CIT. Journal of Computing and Information Technology, Vol. 31, No. 2, June 2023, 123–136
doi: 10.20532/cit.2023.1005754

Research on Small Target Detection Method for Industrial Safety Helmets Based on Improved YOLOv8

Lan Yuanshuai[1], Chen Mo[1], Li Chuan[2], Wang Qian[1] and Liao Min[1]

[1]School of Electronic Information Engineering, Geely University, Chengdu, Sichuan, China
[2]Chengdu College, University of Electronic Science and Technology of China, Chengdu, Sichuan, China

Industrial safety helmets are crucial personal protective gear but detecting them as small targets in complex environments is challenging. This work proposes enhancements to the YOLOv8 object detection framework, specifically incorporating a spatial-to-depth (SPD) convolution module and a large selective kernel network (LSKNet). SPD-Conv combines spatial-to-depth layers and non-strided convolutions to retain fine-grained information when downscaling feature maps, while LSKNet introduces dynamic spatial selection and attention for refined context modeling. Our customized model is trained on a dataset of construction hardhat images captured via drones. Quantitative results showcase higher precision and recall over baseline YOLOv8, surpassing competing YOLOv5 versions. An optimized final model outcomes demonstrate accuracy exceeding 90% validation in mAP metric after 200 training rounds. By tackling limitations posed by small, obscured industrial safety gears, this enhanced real-time detection approach provides indispensable technological support for bolstering workplace hazard identification and prevention.

ACM CCS (2012) Classification: Computing methodologies → Artificial intelligence → Computer vision → Computer vision problems → Object detection

Keywords: Small Target Detection, Target Detection, safety helmet, YOLOv8

1. Introduction

In recent years, the popularization of industrial safety helmets has become a key issue in the field of industrial safety. As an important personal protective equipment, industrial safety helmets can significantly reduce the risk of head injuries in industrial accidents and are crucial to safeguarding workers' lives, health, and safety. However, detecting small targets in complex industrial environments has always been a challenging task, especially when those small targets are the safety helmets worn by workers on their heads. Therefore, this study aims to explore methods for detecting small objects – industrial safety helmets to improve industrial safety and production efficiency.

With the rapid development of computer vision and deep learning technology, significant progress has been made in the field of target detection. YOLO (You Only Look Once) algorithm, as an end-to-end real-time target detection algorithm, has attracted widespread attention. Among them, YOLOv8, as its latest version, has achieved an ideal balance between target detection accuracy and speed, making it a potential application in the industrial sector.

Shan *et al.* (2023) [1] studied the method of enhancing the YOLOv5 target detection algorithm to improve the real-time alert ability for individuals not wearing safety helmets. By adopting ECA (Enhanced Channel Attention) mechanism [2] and introducing a weighted bidirectional feature pyramid network structure (BiFPN), the decoupling head in YOLOv5 improves detection performance and convergence speed. The experimental results show that the enhanced YOLOv5 model achieves an average accuracy of 95.9% on the custom helmet dataset, which is 3.0 percentage points higher than the original model.

Chen *et al.* (2023) [3] proposed an improved convolutional neural network model called YOLOv7 WFD for detecting unhelmeted workers in the construction industry. The improvements include adding a new module DBS to enhance target feature extraction ability, introducing CARAFE (channel-wise attention refinement module) to improve detail reconstruction ability during image upsampling, and adopting a dynamic focus mechanism with a Wise IoU loss function to improve generalization ability and detection accuracy. The experimental results show that when tested on the SHEL5K dataset, the improved YOLOv7 WFD achieves 92.6% mAP and 79.3 FPS.

Deng *et al.* (2022) [4] proposed a lightweight target detection algorithm called ML-YOLOv3, which addresses the problems of complex networks, high FLOPs, and large parameter sizes in YOLOv3. The improvements include integrating CSPNet and GhostNet to design an efficient residual network CSP-Ghost-Resnet, designing a new backbone network ML Darknet to implement gradient shunting, and designing a lightweight multi-scale feature extraction network using PAN CSP network. The experimental results showed that compared with YOLOv3, ML-YOLOv3 has only 29.7% and 29.4% more FLOPs and parameters, respectively, and is significantly better than YOLOv5 in terms of computational cost and detection effect.

Zhou *et al.* (2021) [5] used a YOLOv5-based safety helmet detection method. By annotating 6045 datasets and training and testing different parameter versions of YOLOv5 models, they obtained experimental results. Among them, the average detection speed of YOLOv5s reached 110 FPS, fully meeting the requirements of real-time detection. The mAP of YOLOv5x, using trainable target detection engines, reached 94.7%, proving the effectiveness of helmet detection based on YOLOv5.

These studies provide us with important references, but they still have some limitations, such as false positive problems in complex backgrounds, insufficient adaptability to different postures and obstructions for small targets, *etc.* Therefore, the main goal of this study is to improve the YOLOv8 model and realize small object detection on industrial safety helmets to improve industrial safety and production efficiency. By combining the YOLOv8 algorithm with industrial safety helmet small object detection, we aim to achieve the following goals:

1. improve the detection accuracy of small targets, including accuracy in complex backgrounds;
2. improve detection speed to achieve real-time monitoring and rapid response;
3. improve adaptability to different postures and obstructions for workers' head safety helmets.

By addressing these issues, we aim to achieve rapid and accurate detection of workers wearing safety helmets in industry, ensuring worker safety while providing strong support for industrial safety helmet detection and worker safety in industry.

2. Literature Review

In the field of industrial safety, the development of small object detection technology has always attracted much attention. Especially for the identification and monitoring of helmet wearers in industrial scenes, researchers have proposed various solutions based on different technical means. These methods include statistics-based methods, traditional machine learning methods, and deep learning methods.

In the research method based on statistical analysis, Smith *et al.* proposed a Gaussian model-based industrial safety helmet small object detection method in their 2019 study [6]. By conducting statistical analysis on the pixel grayscale values of different regions in the image, effective detection of the safety helmet target was achieved. However, the recognition accuracy of this method in complex backgrounds has certain limitations.

On the other hand, traditional machine learning-based research methods also have their unique contributions. Wang *et al.* proposed a support vector machine-based industrial safety helmet small object detection method in their 2021 study [7]. By feature extraction and classifier design, accurate detection and classification of small targets were achieved. However, the recognition of this method still faces challenges in illumination changes and occlusion situations.

The most striking aspect is the research methodology based on deep learning, particularly the small object detection method using YOLOv8. In recent years, Zhang *et al.*, in their 2023 proposal addressing the challenge of detecting small objects in drone images [8], have improved upon the YOLOv5 algorithm by incorporating a spatial-to-depth (SPD) convolutional module, adding various attention mechanisms, and refining the multi-scale detection module to enable recognition of small targets in drone imagery.

Qian *et al.*, in their 2023 introduction of a novel safety helmet detection method based on Convolutional Neural Networks [9], optimized the BottleneckCSP structure in the YOLOv5 backbone network to reduce model complexity. They also designed an upsampling feature enhancement module to eliminate information loss and introduced self-attention mechanisms to avoid redundant information resulting from feature fusion. This allows for faster inference speed and better performance under the same computational capacity.

Tan *et al.*, in their 2022 proposal of a deep learning method based on YOLOv8 [10], combined attention mechanisms with multi-scale feature fusion [11], effectively enhancing the precision and robustness of industrial safety helmet small object detection. Furthermore, Chen *et al.*, in their 2023 research, presented a residual network-based approach to industrial safety helmet small object detection [12]. By incorporating the residual network structure, they effectively improved the model's detection performance in complex scenarios.

Although deep learning-based methods have made significant progress, they also face challenges such as complex models, large computation volumes, and high data demand. In addition, problems such as high labeling cost for datasets and low model robustness still need further research and improvement. This paper aims to gain insights from these studies and propose a new YOLOv8 model to solve these problems in small object recognition.

3. Experimental Methods

3.1. Yolov8 Model Introduction

YOLOv8 [13] is a deep learning model that serves as a prevalent real-time object detection algorithm in the field of object detection. As the latest iteration in the YOLO series of models, YOLOv8 builds upon and improves its predecessors with the aim to realize both real-time object detection and localization. YOLOv8 introduces enhancements and optimizations over previous versions to boost detection accuracy and speed, achieving favorable results across a variety of application scenarios.

3.1.1. Model Input and Output

The input of the Yolov8 model is a single image at a time. Typically, the resolution of this image is fixed, and YOLOv8 will divide the image into a fixed grid. The typical input size may be 416x416, 608x608, or 640x640 pixels.

The output of the YOLOv8 model comprises bounding boxes enclosing objects detected within an image, along with their respective categories [14]. Specifically, the output includes class labels for the targets, positions of the bounding boxes (commonly represented by the coordinates of their top-left and bottom-right corners), as well as confidence scores for each bounding box, which indicate the model's certainty about the presence of a target.

3.1.2. Model Architecture

The core structure of the YOLOv8 model consists of a backbone network, a detection head, and a loss function.

Input layer: Typically, it is a predefined-size image with an image size that is usually a multiple of 32.

Backbone network: YOLOv8 typically employs a backbone network, such as Darknet, to extract image features. Specifically, YOLOv8 references the CSPDarkNet-53 network [15]. Unlike YOLOv5, YOLOv8 utilizes C2f (CSPLayer_2Conv) modules instead of C3 modules.

Feature extraction layers: YOLOv8 employs a feature extraction mechanism similar to the PAN-FPN used in YOLOv5, referred to as dual-stream FPN [16]. This mechanism efficiently extracts feature information from the input image through multiple convolutional and pooling operations, resulting in fast and effective detection performance.

Detection layers: This section is responsible for performing object detection based on the features extracted by the backbone network. YOLOv8 utilizes specialized convolutional and connection layers to predict the positions and categories of bounding boxes [17].

Output layer: This layer outputs information regarding the detected objects, including their categories, positions, and confidence scores. After the output, the results must undergo a non-maximum suppression (NMS) algorithm [18] to filter the output. The purpose of NMS is to eliminate overlapping bounding boxes and select the highest confidence scoring bounding box as the final output.

Loss function: The YOLOv8 model uses two aspects of loss functions: positive and negative sample allocation strategies and loss calculation.

3.2. Improvement Methods

3.2.1. SPD-Conv Module

SPD-Conv [19], depicted in Figure 1, consists of an SPDlayer and a convolutional layer with no stride, which can be applied to most CNN architectures, including the YOLOv8 model.

The role of the SPD layer is to reduce each spatial dimension of the input feature map to the channel dimension while preserving information within the channels. This can be achieved by mapping each pixel or feature in the input feature map to a channel [20]. During this process, the size of the spatial dimensions decreases while the size of the channel dimension increases.

The no-stride convolution (cConv) layer is a standard convolution operation performed after the SPD layer. Unlike stride convolutions, which move across the feature map, no-stride convolutions perform convolution operations on each pixel or feature map without shifting [21]. This helps alleviate potential over-downsampling issues that may arise in the SPD layer and retains more fine-grained information.

The combination method of the SPD-Conv module involves sequentially applying the SPD layer and then the Conv layer to the input fea-

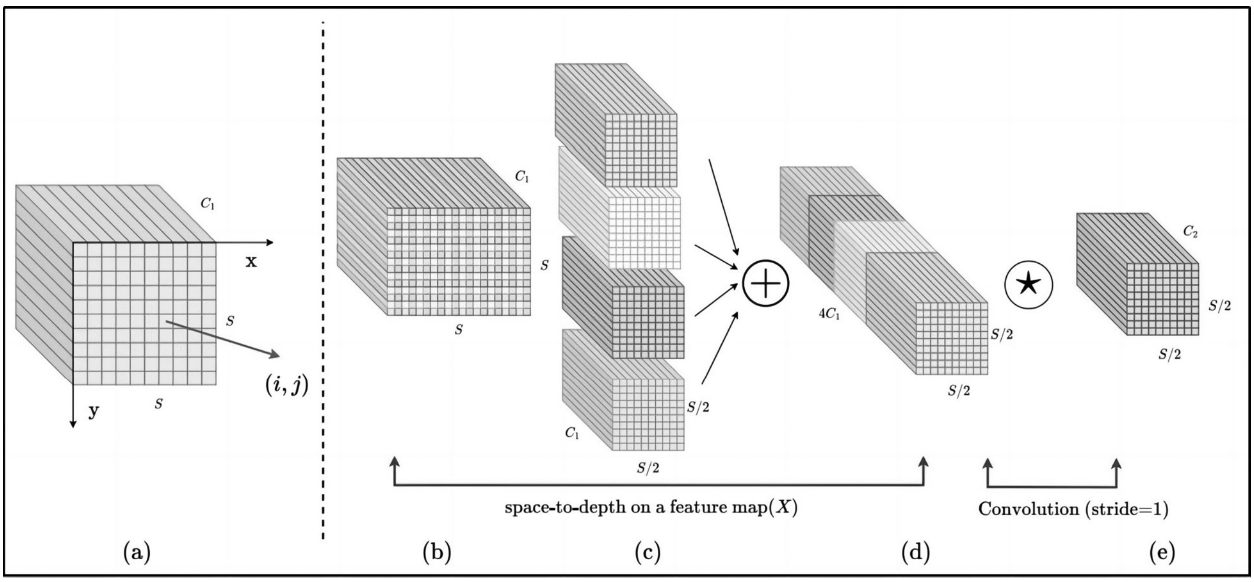

Figure 1. SPD-Conv architecture, scale=2.

ture map. This combination method can reduce the size of the spatial dimensions without losing information while retaining information within the channels, thereby improving the detection performance of CNNs on low-resolution images and small objects. It can also reduce dependence on "good quality" inputs for models.

3.2.2. Data Augmentation

Data augmentation is a crucial technique for small object detection tasks as it improves the robustness and generalization ability of models to small targets. When handling data augmentation for small targets, there are several commonly used methods:

- Random cropping and rescaling: Increasing dataset diversity through random cropping and rescaling is a common data augmentation technique. When dealing with small targets, random cropping and rescaling can help the model better learn contextual information and scale variations of the targets, thereby enhancing detection performance [22].
- Rotation and flip augmentation: By introducing random rotation and flip augmentation, the model can learn the appearance features of safety helmets at different angles and orientations. This is significant for improving the robustness of the model in complex scenarios [23].
- Color jittering: For small targets, color jittering can be introduced to increase data variety, allowing the model to better adapt to target detection under different lighting conditions. This can enhance the model's generalization ability and adaptability [24].

When performing data augmentation, it is important to maintain the integrity of target features to avoid data distortion affecting model performance. Therefore, appropriate strategies for data augmentation should be selected based on specific datasets and tasks, and evaluation and adjustment should be carried out using techniques such as cross-validation.

3.2.3. LSKNet

The design idea of LSKNet [25], shown in Figure 2, is to achieve adaptive receptive fields through a spatial selection mechanism. In traditional convolutional neural networks (CNNs), the size of the convolution kernel is fixed, whereas in LSKNet, the size of the convolution kernel is dynamically determined based on the input. This means that the model can adjust the receptive field for each target as needed, thereby better capturing the features of the target.

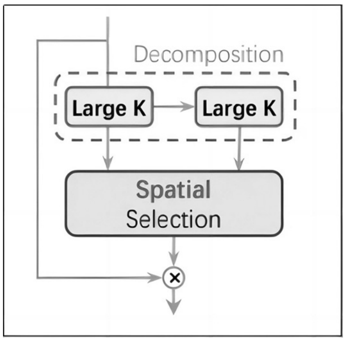

Figure 2. LSKNet network structure.

In addition to rotation-sensitive convolution operations, LSKNet also utilizes attention mechanisms to further enhance the accuracy of object detection [26]. In LSKNet, the attention mechanism is applied to feature representations, enabling the model to better focus on target features and thus improve object detection accuracy. The attention mechanism helps the model better attend to regions relevant to the target and suppress interference from irrelevant regions. By incorporating LSKNet into the YOLOv8 network, it optimizes network performance and improves the detection of small objects.

4. Experimental Verification

4.1. Dataset and Experimental Environment

The training data used in this experiment comes from the Smart Construction project provided by SHWD for headwear detection and human head detection. It includes 7581 images, including 9044 headwear objects (front view) and 111514 normal head objects (not worn or negative). The front objects are from Google or Baidu, and the objects in the dataset are manually labeled with LabelImg. Some negative objects come from SCUT-HEAD. The original SCUT-HEAD errors have been fixed in the dataset, and the data can be loaded directly in the normal Pascal VOC format. Some dataset examples are shown in Figures 3 and 4.

After acquiring the training data, it is necessary to put the data into the model for training. The hardware environment required for model training determines whether the experiment will proceed smoothly. The experimental environment for this paper is shown in Table 1.

Table 1. Configuration of model training environment.

Experimental Environment	**Environment Selection**
Programming Language	Python 3.8
Framework	Pytorch 1.10.0
Operating System	ubuntu20.04
GPU	RTX 3090 (24 GB)
Cuda	11.3
Memory	43 GB

Figure 3. Test set images of Safety Helmets 1.

Figure 4. Test set images of Safety Helmets 2.

4.2. Experimental Analysis

In Chapter 3, the YOLOv8 model was analyzed, and its deficiencies were identified and improved according to the research objectives, resulting in our proposed model, OurModel. As described in section 1, previous studies used the YOLOv5 model for small object detection, which also greatly improved performance, with good accuracy and precision. Therefore, in this experiment, YOLOv5 models, YOLOv8 models, and OurModel models will be used to verify the feasibility of the proposed improvement method, respectively Model 1, Model 2, and Model 3. Below, these three network structures will be tested through ablation and comparison experiments to conclude that the introduced modules can improve the accuracy of the model.

For data augmentation, there is no consideration of performance metrics.

4.2.1. Ablation Experiments

The purpose of ablation experiments is to understand and evaluate the effectiveness of a model's network structure by dissecting its internal principles and components.

Based on the two proposed network structures, experiments will be conducted using the precision and recall plots from the training set, as shown in Figures 5 and 6.

According to Figure 5, the precision of the original YOLOv8 model is gradually increasing, and the learning effect of the model is also gradually improving [27]. After introducing the SPD-Conv and LSKNet network modules, respectively, the accuracy of the new model is higher than that of the Yolov8 model in the early stage. In Figure 5, until around the 60th epoch, the precision of the newly introduced module is higher than that of the original YOLOv8 model. As the number of rounds increases, it starts to decline slowly and becomes lower than the YOLOv8 model. However, the model with the LSKNet module has a precision similar to that of the original YOLOv8 model and is even higher than the original model at some points.

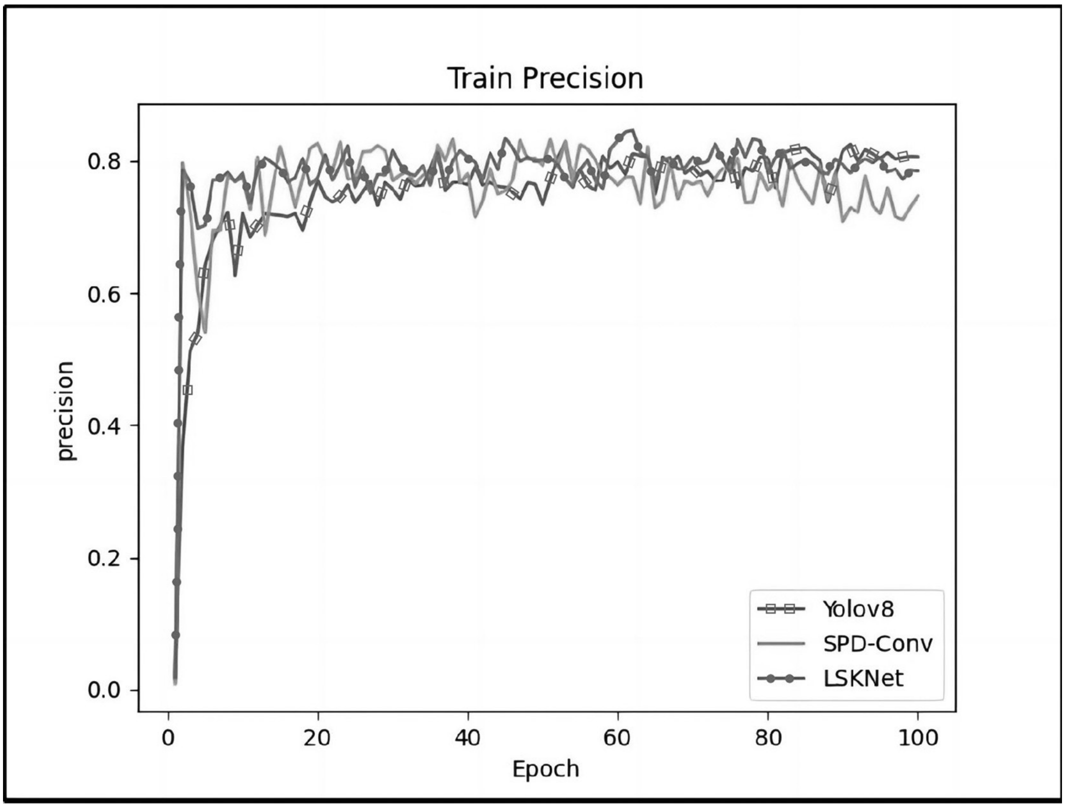

Figure 5. Precision graph of each model.

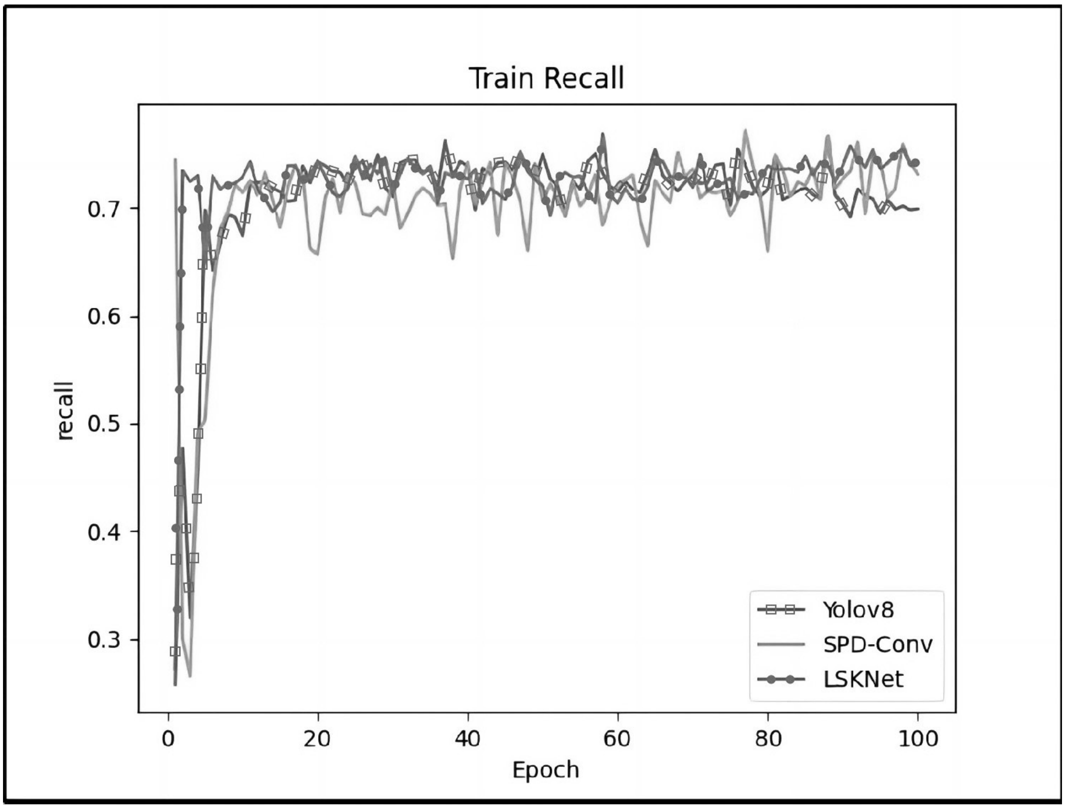

Figure 6. Recall graph of each model.

In Figure 6, the recall rate of the model with the SPD-Conv module is unstable in the early stage, and its performance is not very good. However, as the number of rounds increases, it begins to exceed the YOLOv8 model after the 60th round. The recall rate of the model with the LSKNet module is stable throughout the training process.

Upon analyzing the aforementioned issue, it can be concluded that although the precision of the SPD-Conv and LSKNet models starts to decrease below the YOLOv8 model in the later stages depicted in Figure 5, their precision remains higher than that of the YOLOv8 model around the 80th round. Furthermore, in terms of recall rate at the 80th round in Figure 6, both the SPD-Conv and LSKNet models demonstrate good performance with an improvement over the YOLOv8 model. LSKNet effectively addresses these issues through its rotation-sensitive convolution operations and attention mechanism, enhancing the accuracy of object detection [28]. Therefore, introducing these two modules into the YOLOv8 model is indeed effective.

4.2.2. Comparison Experiments

The purpose of comparison experiments is to obtain conclusions about which model is better, through comparing different models' performance and selecting the best one.

Using the same dataset and performing the same dataset division as Model 1, Model 2, and Model 3 under all parameters being consistent, training results with precision and mAP50 will be plotted, as shown in Figures 7 and 8.

Through the comparison of the two graphs, it can be seen that the OurModel model significantly outperforms the YOLOv5 model. In Figures 7 and 8, the OurModel model consistently demonstrates superiority over the YOLOv5 model in terms of precision and mAP50, and to some extent, it can be compared with the YOLOv8 model. However, as the number of rounds increases, the advantages of the OurModel model become more evident. Its precision starts to improve and remains higher than that of the YOLOv8 model. The mAP50 is also a highlight, consistently staying above both the YOLOv8 and YOLOv5 models.

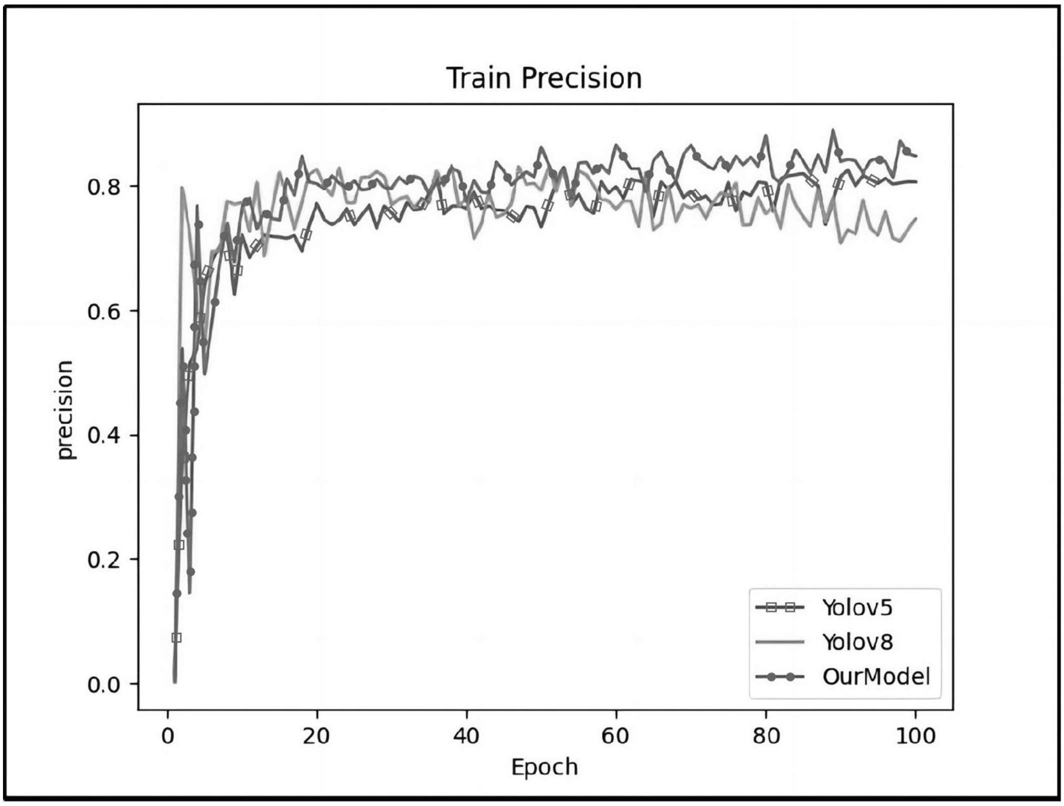

Figure 7. Precision graph of each model.

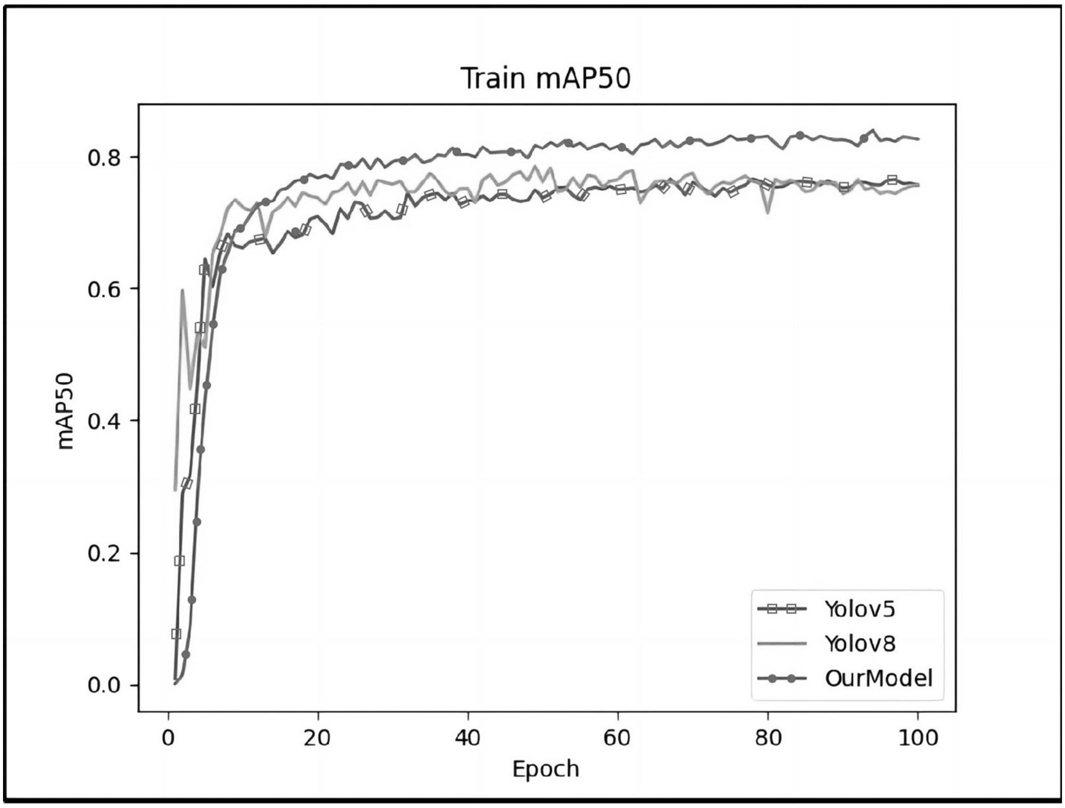

Figure 8. mAP50 of each model.

The mAP50 measure, to some extent, reflects the accuracy of the model when labeling data, and it can enhance detection efficiency and accuracy. Therefore, the OurModel model is superior to both the YOLOv8 and YOLOv5 models and can be used for identifying industrial safety helmets and detecting small target objects.

4.3. Final Model

After the above experiments, it can be concluded that the proposed network structure is effective for recognizing small targets. Therefore, the parameters of OurModel model will be tuned to obtain the optimal solution.

Training is performed on a GPU with the initial settings given in Table 2. As shown in the training environment and related parameters, different parameters will be set to obtain the trained model (OurModel). Based on the training of the model, the accuracy of each round in the validation set will be saved and plotted. The loss results are shown in Figure 9.

According to the precision of OurModel model, with the increase of training rounds, the model tends to stabilize, and the maximum value of accuracy exceeds 0.9. It can be seen that the improved YOLOv8 model has a much higher performance than the original YOLOv8 model.

When the number of training rounds is adjusted to 200, according to the loss diagrams given in Figure 10 for both the training set and in the validation set, the losses of predicting target boxes and target categories are very low. After reaching a certain number of rounds in the validation set, they basically stabilize. Therefore, the improvement is effective.

The inference visualization of the model is shown in Figures 11 and 12.

Table 2. Training parameters and methods of OurModel.

Related Parameters Settings	**Value**
epochs	100
imgsz	640
Batch size	32
optimizer	AdamW

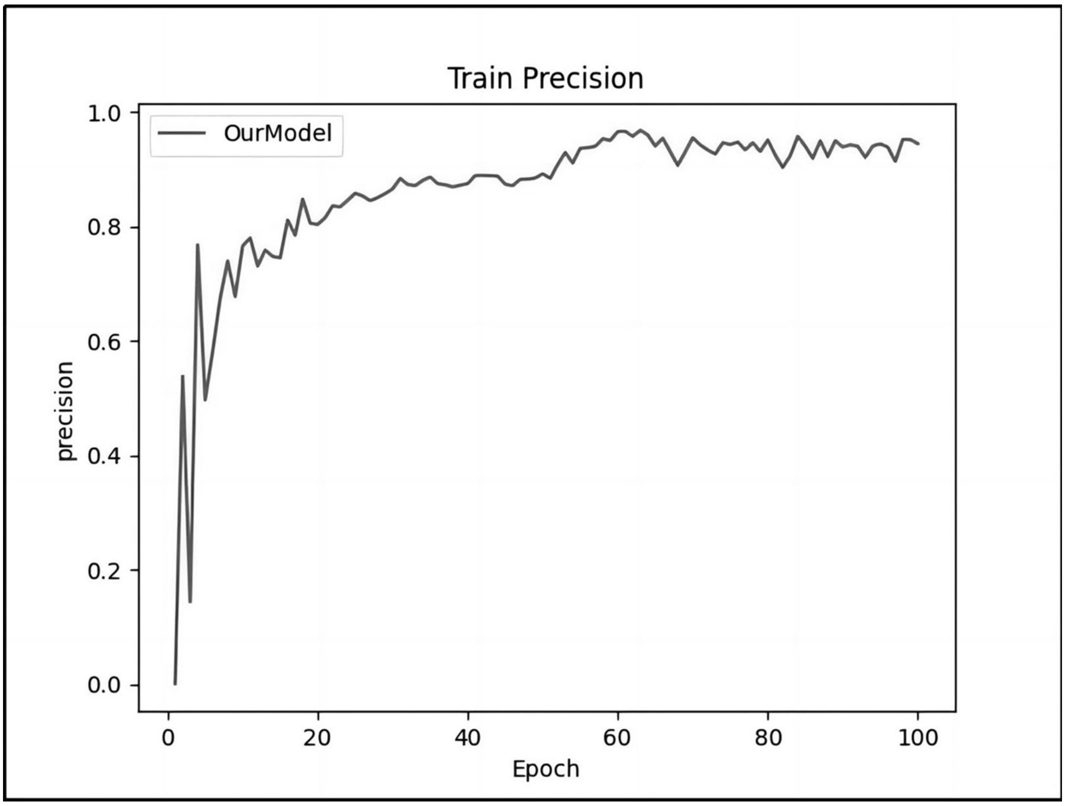

Figure 9. Precision of OurModel.

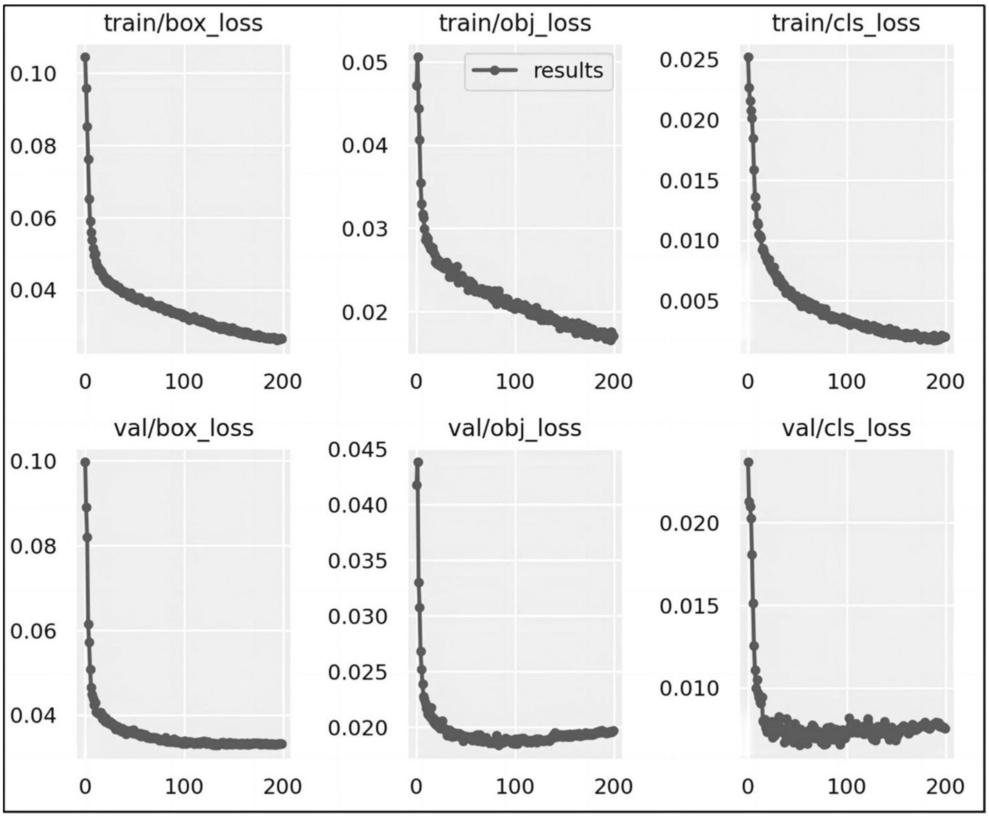

Figure 10. Loss of OurModel.

Figure 11. The Inference Structure of OurModel Model.

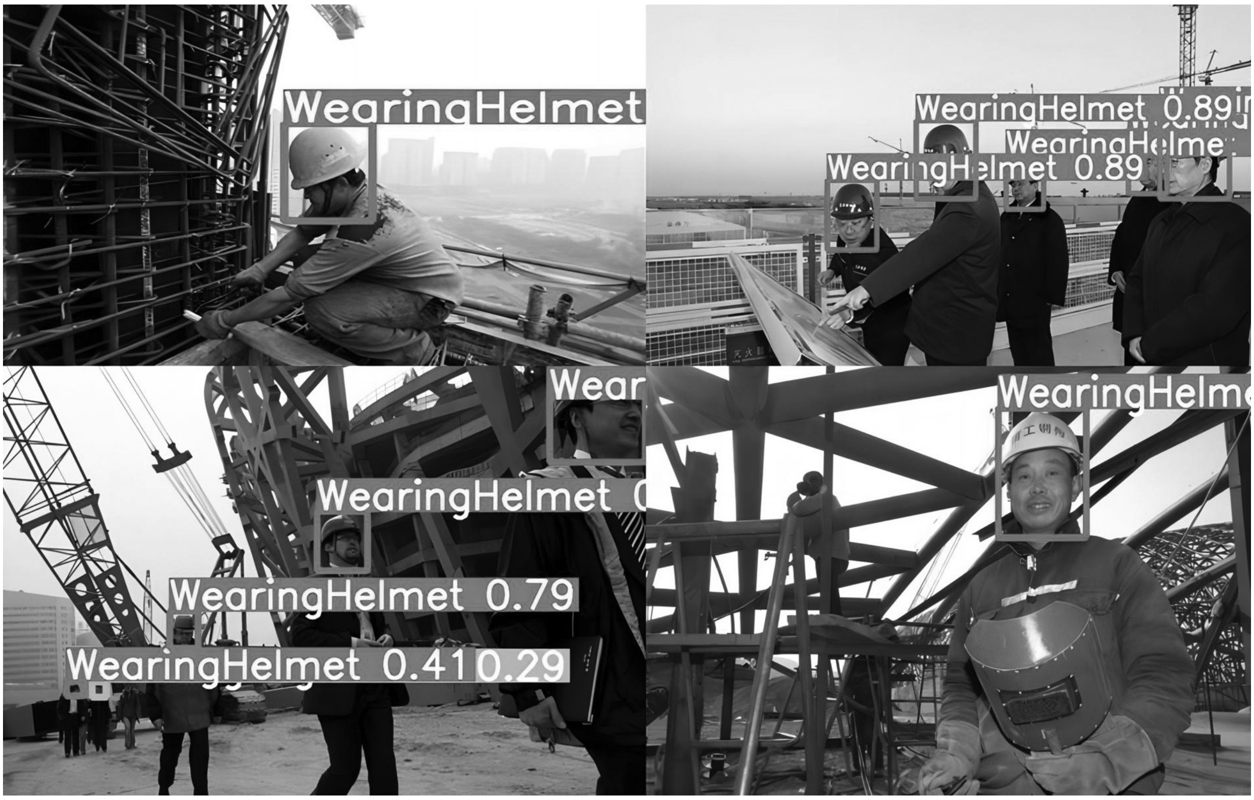

Figure 12. Inference of OurModel Model.

5. Conclusion

In conclusion, this paper presented a robust objection detection technique tailored for accurate identification of small safety helmets in industrial settings by augmenting YOLOv8. Structural improvements via the proposed SPD-Conv and LSKNet modules help retain multi-scale visual information and focus contextual modeling through dynamic spatial selection and attention. Comparative evaluations indicate consistently higher precision and recall over the default YOLO frameworks. The final optimized model achieves over 90% validation accuracy, demonstrating real-world reliability. While constrained to headgear use cases, the methodologies and experiments provide useful insights into enhancing deep neural networks for allied minute object detection challenges in complex environments. As industrial IoT and automated monitoring systems gain traction, such computer vision capabilities form integral components to uphold safety and compliance.

References

[1] C. Shan *et al.*, "Research on Improved Algorithm for Helmet Detection Based on YOLOv5", *Scientific Reports*, vol. 13, no. 1, 2023. http://dx.doi.org/10.1038/s41598-023-45383-x

[2] Y. Pengcheng *et al.*, "Recognition and Location of Coal Gangue Based on BiFPN and ECA Attention Mechanism", *International Journal of Coal Preparation and Utilization*, pp. 1–13, 2023. http://dx.doi.org/10.1080/19392699.2023.2270920

[3] J. Chen *et al.*, "YOLOv7-WFD: A Novel Convolutional Neural Network Model for Helmet Detection in High-Risk Workplaces", *IEEE Access*, 2023. http://dx.doi.org/10.1109/ACCESS.2023.3323588

[4] L. Deng *et al.*, "A Lightweight YOLOv3 Algorithm Used for Safety Helmet Detection", *Scientific Reports*, vol. 12, no. 1, 2022. http://dx.doi.org/10.1038/s41598-022-15272-w

[5] F. Zhou *et al.*, "Safety Helmet Detection Based on YOLOv5", in *Proc. of the 2021 IEEE International conference on power electronics, computer applications (ICPECA). IEEE*, 2021, pp. 6–11. http://dx.doi.org/10.1109/ICPECA51329.2021.9362711

[6] L. Huang *et al.*, "Detection Algorithm of Safety Helmet Wearing Based on Deep Learning", *Concurrency and Computation: Practice and Experience*, vol. 33, no. 13, 2021. http://dx.doi.org/10.1002/cpe.6234

[7] L. Wang *et al.*, "Industrial Helmet Detection Using Support Vector Machines", *Machine Learning in Industry*, 2021.

[8] F. Pang and X. Chen, "MS-YOLOv5: A Lightweight Algorithm for Strawberry Ripeness Detection Based on Deep Learning", *Systems Science & Control Engineering*, vol. 11, no. 1, 2023.

[9] Y. Qian and B. Wang, "A New Method for Safety Helmet Detection Based on Convolutional Neural Network", *PLoS ONE*, vol. 18, no. 10, 2023. https://doi.org/10.1371/journal.pone.0292970

[10] K. Tan *et al.*, "YOLOV8-based Method for Industrial Safety Helmet Small Target Detection", *Deep Learning Research*, 2022.

[11] J. Mi *et al.*, "Three-dimensional Visualization of Thyroid Ultrasound Images Based on Multi-scale Features Fusion and Hierarchical Attention", 2023.

[12] S. Chen *et al.*, "Residual Network-based Approach for Industrial Safety Helmet Small Target Detection", *Industrial Technology Review*, 2023.

[13] J. Redmon *et al.*, "You Only Look Once: Unified, real-time Object Detection", in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016. http://dx.doi.org/10.1109/CVPR.2016.91

[14] J. Wang *et al.*, "Mask OBB: A Semantic Attention-based Mask Oriented Bounding Box Representation for Multi-category Object Detection in Aerial Images", *Remote Sensing*, vol. 11, no. 24, 2019. http://dx.doi.org/10.3390/rs11242930

[15] M. Mahasin and I. A. Dewi, "Comparison of CSPDarkNet53, CSPResNeXt-50, and EfficientNet-B0 Backbones on YOLO V4 as Object Detector", *International Journal of Engineering, Science and Information Technology*, vol. 2, no. 3, pp. 64–72, 2022. http://dx.doi.org/10.52088/ijesty.v2i3.291

[16] S. Zhang and G. Jintian, "Vehicle Detection Algorithm for Highway Driving Scenarios", in *Proc. of the 5th International Conference on Artificial Intelligence and Computer Science (AICS 2023)*, vol. 12803, 2023. http://dx.doi.org/10.1117/12.3009232

[17] T. Juan and D. Cordova-Esparza, "A Comprehensive Review of YOLO: From YOLOv1 to YOLOv8 and Beyond", arXiv preprint arXiv:2304.00501, 2023.

[18] H. Jan *et al.*, "Learning Non-maximum Suppression", in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017. http://dx.doi.org/10.1109/CVPR.2017.685

[19] R. Sunkara and T. Luo, "No More Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects" ECML/PKDD, 2022.

[20] Y. Chen *et al.*, "Image Super-resolution Reconstruction Based on Feature Map Attention Mechanism", *Applied Intelligence*, vol. 51, pp. 4367–4380, 2021. http://dx.doi.org/10.1007/978-3-031-26409-2_27

[21] D. Li *et al.*, "Involution: Inverting the Inherence of Convolution for Visual Recognition", in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. http://dx.doi.org/10.1109/CVPR46437.2021.01214

[22] R. Takahashi *et al.*, "Data Augmentation Using Random Image Cropping and Patching for Deep CNNs", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 9, pp. 2917–2931, 2019. http://dx.doi.org/10.1109/TCSVT.2019.2935128

[23] H. Xu *et al.*, "An Enhanced Framework of Generative Adversarial Networks (EF-GANs) for Environmental Microorganism Image Augmentation with Limited Rotation-invariant Training Data", *IEEE Access 8*, pp. 187455–187469, 2020. http://dx.doi.org/10.1109/ACCESS.2020.3031059

[24] R. Stephens, "Analyzing jitter at High Data Rates", *IEEE Communications Magazine*, vol. 42, no. 2, 2004. http://dx.doi.org/10.1109/MCOM.2003.1267095

[25] Y. Li *et al.*, "Large Selective Kernel Network for Remote Sensing Object Detection", ArXiv abs/2303.09030 (2023): n. pag.

[26] J.-S. Lim *et al.*, "Small Object Detection Using Context and Attention", in *Proc. of the 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). IEEE*, 2021. http://dx.doi.org/10.1109/ICAIIC51459.2021.9415217

[27] Y. Li *et al.*, "A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition", *Drones*, vol. 7, no. 5, p. 304, 2023. http://dx.doi.org/10.3390/drones7050304

[28] Z. Zhong *et al.*, "Random Erasing Data Augmentation", in *Proc. of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, 2020. http://dx.doi.org/10.1609/aaai.v34i07.7000

Received: November 2023
Revised: December 2023
Accepted: December 2023

Contact addresses:
Lan Yuanshuai
School of Electronic Information Engineering
Geely University
Chengdu
Sichuan
China
e-mail: 448916030@qq.com

Chen Mo
School of Electronic Information Engineering
Geely University
Chengdu
Sichuan
China
e-mail: 993287003@qq.com

Li Chuan
Chengdu College
University of Electronic Science and Technology of China
Chengdu
Sichuan
China
e-mail: 2463739729@qq.com

Wang Qian
School of Electronic Information Engineering
Geely University
Chengdu
Sichuan
China
e-mail: 6327752@qq.com

Liao Min
School of Electronic Information Engineering
Geely University
Chengdu
Sichuan
China
e-mail: 859490624@qq.com

LAN YUANSHUAI obtained his M.Eng. degree from Chengdu University of Information Technology, Chengdu, China in 2021. Now, he is a lecturer at Geely University of China, with research interests in the application of artificial intelligence and computer vision technologies.

CHEN MO obtained her M.Sc. degree from Technische Universitaet Clausthal, Lower Saxony, Germany, in 2017. Now, she is a teaching assistant at Geely University of China, with research interests in artificial intelligence and big data technologies.

LI CHUAN obtained his B.Eng. degree from the Chengdu College of University of Electronic Science and Technology of China, Chengdu, China, in 2019. Now he is a teaching assistant at the Chengdu College of University of Electronic Science and Technology of China, with research interests in computer vision technologies.

WANG QIAN obtained her M.Eng. degree from University of Electronic Science and Technology of China, Chengdu, China, in 2006. Now, she is a professor at Geely University of China, with research interests in the field of embedded systems.

LIAO MIN obtained her LL.M. degree from Sichuan University, Chengdu, China. in 2017. Now. she is a teaching assistant at Geely University of China, with research interests in the areas of big data and the Internet of things.