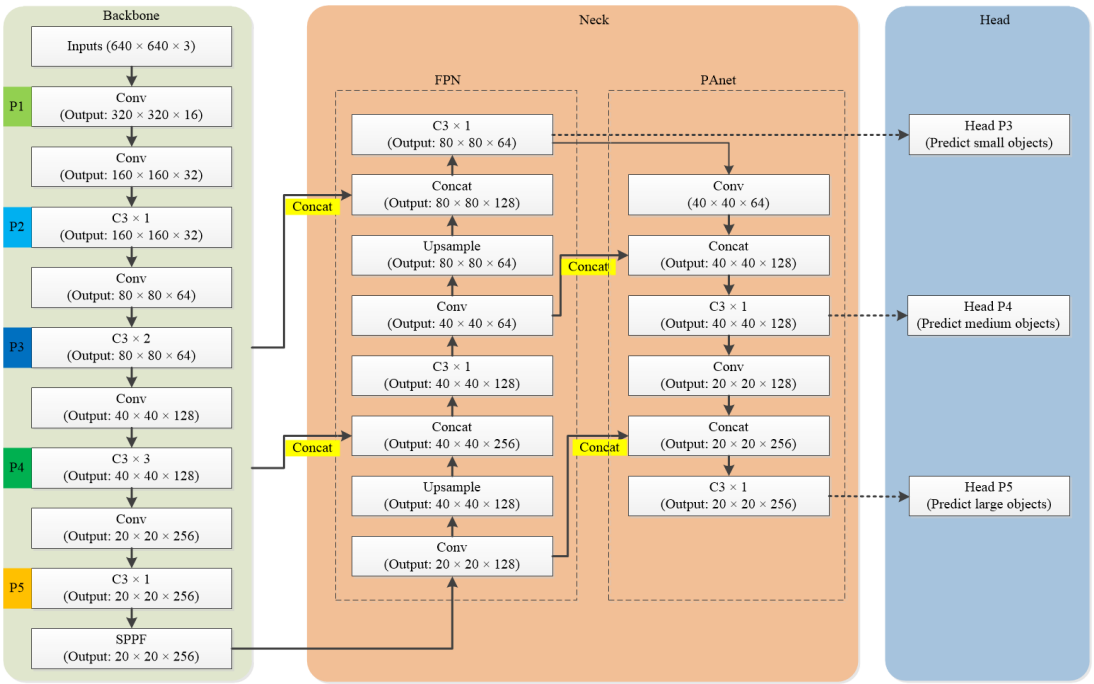


Figure 1. YOLOv5n architecture

The Conv module consists of a convolutional layer, a batch normalization (BN) layer, and the sigmoid linear unit (SiLU) activation function. The C3 module includes Conv and Bottleneck structures, improving the receptive field and reducing computational complexity. The C3 module is repeated two and

three times when generating P4 and P5 feature maps, respectively, and is only performed once at other locations. The Concat module merges two feature maps of the same size. The Upsample module doubles the feature map size while retaining the number of channels. The spatial pyramid pooling fast (SPPF) module enhances model speed and accuracy when processing input images of various sizes.

The backbone generates five feature maps: 320×320 (P1), 160×160 (P2), 80×80 (P3), 40×40 (P4), and 20×20 (P5) through a series of downsampling operations. The neck's FPN uses concatenation to fuse three feature maps from the backbone: P3, P4, and P5. The PANet further extracts features through three downsamplings to generate semantically richer feature maps and fuses them with FPN feature maps. The head uses the fused P3, P4, and P5 feature maps to predict objects.

2.2. CBAM

CBAM was proposed to enhance the representation capability of convolutional neural networks (CNNs) [28]. It helps the network focus on significant features while attenuating inconsequential ones. Thus, CBAM enhances the network's feature extraction capability, improving the model's accuracy. Given an intermediate feature map, CBAM employs a sequential approach to infer the attention maps. This process occurs along two distinct dimensions: channel and spatial. Then, the attention maps are multiplied by the input feature map, resulting in an adaptive refinement of the features. The CBAM module consists of two sub-modules, namely the channel attention module and the spatial attention module. Figure 2 shows the CBAM structure.

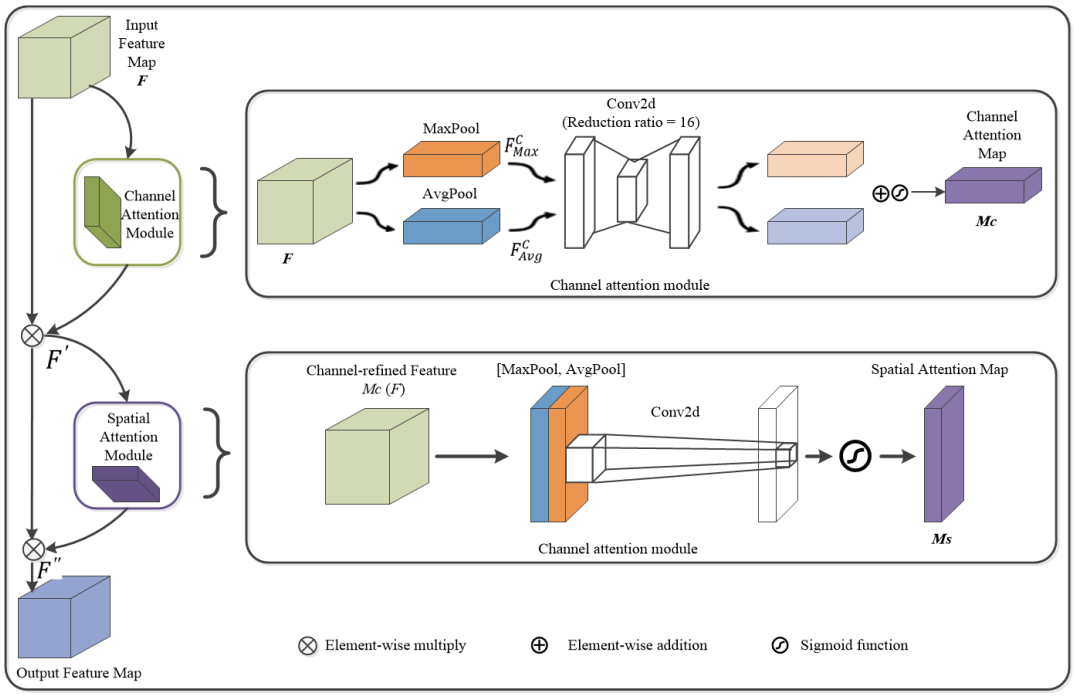


Figure 2. CBAM structure

The spatial dimension of the input feature map is squeezed to compute the channel attention efficiently. Both average-pooling and max-pooling are employed to aggregate spatial information, generating two distinct spatial descriptors: F_{avg}^{c} and F_{max}^{c}. After applying a two-dimensional convolution layer with a kernel size of 1 to each descriptor, the output feature vectors are merged using element-wise summation, and a sigmoid function is subsequently performed to obtain a one-dimensional channel attention map M_c. The M_c is multiplied element-wise with the original feature map to generate the channel-refined feature F'. A reduction ratio of 16 is used to reduce parameter overhead.

Furthermore, the channel-refined feature map F' is aggregated using two pooling operations, generating two-dimensional maps: F_{avg}^{s} and F_{max}^{s}. Those are then concatenated and convolved by a two-dimensional convolution layer, producing a two-dimensional spatial attention map M_s. The M_s is multiplied element-wise with channel-refined feature F' to obtain the final output F''.

2.3. Efficient RepGFPN

The GFPN aims to enhance the fusion and expression capability of the multi-scale feature by improving upon the traditional FPN [29]. It allows the network to establish more direct connections between feature maps of different scales, further enhancing the expressiveness of features. However, the latency of the GFPN-based model is much higher than the FPN-based model. Therefore, the efficient RepGFPN was proposed to address this issue [30].

Concerning topological structure optimization, the efficient RepGFPN uses different numbers of channels under different scale features, which can flexibly control the expressive power of high-level features and low-level features under the constraint of lightweight computation. Concurrently, it removes inefficient upsampling connections in feature fusion. Concerning fusion methods, the efficient RepGFPN uses the fusion block (FB) to perform feature fusion. The FB introduces technologies such as cross-stage-partial (CSP) connection, reparameterization mechanism, and multi-layer aggregation connection to improve the fusion effect further. Figure 3 displays the architecture of the efficient RepGFPN. Given three feature maps with different sizes, the Efficient RepGFPN effectively fused these multi-scale feature maps.

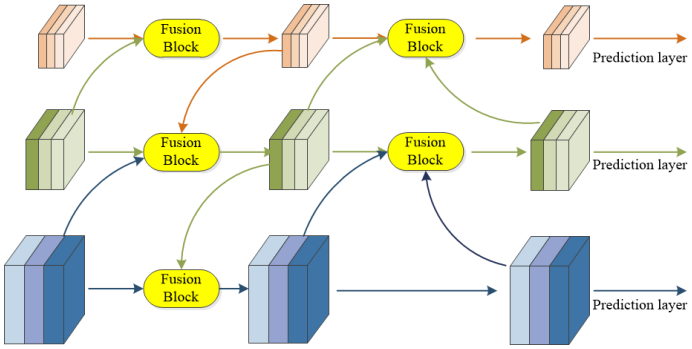


Figure 3. Efficient RepGFPN structure

2.4. DCNv2

The deformable ConvNets (DCN) extends the classic CNN architecture and introduces a deformable convolution operation [31]. This can capture local geometric deformations in the input image to improve the model's adaptability and accuracy to different scenarios, especially object deformation and occlusion. The DCN uses deformable convolution to allow the location of the filter to shift relative to the regular grid to adapt to the local deformation of the image. Concurrently, it uses multi-level pyramid pooling to combine information at multiple scales, enabling the model to handle deformations of different ranges.

However, the visualization results of DCN show that the corresponding position of its receptive field exceeds the object range, resulting in the feature being unaffected by the image content. Therefore, the DCNv2 was proposed to address this issue [32]. The DCNv2 enhances the modelling power for learning deformable convolutions and introduces the modulation mechanism that expands the scope of deformation modelling.

2.5. CIoU

Concerning the evaluation metric for bounding box regression, the most prevalent metric is the IoU. Due to the inherent deficiencies of IoU, the bounding box regression loss function usually aggregates some geometric indicators based on IoU, including the distance, overlapping area, and aspect ratio between the predicted and ground truth boxes. The CIoU loss is employed in the YOLOv5 network, which considers three important geometric factors: the overlap area, the central point distance, and the aspect ratio. Given a predicted box B_p and a ground truth box B_{gt}, the CIoU loss calculation formula is given in (1).

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(B_p, B_{gt})}{(w_c)^2 + (h_c)^2} + \alpha v \quad (1)$$

Where w_c and h_c denote the width and height of the smallest enclosing box that covers two boxes. $\rho(\cdot)$ denotes the Euclidean distance between the center point of the predicted and ground truth boxes. α is a positive trade-off parameter, its calculation formula is given in (2).

$$\alpha = \frac{v}{(1-IoU)+v} \quad (2)$$

Where v is used to measure the consistency of aspect ratio, its calculation formula is given in (3).

$$v = \frac{4}{\pi^2}\left(tan^{-1}\frac{w_{gt}}{h_{gt}} - tan^{-1}\frac{w_p}{h_p}\right)^2 \quad (3)$$

Where w_{gt} and h_{gt} denote the width and height of the ground truth box, while w_p and h_p denote the width and height of the predicted box.

The CIoU is subject to the following limitations. If the aspect ratios of the predicted boxes and the ground truth box are identical, the penalty term associated with the aspect ratio is always zero. Consequently, it is unable to ascertain which prediction box is more accurate. Furthermore, the CIoU enhances the precision of object localization. However, this may result in the missed detection of some objects.

3. METHOD

This paper improved the YOLOv5n network in four components. The first was the integration of three CBAMs into the YOLOv5n backbone network to enhance accuracy. The second improvement was to optimize the YOLOv5n neck using the efficient RepGFPN, which enabled the effective fusion of multi-scale feature maps, thus further improving accuracy. Subsequently, the DCNv2 module was introduced as an improvement for the C3 module of YOLOv5n. This modification enhanced the feature extraction capability of the deformable and occluded objects to decrease the missed detection rate. Finally, focal-EIoU was introduced as a replacement for CIoU to reduce the missed detection rate further. Therefore, the improved algorithm was called CGDE-YOLOv5n.

3.1. Integration of CBAM

Although CBAM can be integrated anywhere in the YOLOv5n network, the different numbers and locations of integrating CBAMs will result in different outcomes. The number of integrating CBAMs directly affects the number of network parameters, particularly when the intermediate feature map contains numerous channels. Through the investigation of the YOLOv5n network, it was found that the YOLOv5n backbone generated feature maps of 80×80, 40×40, and 20×20 sizes and fused the feature maps with the same size obtained by upsampling in the neck. These three fused feature maps predicted small, medium, and large objects. Therefore, this paper integrated CBAMs after the three feature maps in the backbone or neck to extract features efficiently. Figure 4 shows the integration of CBAMs in the backbone.

Three CBAMs are integrated after the P3, P4, and P5 layers, as shown in the yellow area. The methodology employed for the integration of CBAM into the neck was analogous. Three ablation experiments were conducted to identify the most effective method of integrating CBAM.

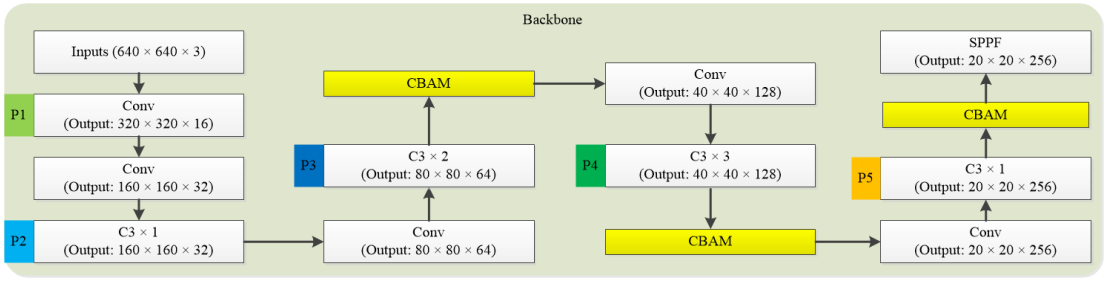


Figure 4. Integrating CBAMs in the backbone

3.2. Improving the YOLOv5n neck

This paper employed the efficient RepGFPN as the YOLOv5 neck to refine and fuse high-level semantic and low-level spatial features. Figure 5 displays the improvement of the YOLOv5n neck using the efficient RepGFPN. The FB includes the concatenation and CSPStage modules, enclosed by a purple line. The concatenation module fuses feature maps from different layers, highlighted in green areas. The CSPStage module performs excellently in deep networks, highlighted in yellow areas. The first FB fuses the feature maps of P5 and P4 from the backbone. Similarly, the second FB fuses the feature maps of P4 and P3.

The third FB fuses only the feature map of P3. The fourth FB further fuses the feature map of P4. The fifth FB further fuses the feature maps of P5 and P4. Figure 6 shows the CSPSatge structure.

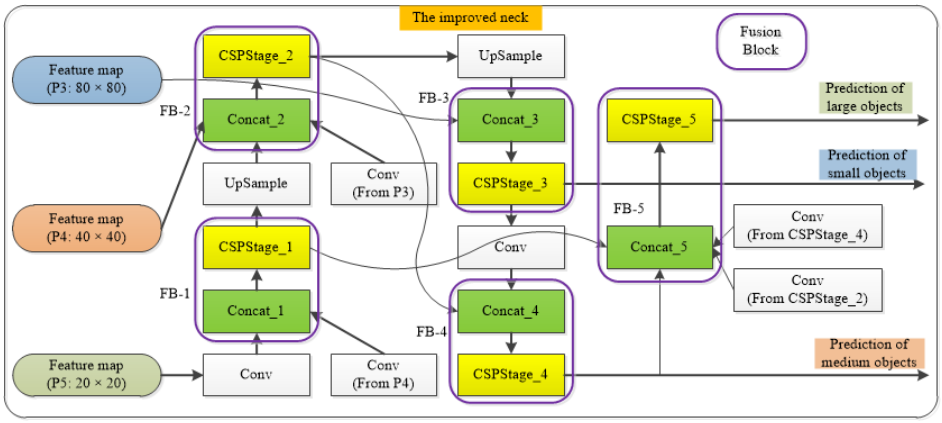


Figure 5. Improvement of the YOLOv5n neck

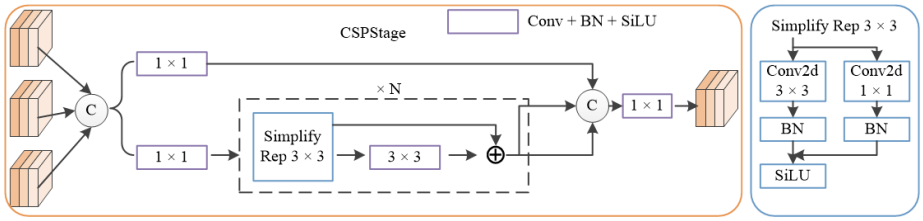


Figure 6. CSPStage structure

BN denotes the batch normalization layer. The N symbol indicates the number of repetitions of the module. The C symbol represents the concatenation module. The CSPStage module receives the concatenated feature map. The feature map splits into two distinct branches to obtain refined feature maps. Then, these feature maps are concatenated and passed through a 1×1 convolution layer to generate the final output.

3.3. Improving the C3 module

The DCNv2 module was introduced into the original C3 module of YOLOv5n to enhance the adaptability to deformed and occluded objects, thereby reducing the missed detection rate of the model. Structures of the C3_DCNv2, Bottleneck, and Bottleneck_DCNv2 modules are shown in Figure 7. The improved C3 module, C3_DCNv2, is shown in Figure 7(a). The input feature map was subjected to two refinement processes. In the left branch, the input feature map passed through a ConvBNSiLU structure to reduce the number of channels to half. In the right branch, the input feature map sequentially passed through a ConvBNSiLU structure and a Bottleneck_DCNv2 structure. After that, the two refined feature maps were concatenated. The concatenated feature map was passed through a ConvBNSiLU structure to generate the final output.

Figure 7(b) shows the Bottleneck structure of the original C3 module. In the left branch, the Shortcut represented the residual connection, which was used to accelerate the convergence of the model and improve the model's accuracy. In the right branch, the input feature map sequentially passed through two ConvBNSiLU structures with kernel sizes of 1 and 3. Finally, the two refined feature maps were added element-wise to generate the output feature map.

Figure 7(c) shows the Bottleneck_DCNv2 structure of the C3_DCNv2 module. The DCNv2 structure superseded the ConvBNSiLU structure to enhance the capacity for feature extraction. Figure 8 displays the improved architecture of the YOLOv5n backbone with C3_DCNv2 modules. Since the backbone comprised four C3 modules, four ablation experiments were conducted to ascertain the influence of the number of C3_DCNv2 modules and the configuration of these modules on the model performance.

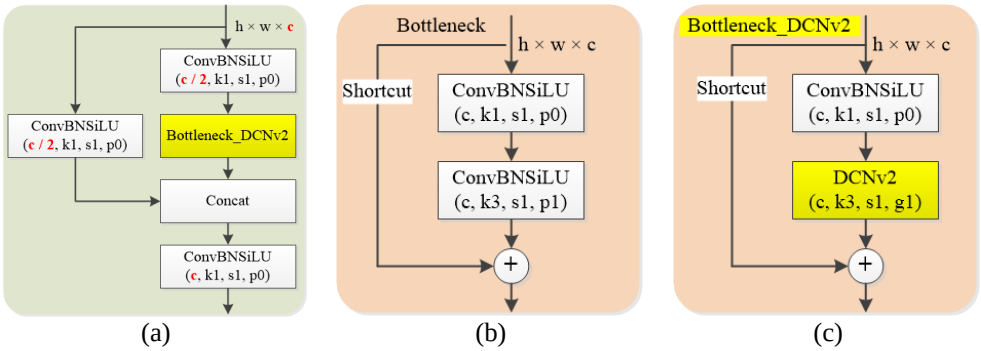


Figure 7. Three modules' structures: (a) C3_DCNv2, (b) Bottleneck, and (c) Bottleneck_DCNv2

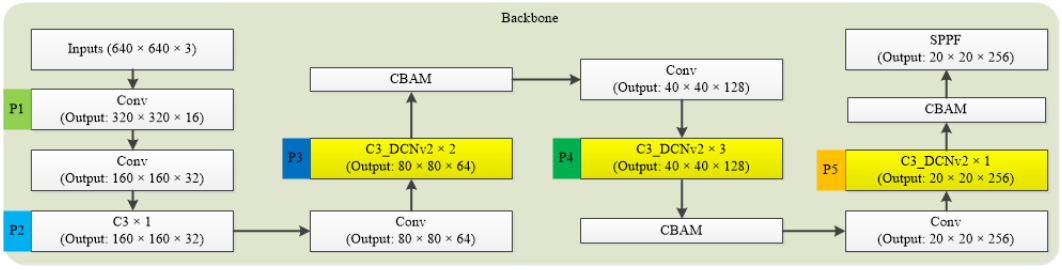


Figure 8. YOLOv5n backbone with C3_DCNv2 modules

3.4. Replacing the CIoU with Focal-EIoU

Compared to the CIoU loss, the EIoU loss directly minimized the discrepancies in width and height between the ground truth box and the predicted box by splitting the aspect ratio. This may reduce the localization precision to increase the recall, thereby reducing the missed detection ratio. The EIoU loss calculation formula is given in (4).

$$L_{EIoU} = 1 - IoU + \frac{\rho^2(B_p, B_{gt})}{(w_c)^2 + (h_c)^2} + \frac{\rho^2(w_p, w_{gt})}{(w_c)^2} + \frac{\rho^2(h_p, h_{gt})}{(h_c)^2} \quad (4)$$

Furthermore, inspired by the focal loss's idea, a reweighted EIoU loss, Focal-EIoU, was designed to enhance the contributions of high-quality predicted boxes in model optimization, while simultaneously suppressing the contributions of low-quality ones. The Focal-EIoU loss calculation formula is given in (5).

$$L_{Focal-EIoU} = IoU^{\gamma} \cdot L_{EIoU} \quad (5)$$

Where $r = 0.5$ denotes the degree of inhibition of outliers.
To ensure a fair comparison, ablation experiments were conducted to evaluate the performance of CIoU, focal-CIoU, EIoU, and focal-EIoU losses.

4. RESULTS AND DISCUSSION

Since the detection rate of workers not wearing helmets in the construction industry takes priority over the precision of locating the object, the precision can be appropriately reduced to increase the recall. Accordingly, the mAP50 and recall metrics of the YOLOv5n and improved models are primarily subjected to comparison. This paper used the YOLOv5n model as the baseline for performance comparison. Subsequently, the final improved model was deployed on the Jetson Orin Nano terminal for real-time detection of safety helmet usage.

4.1. Experimental dataset and environment

The dataset employed in this paper was derived from the open-source dataset SHEL5K [33]. The dataset comprised six classes: "helmet", "head-with-helmet", "person-with-helmet", "head", "person-no-helmet", and "face". However, an increase in the number of categories results in a notable reduction in the

accuracy of the trained model. Therefore, the "helmet", "head", and "face" categories of the original SHEL5K dataset were unlabeled. Furthermore, this labelling strategy addressed the issue of the unreasonable labelling of some datasets. A total of 5,000 images were collected, comprising three classes. The categories were as follows: "person-with-helmet", "head-with-helmet", and "person-no-helmet". Finally, the dataset was randomly divided into a training set and a validation set in a ratio of 9:1.

The experimental environment for training described in this paper was configured with the following hardware and software: the CPU processor was an Intel i7-13400KF, and the graphics card was an NVIDIA GeForce RTX 3060 with 12G graphic memory. The software system was Windows 11, with CUDA version 11.8, Pytorch version 2.0.1, torchvision 0.15.2 and Python version 3.8.17. The training settings were batch size 32, input image size 640 × 640, and 200 epochs. The deployment environment was configured with the following hardware and software: the Jetson Orin Nano had 20 TOPs AI performance, a 1.5 GHz CPU, and a 4G memory. The software system was Ubuntu 20.04, with CUDA version 11.4, Jetpack 5.1.1, TensorRT 8.5.2.2, torch version 1.11.0, torchvision 0.12.0 and Python version 3.8.13.

4.2. Experimental results of improving YOLOv5n

4.2.1. Integration of CBAMs

Three ablation experiments are conducted to verify the effectiveness of the integrated CBAMs. Table 1 presents the performance comparison of the YOLOv5n model and models integrated CBAMs. CBAM-1 indicates that CBAMs are solely incorporated into the backbone of YOLOv5n. CBAM-2 presents that CBAMs are only integrated into the neck of YOLOv5n. CBAM-3 indicates that CBAMs are integrated into both the backbone and neck of YOLOv5n.

The experimental results demonstrated that integrating CBAMs improved the model's accuracy. Compared to the YOLOv5n model, the CBAM-1 model achieved the highest mAP50 of 89.1%, representing a 0.6% increase and a slight rise of 11,049 in parameters. Furthermore, precision and recall were increased by 0.4% and 0.3%, respectively. Therefore, the CGDE-YOLOv5n algorithm selected the first method for integrating CBAMs.

Table 1. Performance comparison of four models

Model	Precision (%)	Recall (%)	mAP50 (%)	Parameters
YOLOv5n	89.2	83.2	88.5	1,763,224
+ CBAM-1	89.6	83.5	**89.1 (+0.6)**	1,774,273 (+11,049)
+ CBAM-2	88.5	82.8	89.0 (+0.5)	1,774,273 (+11,049)
+ CBAM-3	87.7	83.1	88.6 (+0.1)	1,785,322 (+22,098)

4.2.2. Improving the YOLOv5n neck

Following the integration of CBAMs, the Efficient RepGFPN was chosen to improve the YOLOv5 neck to refine and fuse high-level semantic and low-level spatial features. Table 2 shows the performance comparison of three models. The third model incorporating the CBAM and efficient RepGFPN achieved a mAP50 of 89.3%, representing an improvement of 0.2% compared to other models. Compared to the first model with CBAM, the third model reduced the recall by 0.5% to increase the precision by 0.6% with an increase of 557,312 parameters. Overall, the third model with the CBAM and Efficient RepGFPN slightly outperformed the YOLOv5n model with CBAM.

Table 2. Performance comparison of three models

Model	Precision (%)	Recall (%)	mAP50 (%)	Parameters
YOLOv5n + CBAM	89.6	83.5	89.1	1,774,273
YOLOv5n + RepGFPN	90.1	83.0	89.1	2,320,536 (+546,263)
YOLOv5n + CBAM + RepGFPN	90.2	83.0	**89.3 (+0.2)**	2,331,585 (+557,312)

4.2.3. Improving the C3 module

Four ablation experiments were conducted to ascertain the efficacy of the C3_DCNv2 module and to determine the influence of its number on the model performance. Table 3 shows the performance comparison of the YOLOv5n model and the models that improved the C3 module for all classes. C3_DCNv2-1 indicates that the last C3 module within the YOLOv5n backbone is replaced with the C3_DCNv2 module. C3_DCNv2-2 indicates that the later two C3 modules within the YOLOv5n backbone are replaced with the C3_DCNv2 modules. Similarly, C3_DCNv2-3 and C3_DCNv2-4 adhere Table 3.

Table 3. Performance comparison of five models

Model	Precision (%)	Recall (%)	mAP50 (%)	Parameters
YOLOv5n + CBAM + RepGFPN	90.2	83.0	**89.3**	2,331,585
YOLOv5n + CBAM + RepGFPN + C3_DCNv2-1	90.2	82.8	89.2 (-0.1)	2,362,972 (+31,387)
YOLOv5n + CBAM + RepGFPN + C3_DCNv2-2	88.7	83.2	88.8 (-0.5)	2,410,093 (+78,508)
YOLOv5n + CBAM + RepGFPN + C3_DCNv2-3	90.0	83.2	**89.3**	2,425,827 (+94,242)
YOLOv5n + CBAM + RepGFPN + C3_DCNv2-4	90.2	83.0	88.9 (-0.4)	2,429,774 (+98,189)

The experimental results demonstrated that the model with three C3_DCNv2 modules achieved a mAP50 of 89.3% with an increase of 94,242 parameters. The mAP50s of the other methods were found to be inferior. Despite the absence of an improvement in the mAP50 relative to the baseline model, the enhanced model, comprising three C3_DCNv2 modules, exhibited a notable enhancement in the recall for the "person_no_helmet" class. Table 4 shows the validation results of three models for each class.

Table 4. Validation results of three models for each class

Model	YOLOv5n			+ CBAM + RepGFPN			+ CBAM + RepGFPN + C3_DCNv2		
Class	Precision (%)	Recall (%)	mAP50 (%)	Precision (%)	Recall (%)	mAP50 (%)	Precision (%)	Recall (%)	mAP50 (%)
All	89.2	83.2	88.5	90.2 (+1.0)	83.0 (-0.2)	89.3	90.0 (+0.8)	83.2	89.3
Head_with_helmet	92.3	89.5	93.5	92.6	89.5	94.2	93.0	89.5	94.0
Person_with_helmet	89.6	86.4	90.8	91.7 (+1.1)	85.1 (-0.7)	91.1	91.3 (+1.7)	84.6 (-1.8)	91.3
Person_no_helmet	85.7	74.8	81.8	86.3 (+0.6)	74.5 (-0.3)	82.6	85.7	**75.6 (+0.8)**	82.6

Compared to the YOLOv5n model, the second model reduced the recall to a slight degree to increase the precision, resulting in an increase of 0.8% from 88.5% to 89.3% in mAP50. The third method also increased the mAP50 to 89.3% with an increase of 0.8% in precision. For the "head_with_helmet" class, the three models achieved comparable precision, recall, and mAP50. For the "person_with_helmet" class, the second and third models reduced recalls to increase precisions compared to the YOLOv5n model. For the "person_no_helmet" class, compared to the YOLOv5n model, the second model also reduced the recall to increase the precision, while the third model increased the recall by 0.8% from 74.8% to 75.6%. Overall, the enhanced model incorporating three C3_DCNv2 modules demonstrated a reduction in recall for the "person_with_helmet" class and an increase in recall for the "person_no_helmet" class.

4.2.4. Introducing the focal-EIoU

The focal-EIoU loss function replaced the CIoU loss function of YOLOv5n to split the aspect ratio penalty into the width and height penalties. Table 5 shows the performance comparison of the models using different loss functions for all classes. As the replacement of the loss function did not result in a modification of the network structure, the parameters remained unaltered. The EIoU model had the lowest mAP50 of 88.7%. The reweighted CIoU, focal-CIoU, did not outperform compared to the CIoU model, with a mAP50 of 89.3%. The focal-EIoU model reduced the precision by 0.8% (from 90% to 89.2%) to increase the recall by 0.8% (from 83.2% to 84%), finally achieving the highest mAP50 of 89.5%. Table 6 shows the validation results of three models for each class.

Table 5. Performance comparison of four models

Model	Precision (%)	Recall (%)	mAP50 (%)	Parameters
YOLOv5n + CBAM + RepGFPN + C3_DCNv2 (CIoU)	90.0	83.2	89.3	2,425,827
YOLOv5n + CBAM + RepGFPN + C3_DCNv2 (EIoU)	88.5	83.2	88.7(-0.6)	2,425,827
YOLOv5n + CBAM + RepGFPN + C3_DCNv2 (Focal-CIoU)	89.0	83.3	89.3	2,425,827
YOLOv5n + CBAM + RepGFPN + C3_DCNv2 (Focal-EIoU)	89.2	**84.0**	**89.5(+0.2)**	2,425,827

Table 6. Validation results of two models for each class

Model	Precision (%) (head_with_helmet)	Recall (%) (head_with_helmet)	Precision (%) (person_with_helmet)	Recall (%) (person_with_helmet)	Precision (%) (person_no_helmet)	Recall (%) (person_no_helmet)
YOLOv5n	92.3	89.5	89.6	86.4	85.7	74.8
CIoU	93.0 (+0.7)	89.5	91.3 (+1.7)	84.6 (-1.8)	85.7	75.6 (+0.8)
Focal-EIoU	91.2 (-1.1)	89.7 (+0.2)	90.2 (+0.6)	85.9 (-0.5)	86.1 (+0.4)	**76.5 (+1.7)**

Compared to the YOLOv5n model, for the "head_with_helmet" class, the CIoU model achieved the highest precision of 93%, increasing by 0.7%, while the focal-EIoU model reduced the precision by 1.1% (from 92.3% to 91.2%) to increase the recall by 0.2% slightly. For the "person_with_helmet" class, the CIoU and focal-EIoU models reduced the recall to improve the precision. For the "person_no_helmet" class, the CIoU model increased the recall by 0.8%. In particular, the focal-EIoU model increased the precision and recall by 0.4% and 1.7%, respectively. Overall, the focal-EIoU model achieved the highest recall of 76.5% and precision of 86.1% for the "person_no_helmet" class.

4.2.5. Summary

The YOLOv5n model was improved in four areas to develop the CGDE-YOLOv5n model. Table 7 presents a summary of the comparative performance of each model. The CBAM and efficient RepGFPN enhanced the accuracy of the original YOLOv5n model. The C3_DCNv2 exhibited an enhanced recall rate for the "person_no_helmet" class. The focal-EIoU further improved accuracy and recall. Finally, the CGDE-YOLOv5n model achieved the highest mAP50 of 89.5% and recall of 84% compared to other models. Compared to the YOLOv5n model, the mAP50 and recall increased by 1% and 0.8% respectively. In particular, the recall of the "person_no_helmet" class increased by 1.7%, reaching 76.5%. Figure 9 shows the experimental curves of five models.

Table 7. Performance comparison of five models

Model	Precision (%) (all classes)	Recall (%) (all classes)	mAP50 (%) (all classes)	Parameters	Recall (%) (person_no_helmet)
YOLOv5n (baseline)	89.2	83.2	88.5	1,763,224	74.8%
+ CBAM	89.6	83.5	89.1 (+0.6)	1,774,273 (+11,049)	74.9%
+ CBAM + RepGFPN	90.2	83.0	89.3 (+0.8)	2,331,585 (+568,361)	74.5 (-0.3)
+ CBAM + RepGFPN + C3_DCNv2	90.0	83.2	89.3 (+0.8)	2,425,827 (+662,603)	75.6 (+0.8)
+ CBAM + RepGFPN + C3_DCNv2 + Focal-EIoU (**CGDE-YOLOv5n)**	89.2	**84.0**	**89.5 (+1.0)**	2,425,827 (+662,603)	**76.5 (+1.7)**

Figure 9(a) displays the precision-recall curves of five models. The value of mAP50 is the area enclosed by the precision-recall curve and the two coordinate axes. The CGDE-YOLOv5n model with the purple precision-recall curve had the largest area. Figure 9(b) shows the mAP50 curves of five models in 200 training epochs. Before the 75th epoch, the mAP50 of the YOLOv5n model increased faster than other models. After that, the CGDE-YOLOv5n model outperformed other models in most epochs, achieving the highest mAP50 of 89.5%.

Figure 9(c) shows the precision curves of five models in 200 training epochs. In the early stage of training, the YOLOv5n model achieved better precision. In the middle of training, the CGDE-YOLOv5n model performed better. In the later stage of training, the YOLOv5n model with CBAM and efficient RepGFPN outperformed other models, achieving the highest precision of 90.2%. Figure 9(d) shows the recall curves of five models in 200 training epochs. Before the 100th epoch, the YOLOv5n model achieved better recall. After that, the CGDE-YOLOv5n model and the YOLOv5n model with CBAM, efficient RepGFPN, and DCNv2 outperformed other models.

4.3. Comparative experiments with other common models

To facilitate a comparison with the CGDE-YOLOv5n model, several lightweight backbones, including MobileNetv3, ShuffleNetv2, and GhostNetv2, were replaced with the YOLOv5n backbone. Furthermore, the YOLOv5s model with numerous parameters was employed to compare performance. Table 8 compares the performance of five models.

Compared to three lightweight backbones, the CGDE-YOLOv5n model exhibited superior performance, as evidenced by higher precision, recall, and mAP50. In particular, the CGDE-YOLOv5n model achieved the same mAP50 of 89.5% compared to the YOLOv5s model. However, the parameters of the CGDE-YOLOv5n model were significantly reduced. Despite the YOLOv5s model demonstrating the highest precision of 91.3%, the CGDE-YOLOv5n model exhibited the highest recall of 84%.

4.4. Deployment on the Jetson Orin Nano

The trained model was unsuitable for direct deployment on embedded devices. TensorRT was employed to accelerate the model's inference speed to address this issue. Additionally, the data precision of the inference model was reduced from 32-bit floating point numbers (FP32) to FP16, further enhancing

inference speed. Generally, the model's detection speed is measured in FPS. Accordingly, the validation dataset comprising 500 images was employed to evaluate the model's average detection speed. Table 9 compares the detection speed between the YOLOv5n and CGDE-YOLOv5n models.

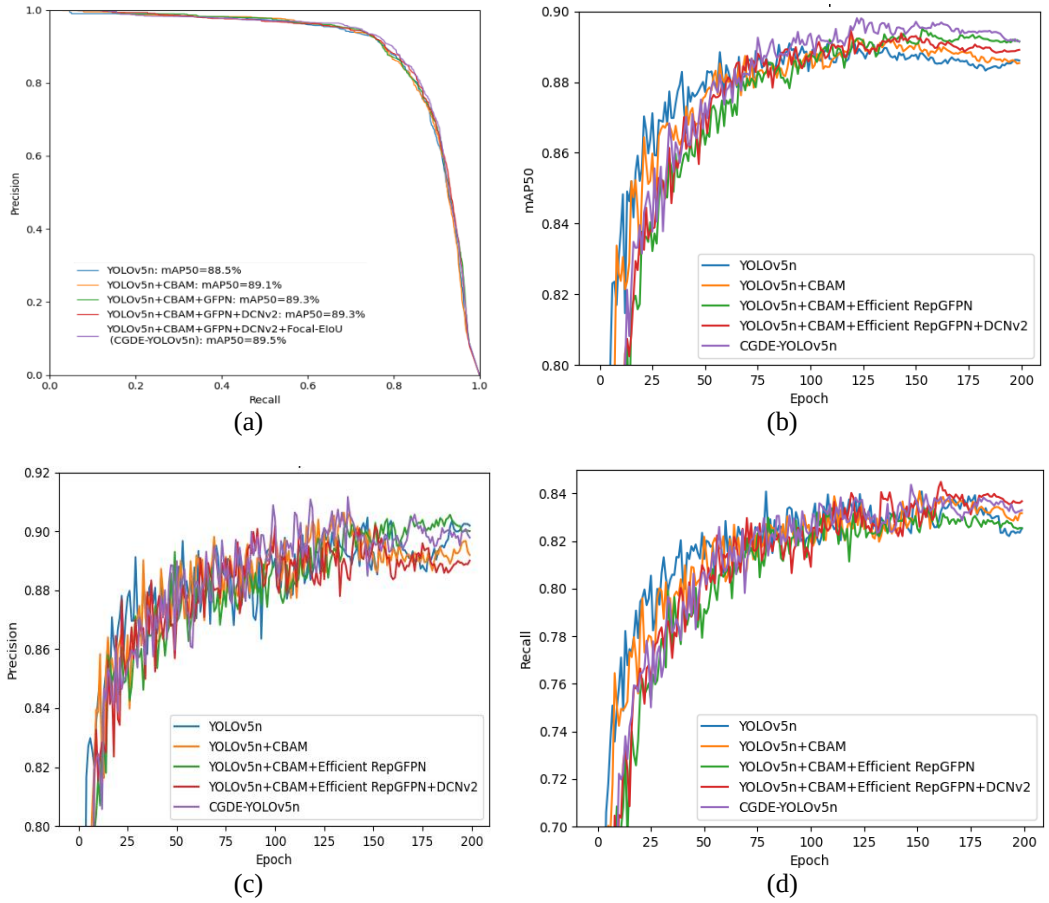


Figure 9. Experimental curves of five models in 200 epochs: (a) precision-recall curves, (b) mAP50 curves, (c) precision curves, and (d) recall curves

Table 8. Performance comparison of five models

Model	Precision (%)	Recall (%)	mAP50 (%)	Parameters
YOLOv5n-MobileNetv3	87.8	79.5	85.8	2,095,048
YOLOv5n-ShuffleNetv2	87.8	82.1	87.2	1,858,260
YOLOv5n-GhostNetv2	88.7	81.4	88.0	2,057,916
YOLOv5s	**91.3**	83.7	**89.5**	7,018,216
CGDE-YOLOv5n	89.2	**84.0**	**89.5**	2,425,827

Table 9. Detection speed comparison of two models on the Jetson Orin Nano

Model	Pre-process (millisecond)	Inference (millisecond)	NMS (millisecond)	FPS
YOLOv5n	1.4	8.1	4.3	72.5
CGDE-YOLOv5n	1.4	9.4	3.8	68.5

The model's detection time is divided into three phases: preprocessing, inference, and NMS. FPS is the inverse of the detection time. The YOLOv5n model achieved a faster FPS of 72.5 compared to the CGDE-YOLOv5n model with 68.5 FPS. Figure 10 shows the detection results of the CGDE-YOLOv5n model for sample images containing single or multiple workers.

Where the bounding boxes with red, pink, and orange colors denote the detection results of the "head_with_helmet" class, the "person_with_helmet" class, and the "person_no_helmet" class. Figures 10(a) to 10(d) show the detection results of workers wearing safety helmets. All "head_with_helmet"

and “person_with_helmet” objects in sample images were correctly detected. Figures 10(e) and 10(f) show the detection results of multiple workers wearing safety helmets. All “head_with_helmet” and “person_with_helmet” objects were correctly detected. It is noteworthy that several occluded “person_with_helmet” objects were correctly detected. Figures 10(g) and 10(h) show the detection results of multiple workers not wearing safety helmets. All “person_no_helmet” objects were correctly detected.

Although both models satisfied the requisite for real-time detection, the CGDE-YOLOv5n model achieved superior performance in terms of accuracy and recall compared to the YOLOv5n model, as evidenced in practical applications. Table 10 shows a cost-benefit comparison analysis of five terminals. Where the data comes from the Amazon website. The Jetson Orin Nano terminal used in this study outperforms other terminals in function, price, and power consumption. Therefore, the CGDE-YOLOv5n algorithm has good potential application value in the industrial field.

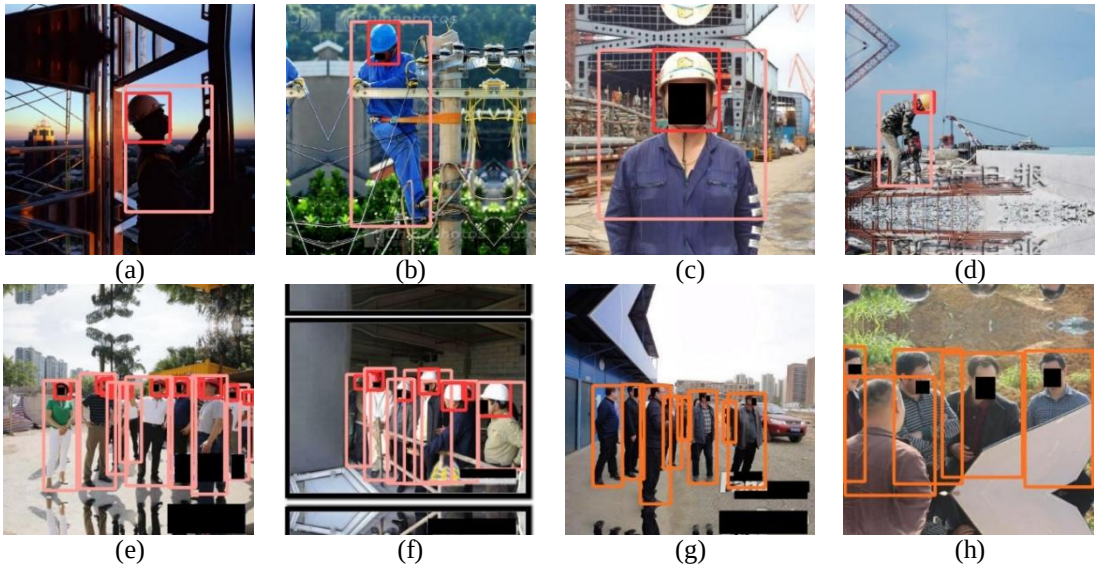


Figure 10. Detection results of the CGDE-YOLOv5n model: (a)–(d) single worker wearing safety helmets, (e)–(f) multiple workers wearing safety helmets, and (g) – (h) multiple workers not wearing safety helmets

Table 10. Cost-benefit comparison of five terminals

Brand	Model	Function	Price (USD)	Power consumption
YAHBOOM	Jetson Orin Nano 4G camera kit	Robotics, edge computing and AI, and vision AI	280	7-10 watts
HIKVISION	DS-2DE4425IW-DE(T5)	Intrusion, line crossing, area entry and exit detections	415	25-30 watts
318NETECH	iDS-2CD7A46G0/P-IZHS	License plate recognition based on deep learning	690	11 watts
AMCREST	4MP AI PTZ POE IP	Human and vehicle detection, face detection, tripwire, intrusion, abandoned object, missing object	560	6.5-12 watts
REKOR	Edge Pro	License plate recognition	999	26 watts

4.5. Discussion

It is acknowledged that deploying the detection model on embedded devices with limited resources is a challenge. Previous models achieved high accuracy but suffered from slow detection speed. Therefore, YOLOv5n is a suitable choice due to its fewer parameters and calculations. However, the lightweight YOLOv5n model has shortcomings in feature extraction and recall improvement. Consequently, the YOLOv5n model was improved through four approaches to obtain the CGDE-YOLOv5n model.

First, it was found that CBAM can adjust the intermediate feature map’s channel and pixel weights to improve CNN performance. After integrating CBAMs, accuracy increased by 0.6%. Second, Efficient RepGFPN could better refine and fuse high-level semantic and low-level spatial features to enhance accuracy. After improving the neck, accuracy increased by 0.2%. Third, it was also found that object shape and size in the helmet dataset varied greatly, DCNv2 was used to enhance feature extraction for deformable

and occluded objects. Although accuracy remained unchanged after improving the C3 module, recall for the "person_no_helmet" class increased by 0.8%. Fourth, focal-EIoU split the aspect ratio penalty into width and height penalties to decrease localization precision and improve recall. After replacing CIoU with focal-EIoU, accuracy increased by 0.2%. Furthermore, recall increased by 0.8%, and recall of the "person_no_helmet" class increased by 1.7%.

With 89.5% accuracy and 68.5 FPS, the CGDE-YOLOv5n model achieved an optimized balance between accuracy and speed. Compared with previous studies, it achieved the same high level of accuracy but with significantly improved detection speed. Therefore, experimental findings will serve as a reference point for researchers seeking to enhance the precision and recall of detection models. Furthermore, they also support the hypotheses of this study.

Experimental findings also indicate that the recall of the "person_no_helmet" class is relatively low. The possible reasons are the insufficient number of images in this category and the insufficient feature extraction capability of the algorithm for this category. Therefore, future research will conduct experiments from two perspectives: dataset diversity and algorithm optimization. Finally, a real-time safety helmet-wearing compliance detection system will be developed for real-world construction sites.

In conclusion, improvement methods integrating CBAM and efficient RepGFPN are beneficial for improving accuracy in this paper. Furthermore, improvement methods using DCNv2 and focal-EIoU facilitate an enhancement in the recall and a reduction in the missed detection rate. Furthermore, the YOLOv5n-based improved model fully meets the requirements of real-time detection.

5. CONCLUSION

This paper has addressed the challenge of deploying a real-time safety helmet-wearing compliance detection model on embedded devices. The lightweight algorithm, YOLOv5n, was selected to minimize the model's parameters and calculations. To address the problem of decreased accuracy in the lightweight network, CBAMs were incorporated into the YOLOv5n backbone. Additionally, Efficient RepGFPN was employed to enhance the YOLOv5n neck to improve accuracy further. Furthermore, DCNv2 was used to improve the C3 module of YOLOv5n to reduce the missed detection rate. Then, focal-EIoU replaced CIoU to increase recall further. Finally, CGDE-YOLOv5n increased accuracy by 1.0% and recall by 0.8% compared to YOLOv5n. Furthermore, it achieved the same accuracy of 89.5% as YOLOv5s with significantly fewer parameters and computational overhead. When deployed on the Jetson Orin Nano, the CGDE-YOLOv5n model achieved a real-time detection speed of 68.5 FPS. The experimental results demonstrated that the CGDE-YOLOv5n algorithm is effective in real-time detection of safety helmet-wearing compliance.

ACKNOWLEDGEMENTS

The authors thank to the College of Engineering and Universiti Teknologi MARA(UiTM) for all their help and support in this research.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Wanbo Luo	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Ahmad Ihsan Mohd Yassin	✓	✓	✓	✓	✓					✓		✓	✓	
Khairul Khaizi Mohd Shariff	✓	✓			✓					✓		✓		
Rajeswari Raju	✓	✓			✓					✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in [CGDE-YOLOv5n] at https://github.com/bobo504/CGDE-YOLOv5n/tree/master.

REFERENCES

[1] M. I. B. Ahmed *et al.*, "Personal protective equipment detection: a deep-learning-based sustainable approach," *Sustainability*, vol. 15, no. 18, p. 13990, Sep. 2023, doi: 10.3390/su151813990.

[2] A. Mansoor, S. Liu, G. M. Ali, A. Bouferguene, and M. Al-Hussein, "Scientometric analysis and critical review on the application of deep learning in the construction industry," *Canadian Journal of Civil Engineering*, vol. 50, no. 4, pp. 253–269, Apr. 2023, doi: 10.1139/cjce-2022-0379.

[3] Q. Li, K. Yu, H. Wang, Q. Guan, S. Gao, and J. Jiang, "Detection algorithm for safety helmet wearing of chemical plant personnel based on improved YOLOv5m," in *2023 International Conference on the Cognitive Computing and Complex Data (ICCD)*, Oct. 2023, pp. 63–67, doi: 10.1109/ICCD59681.2023.10420558.

[4] X. Liang, T. Jiang, Q. Fu, and Q. Wang, "Multi-object detection and classification in construction sites based on YOLOv5," in *Proceedings of the 2023 5th International Conference on Video, Signal and Image Processing*, Nov. 2023, pp. 79–85, doi: 10.1145/3638682.3638694.

[5] H. Guo *et al.*, "Enhancing helmet and cigarette detection in electricity power construction based on YOLOv5s-I algorithm," in *Fourth International Conference on Machine Learning and Computer Application (ICMLCA 2023)*, May 2024, p. 115, doi: 10.1117/12.3029269.

[6] Q. Liu and F. Han, "Research on an improved YOLOv5s algorithm for detecting helmets on construction sites," in *Proceedings of the 2023 7th International Conference on Electronic Information Technology and Computer Engineering*, Oct. 2023, pp. 104–110, doi: 10.1145/3650400.3650418.

[7] Q. Ren, H. Zhu, C. Chen, H. Lan, and R. Luo, "Safety helmet wearing detection based on improved YOLOv5s," in *Proceedings of the 2023 6th International Conference on Image and Graphics Processing*, Jan. 2023, pp. 148–154, doi: 10.1145/3582649.3582654.

[8] F. Syah, H. H. Marfuah, and A. Wahana, "Detection of safety helmet using principal component analyst (PCA) method," in *AIP Conference Proceedings*, 2023, vol. 2491, p. 040017, doi: 10.1063/5.0116216.

[9] J. Tan, H. Wang, X. Li, and X. Zhang, "Improved object detection algorithm for workshop environments based on YOLOv5," in *2023 9th International Conference on Computer and Communications (ICCC)*, Dec. 2023, pp. 1714–1718, doi: 10.1109/ICCC59590.2023.10507328.

[10] G. Wu, J. Chen, W. Kang, J. Chi, H. Ma, and W. Song, "AI recognition method for multiple time series abnormal behavior images in electricity safety scenarios," in *Fourth International Conference on Machine Learning and Computer Application (ICMLCA 2023)*, May 2024, p. 118, doi: 10.1117/12.3029283.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[12] W. Liu *et al.*, "SSD: single shot multibox detector," in *Computer Vision--ECCV 2016: 14th European Conference*, 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.

[13] H. Feng and J. Hu, "Helmet wearing detection using improved single shot multibox detector," in *2022 5th World Conference on Mechanical Engineering and Intelligent Manufacturing (WCMEIM)*, Nov. 2022, pp. 444–447, doi: 10.1109/WCMEIM56910.2022.10021515.

[14] H. Zhan and X. Pei, "Based on improved single shot multibox detector construction site helmet detection algorithm," in *2023 5th International Conference on Communications, Information System and Computer Engineering (CISCE)*, Apr. 2023, pp. 425–428, doi: 10.1109/CISCE58541.2023.10142555.

[15] R. M. Calle Quispe, M. Aghaei Gavari, and E. Aguilar Torres, "Towards accurate real-time safety helmet detection through a deep learning-based method," *Ingeniare. Revista chilena de ingeniería*, vol. 31, Jul. 2023, doi: 10.4067/S0718-33052023000100212.

[16] W. Chen, M. Liu, X. Zhou, J. Pan, and H. Tan, "Safety helmet wearing detection in aerial images using improved YOLOv4," *Computers, Materials & Continua*, vol. 72, no. 2, pp. 3159–3174, 2022, doi: 10.32604/cmc.2022.026664.

[17] Z. Huang, Y. Zhang, Y. Zhang, and K. Ren, "Indoor safety helmet-wearing detection algorithm based on improved YOLOv4," *Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/Journal of Tianjin University Science and Technology*, vol. 56, no. 1, pp. 64–72, 2023, doi: 10.11784/tdxbz202111026.

[18] X. Guobo, T. Jingjing, L. Zhiyi, Z. Xiaofeng, and F. Ming, "Improved YOLOv4 helmet detection algorithm under complex scenarios," *Laser & Optoelectronics Progress*, vol. 60, no. 12, p. 1210011, 2023, doi: 10.3788/LOP221388.

[19] H. Li, D. Wu, W. Zhang, and C. Xiao, "YOLO-PL: helmet wearing detection algorithm based on improved YOLOv4," *Digital Signal Processing*, vol. 144, p. 104283, Jan. 2024, doi: 10.1016/j.dsp.2023.104283.

[20] H.-P. Wan, W.-J. Zhang, H.-B. Ge, Y. Luo, and M. D. Todd, "Improved vision-based method for detection of unauthorized intrusion by construction sites workers," *Journal of Construction Engineering and Management*, vol. 149, no. 7, Jul. 2023, doi: 10.1061/JCEMD4.COENG-13294.

[21] Q. An, Y. Xu, J. Yu, M. Tang, T. Liu, and F. Xu, "Research on safety helmet detection algorithm based on improved YOLOv5s," *Sensors*, vol. 23, no. 13, p. 5824, Jun. 2023, doi: 10.3390/s23135824.

[22] Z. Deng, C. Yao, and Q. Yin, "Safety helmet wearing detection based on Jetson Nano and improved YOLOv5," *Advances in Civil Engineering*, vol. 2023, pp. 1–12, May 2023, doi: 10.1155/2023/1959962.

[23] G. Dong, Y. Zhang, W. Xie, and Y. Huang, "A safety helmet-wearing detection method based on cross-layer connection," *Journal of Real-Time Image Processing*, vol. 21, no. 3, p. 72, Jun. 2024, doi: 10.1007/s11554-024-01437-5.

[24] G. Hou, Q. Chen, Z. Yang, Y. Zhang, D. Zhang, and H. Li, "Helmet detection method based on improved YOLOv5," *Gongcheng Kexue Xuebao/Chinese Journal of Engineering*, vol. 46, no. 2, pp. 329–342, 2024, doi: 10.13374/j.issn2095-9389.2022.12.07.002.

[25] O. Iparraguirre-Villanueva, J. Gonzales-Huaman, J. Machuca-Solano, and J. Ruiz-Alvarado, "Improving industrial security device detection with convolutional neural networks," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 34, no. 3, pp. 1935–1743, Jun. 2024, doi: 10.11591/ijeecs.v34.i3.pp1935-1943.

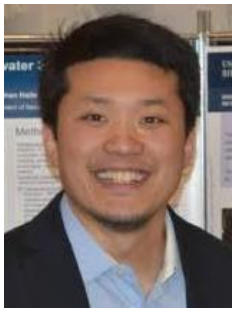
[26] Kisaezehra, M. U. Farooq, M. A. Bhutto, and A. K. Kazi, "Real-time safety helmet detection using YOLOv5 at construction sites," *Intelligent Automation & Soft Computing*, vol. 36, no. 1, pp. 911–927, 2023, doi: 10.32604/iasc.2023.031359.

[27] N. Kwak and D. Kim, "Detection of worker's safety helmet and mask and identification of worker using deeplearning," *Computers, Materials & Continua*, vol. 75, no. 1, pp. 1671–1686, 2023, doi: 10.32604/cmc.2023.035762.
[28] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.
[29] Y. Jiang, Z. Tan, J. Wang, X. Sun, M. Lin, and H. Li, "Giraffedet: a heavy-neck paradigm for object detection," *arXiv preprint arXiv:2202.04256*, 2022, doi: 10.48550/arXiv.2202.04256.
[30] X. Xu, Y. Jiang, W. Chen, Y. Huang, Y. Zhang, and X. Sun, "DAMO-YOLO : a report on real-time object detection design," *arXiv preprint 2211.15444*, Nov. 2022, doi: 10.48550/arXiv.2211.15444.
[31] J. Dai *et al.*, "Deformable convolutional networks," *arXiv preprint 1703.06211*, Mar. 2017, doi: 10.48550/arXiv.1703.06211.
[32] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable ConvNets v2: more deformable, better results," *arXiv preprint 1811.11168*, Nov. 2018, doi: 10.48550/arXiv.1811.11168.
[33] M.-E. Otgonbold *et al.*, "SHEL5K: an extended dataset and benchmarking for safety helmet detection," *Sensors*, vol. 22, no. 6, p. 2315, Mar. 2022, doi: 10.3390/s22062315.

BIOGRAPHIES OF AUTHORS

Wanbo Luo received his M.S. degree from the University of Chongqing University of Posts and Telecommunications, China, in 2009. He is currently an associate professor at Leshan Vocational and Technical College. He is currently doing a Ph.D. degree at Universiti Teknologi MARA. His research interests are mainly in deep learning and object detection. He has authored or coauthored more than 12 publications, with 4 H-index and more than 41 citations. He can be contacted at email: boboluo504@gmail.com.

Ahmad Ihsan Mohd Yassin received his M.S. degree in Electrical Engineering from the Universiti Teknologi MARA, Malaysia, and his Ph.D. degree in Electrical Engineering from the Universiti Teknologi MARA, Malaysia, respectively. He is currently an associate professor at the Microwave Research Institute, Universiti Teknologi MARA, Malaysia. His research interests are mainly in system identification, optimization, and artificial intelligence. He has supervised and co-supervised more than 10 masters and 23 Ph.D. students. He has authored or coauthored more than 156 publications, with 24 H-index and more than 2560 citations. He can be contacted at email: ihsan_yassin@uitm.edu.my.



Khairul Khaizi Mohd Shariff received his M.S. degree in communications and computers from the Universiti Kebangsaan Malaysia, Malaysia, and his Ph.D. degree from the University of Birmingham, UK, respectively. He is currently working as a senior lecturer at Universiti Teknologi MARA, Malaysia, and serves as a researcher at the Microwave Research Institute at the same university. His research interests are mainly in RF and wireless systems. He has supervised and co-supervised more than 3 masters and 1 Ph.D. student. He has authored or coauthored more than 44 publications, with 7 H-index and more than 193 citations. He can be contacted at email: khairulkhaizi@uitm.edu.my.

Rajeswari Raju received her M.S. degree in computer science from the University of Malaya, Malaysia, and her Ph.D. degree from the University of Nottingham Malaysia Campus, Malaysia, respectively. She is currently a senior lecturer at the School of Computing and Mathematics, Universiti Teknologi MARA, Malaysia. Her research interests are mainly in image processing, optimization, computer security, and networking. She has supervised and co-supervised more than 2 masters and 1 Ph.D. student. She has authored or coauthored more than 53 publications, with 5 H-index and more than 100 citations. She can be contacted at email: rajes332@uitm.edu.my.