Journal of Real-Time Image Processing (2024) 21:166
https://doi.org/10.1007/s11554-024-01548-z

RESEARCH

A real-time foreign object detection method based on deep learning in complex open railway environments

Binlin Zhang[1] · Qing Yang[1] · Fengkui Chen[1] · Dexin Gao[2]

Received: 3 July 2024 / Accepted: 23 August 2024

Abstract
In response to the current challenges of numerous background influencing factors and low detection accuracy in the open railway foreign object detection, a real-time foreign object detection method based on deep learning for open railways in complex environments is proposed. Firstly, the images of foreign objects invading the clearance collected by locomotives during long-term operation are used to create a railway foreign object dataset that fits the current situation. Then, to improve the performance of the target detection algorithm, certain improvements are made to the YOLOv7-tiny network structure. The improved algorithm enhances feature extraction capability and strengthens detection performance. By introducing a Simple, parameter-free Attention Module for convolutional neural network (SimAM) attention mechanism, the representation ability of ConvNets is improved without adding extra parameters. Additionally, drawing on the network structure of the weighted Bi-directional Feature Pyramid Network (BiFPN), the backbone network achieves cross-level feature fusion by adding edges and neck fusion. Subsequently, the feature fusion layer is improved by introducing the GhostNetV2 module, which enhances the fusion capability of different scale features and greatly reduces computational load. Furthermore, the original loss function is replaced with the Normalized Wasserstein Distance (NWD) loss function to enhance the recognition capability of small distant targets. Finally, the proposed algorithm is trained and validated, and compared with other mainstream detection algorithms based on the established railway foreign object dataset. Experimental results show that the proposed algorithm achieves applicability and real-time performance on embedded devices, with high accuracy, improved model performance, and provides precise data support for railway safety assurance.

Keywords YOLOv7-tiny · Object detection · Deep learning · Complex environment

1 Introduction

In recent years, railways have remained the primary mode of coal transportation [1]. To meet railway transportation goals and enhance efficiency, the safety requirements for railway transportation have become increasingly stringent. Current research mainly focuses on special enclosed environments such as high-speed rail and subways. However, research on safety warnings for open railways remains a significant and challenging issue. This project aims to develop a safe and reliable foreign object detection algorithm for open railway perimeters to ensure an efficient, safe, and orderly transportation environment for open railways. In recent years, the number of railway safety personnel has significantly decreased due to a large wave of retirements in mining areas. Additionally, the awareness of safety and traffic regulations among people living along railway lines is weak. There are frequent incidents of people and vehicles rushing through nearby crossings, especially during school hours or busy farming seasons, leading to conflicts between moving trains and pedestrians or social vehicles. These conflicts often result in emergency braking and stopping of trains, particularly in curved sections. The unexpected, irregular, and unpredictable issues along the railway lines trigger safety incidents that severely affect normal railway operations. Moreover, foreign objects such as pedestrians, vehicles, animals, and falling rocks infringe upon the railway during

✉ Qing Yang
03390@qust.edu.cn

[1] School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China

[2] School of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, China

Published online: 06 September 2024

locomotive operations. Combined with varying environmental backgrounds along the tracks and the impact of extreme weather conditions over long periods, these factors create a complex and dynamic open railway operating environment. Real-time detection of foreign objects on railway perimeters is key to addressing the frequent conflict incidents in open environments. In recent years, railway departments have been actively exploring new methods [2] for railway line inspection. To ensure the safety of locomotives in mining areas, fixed surveillance cameras are deployed at critical locations along the railway lines, such as railway-road intersections, bridges, and curves, to monitor railway perimeter safety in real-time. Simultaneously, real-time monitoring of the front and rear of the train is conducted during locomotive operations to ensure safe passage through unmonitored sections. Therefore, both onboard and fixed detection along the railway line will be the focus of this paper.

Since the development of target recognition, intelligent detection of foreign objects primarily employs two methods: traditional object detection algorithms and deep learning algorithms. Traditional algorithms have slower computation and detection speeds and lower accuracy, making them unsuitable for practical applications. In contrast, deep learning-based object detection algorithms can detect foreign objects more quickly and accurately, thus gaining increasing attention and application. Deep learning-based object detection algorithms are divided into two-stage and single-stage forms. Two-stage detection algorithms offer higher accuracy but slower detection speeds, which may not meet the need for rapid detection. Single-stage detection algorithms produce final results through a single forward pass, offering faster speeds. Although their accuracy is slightly lower, practical applications can meet accuracy requirements by improving network structures. Key single-stage algorithms include Single Shot multibox Detector (SSD) [3], RetinaNet [4], and You Only Look Once (YOLO) series algorithms [5, 6].

With the rapid development of deep learning, numerous road safety detection methods based on deep learning and machine vision techniques have emerged, such as Faster R-CNN, SSD, and YOLO algorithms. The focus of railway safety work has increasingly shifted to machine vision detection of foreign objects on railways. For example, literature [7] proposed a Faster R-CNN-based network model that replaces the fully connected layer with a global average pooling layer, increasing the number of anchors, and introducing transfer learning concepts, significantly improving the detection accuracy of people and vehicles. Literature [8] proposed a YOLOv3-based detection model with ResNet-18 as the backbone network, using a row anchor box segmentation algorithm and integrating a multi-scale residual module based on Octave. This model doubled the detection speed while ensuring accuracy, meeting real-time detection requirements for foreign objects. Literature [9] improved the YOLOv5s model by integrating the DW-Decoupled Head to construct hybrid feature channels and applying large convolution kernels to increase the receptive field, thereby enhancing the model's localization, classification, and feature extraction capabilities. Literature [10] also improved YOLOv5s by adding the ECA-Net channel attention mechanism, using the SPD-Conv module, and applying the EIOU loss function, which focused more on small object targets, improved detail extraction capabilities, and enhanced overall model accuracy with minimal time cost loss. YOLOv7, due to its efficient detection capabilities, has been widely used in engineering detection [11–15]. However, significant issues remain when directly applying this network to foreign object detection along open railway lines. Firstly, the complex open environment of railways increases interference in target recognition due to the background. Secondly, the long railway lines and varying adverse weather conditions during the journey increase the difficulty of feature extraction. Whether using fixed or mobile monitoring methods, cameras are key perception sensors due to their low deployment cost and large detection range. Therefore, research on fixed-end machine vision-based railway perimeter foreign object detection has become a hotspot in the field of proactive railway safety.

To address the performance and deployment issues of foreign object intrusion detection algorithms in the context of open railways, an improved network based on the YOLOv7-tiny algorithm is proposed, namely the SBG-YOLO network. This network adopts three significant improvements. It is noteworthy that "SBG-YOLO" is an acronym for SimAM Attention Mechanism, BiFPN Network Structure, and GhostNetV2 Module, highlighting the unique enhancements made to YOLO. The main improvements to the algorithm are as follows:

1. The integration of the SimAM attention mechanism into the Neck, enhancing the representation capability of ConvNet without introducing new parameters.
2. Inspired by the BiFPN structure, connections have been added connections between the Backbone and Neck to achieve cross-level feature fusion, improving feature fusion effects at different scales and enhancing object detection performance.
3. Introducing the GhostNetV2 module into the network structure significantly reduces the computational load of the network, achieving model lightweighting.

The structure of this article is as follows. The second section introduces the YOLOv7-tiny network structure used for foreign object detection and target recognition. The third section proposes the improved SBG-YOLO network based on YOLOv7-tiny, detailing the three improvements:

incorporating the SimAM attention mechanism, adopting cross-region fusion channels, and introducing the GhostNetV2 module. The fourth section describes the experimental platform and model evaluation metrics. The fifth section presents the comparison of experimental results between SBG-YOLO and other common models, including model parameters, computational load, Frames Per Second (FPS), recall rate, and mean Average Precision (mAP@0.5), as well as the depth and scale of different versions of the YOLOv7 model. Finally, the sixth section draws conclusions.

2 Related work

2.1 Current situation of open railway

Open railway foreign object detection encompasses not only target detection during locomotive operation but also target detection at unmanned crossings along the railway lines. These two components are interconnected in real-time and can switch seamlessly, enabling real-time detection and early prediction of railway obstructions. When a locomotive approaches within 2 km of a crossing, the onboard video automatically switches to display the target recognition signal from the nearest crossing in the direction of travel. This works in tandem with the locomotive's internal target recognition signals to ensure the safety of railway operations and the detection of foreign objects (such as people and animals) that may obstruct the railway.

Given the high complexity of foreign objects on railway tracks in open environments, there are significant challenges in deployment, as well as issues with false positives and missed detections. To address these problems in open railway environments, a novel method for railway foreign object detection is proposed. This method is designed to better recognize a variety of obstructions in complex open environments under embedded conditions. The performance of obstruction recognition in open railway environments is enhanced through three main approaches:

1. To address the issues of false positives and missed detections, an attention mechanism is embedded into the target recognition network, which enhances the representation capability of the convolutional layers. Additionally, the use of the BiFPN structure allows for the fusion of multi-level features, thereby improving target detection accuracy.
2. Building on railway obstruction image processing, LabelImg software is used to manually label and locate all foreign objects on the tracks to ensure the effective training of the target detection model, particularly for the detection of people and animals.
3. Given the limited computing power of embedded real-time detection systems [16], the smallest possible network structure is selected, significantly reducing computational requirements and enhancing embedded system performance. Improvement and optimization of deep learning models are crucial to ensuring detection performance.

By implementing these strategies, the proposed method aims to significantly improve the identification and prediction of railway foreign objects, thus enhancing the overall safety and efficiency of railway operations in open environments.

2.2 YOLOv7-tiny network structure

YOLOv7-tiny is the most concise model in the YOLOv7 series, with fewer parameters and faster detection speed. It is a network model designed for edge GPUs. YOLOv7-tiny uses Mosaic data enhancement and adaptive anchor frame computation for image preprocessing. Retain the model scaling strategy based on cascades and simplify the ELAN module, represented by ElAN-S. Backbone extracts features based on CBL, ELAN-S and MP structures. SPPCSPC module refers to the module of Spatial Pyramid Pooling (SPP) and Cross Stage Partial Connections (CSP), which combines the spatial pyramid method with the cross-stage partial connections method. It can be used to connect Backbone and Neck, and reduce the computation by half while maintaining the accuracy of the model. Neck uses Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) structure for feature fusion of objects of different scales. The CBL is composed of Convolution, Batch Normalization and Randomized LeakyReLU for improved feature fusion. In the Head part, standard convolution was used to replace RepConv, IDetect test head in YOLOR was introduced, implicit representation strategies were used to refine the prediction results, and large, medium and small images were classified according to the fused feature values. The output uses the combination of Focal_loss and CIoU_loss as a bounding box loss function, alleviating the class imbalance problem and better measuring the distance from the target box. As a result, YOLOv7-tiny has the ability to identify and locate objects and is able to handle many different scales and sizes with high precision and robustness. Its original structure is shown in Fig. 1.

Firstly, three improvements of YOLOv7-tiny are introduced, and then the overall framework of the improved SBG-YOLO is introduced.

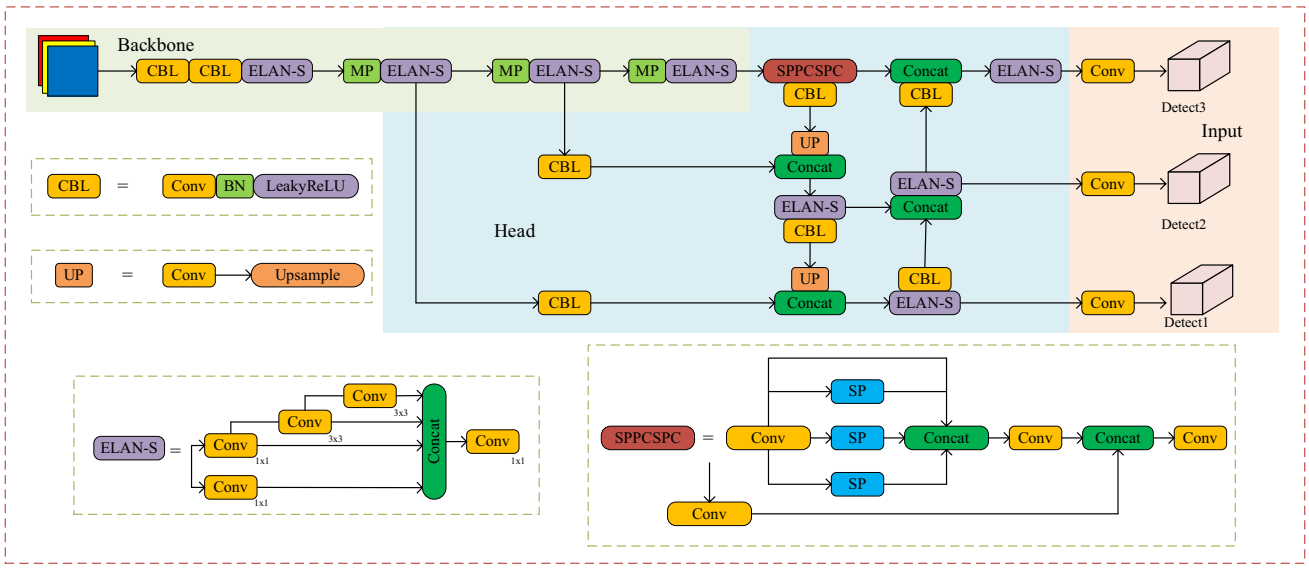


Fig. 1 Network structure of YOLOv7-tiny

3 SBG-YOLO

3.1 SimAM attention mechanism

Under the conditions of open railways, the complexity of background information and the ease with which railway foreign object information can be obscured by redundant information are major factors affecting foreign object detection performance. To enhance target detection performance, this paper tested mainstream attention mechanisms and selected the three-dimensional parameter-free attention mechanism. As a flexible and effective attention module for convolutional neural networks, SimAM does not add extra parameters to the original network. It provides 3D attention [17] weights to the feature maps in the detection layer, thereby enhancing the representational capability of Convolution networks.

SimAM module improves the feature fusion network. Compared with one-dimensional [18] and two-dimensional attention mechanisms [19], three-dimensional attention mechanisms can balance the importance of features more comprehensively and efficiently without introducing parameters, thus enhancing the feature weights of target regions. Through the operation of neurons, neurons with key information are given higher weights to improve the recognition and positioning accuracy of the network. Figure 2 shows a comparison of the different attention mechanisms.

The SimAM module looks for important neurons and defines the energy function. It takes a binary label and adds regular entries. Therefore, the minimum energy can be obtained by the following formula:

$$e_t^* = \frac{4\left(\sigma^2 + \lambda\right)}{(t-u)^2 + 2\sigma^2 + 2\lambda} \tag{1}$$

$$u_t = \frac{1}{M}\sum_{i=1}^{M-1} x_i \tag{2}$$

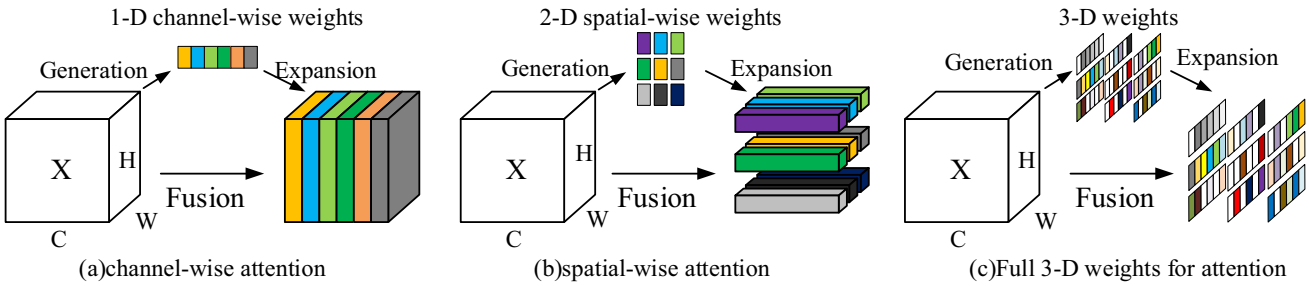


Fig. 2 Comparisons of different attention steps

$$\sigma_t^2 = \frac{1}{M-1}\sum_{i=1}^{M-1}\left(x_i - u_t\right)^2 \tag{3}$$

where u_t is the average of all neurons, σ_t^2 is the variance of all neurons, t is the target neuron, x_i is the other neurons in the input feature channel, λ is the regularization coefficient. Each channel has $M = H \times W$ energy function. The lower the energy, the more distinct the neuron t is from the surrounding neurons and the more important it is. Therefore, the importance of neurons can be obtained by $1/e_t^*$.

SimAM can bring good detection performance improvement. In this study, the attention mechanism is added to the feature fusion network in front of the detection head, as shown in Fig. 3.

3.2 NECK partial improvement

The PANet structure is adopted in YOLOv7-tiny, and a simple bidirectional fusion is formed by adding secondary fusion to improve the feature fusion capability. However, the introduced secondary fusion will interfere with the original feature information and affect the effect of feature fusion. Therefore, we first refer to the BiFPN structure that retains the original information, and then modify and optimize the network structure of PANet. The improved method optimizes the fusion effect of different feature layers, so as to achieve the purpose of improving the detection effect without significantly increasing the amount of computation.

Figure 4c shows the improved BiFPN feature fusion network. Bidirectional networks can be simplified by removing nodes with only one input edge, as they contribute little to a feature network that fuses different features. In addition, an additional channel is added between the first and last node of each element layer. Retain the original features for better fusion without significantly increasing computational costs. The blue arrow represents a top-down path that conveys high-level semantic feature information; the red arrow represents the bottom-up path, conveying the location information of the underlying feature; finally, the purple arrow is the new edge added between the first and last node of each layer. The original BiFPN [20] carries weights in the fusion process, and the introduced weights will not only cause the

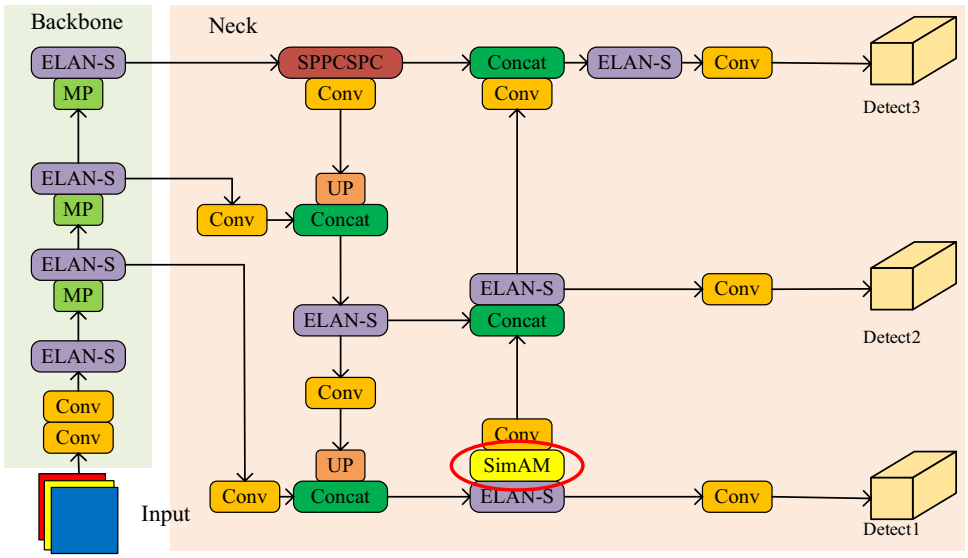


Fig. 3 One modified area of YOLOv7-tiny with embedded SimAM module

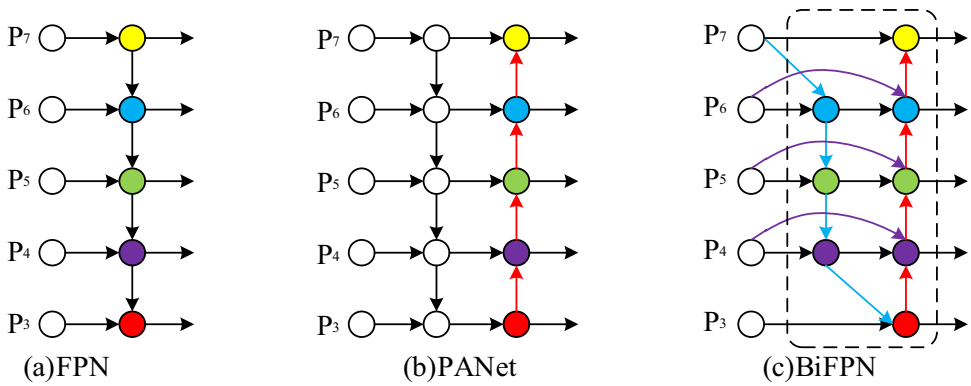


Fig. 4 FPN, PANet and BiFPN structures

network to be large and complex, but also reduce the recognition accuracy. Therefore, our structure does not introduce the weight part, but retains the original idea of building the fusion channel. Figure 5 shows our proposed structure. This improved approach is more stringent and can improve the performance of the network.

3.3 Architecture of GhostNetV2

The lightweight GhostNet [21] module divides the input feature map into two parts; one part generates the feature map through convolution, the other part directly performs linear convolution operation, and finally concatenates it. This can greatly reduce the computational cost, but some subtle and important features may be lost in the process, increasing risk of feature distortion. GhostNetV2 proposes the Decoupage Fully Connected attention mechanism (DFC attention) [22], which has the ability to dynamically calibrate and capture the production distance information. It is easier to deploy on hardware. Directly connecting DFC attention and Ghost module in parallel will introduce additional computing costs. But scaling the width and height of the feature to half the original reduces the computational effort of DFC attention by 75.0%. Then the obtained feature map is restored to the original size through the up-sampling operation to match the resolution of Ghost branch features. Figure 6 illustrates the principle of the GhostNetV2 module. GhostNetV2 adopts reverse bottleneck design, using two Ghost Modules to increase the feature dimension first and then reduce it. This design strategy naturally decoupled the model's performance

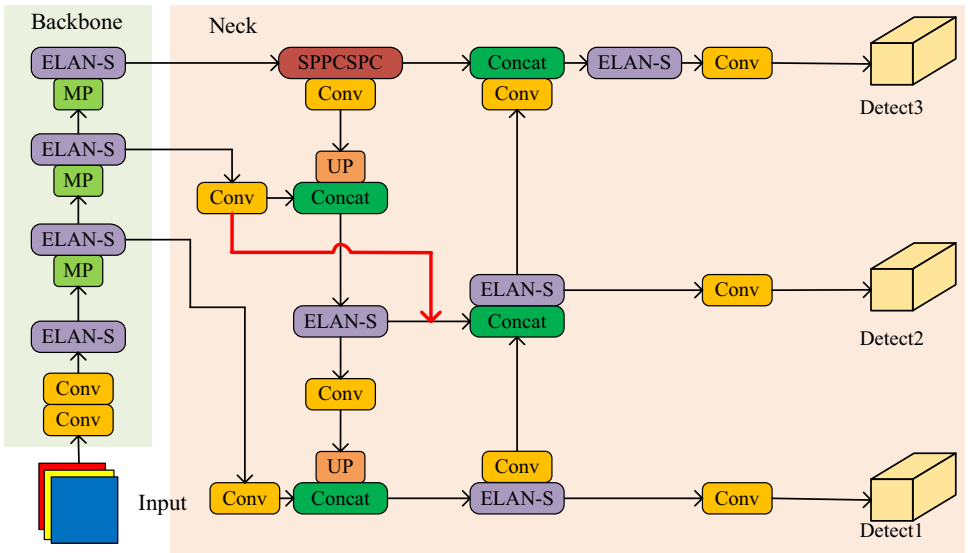
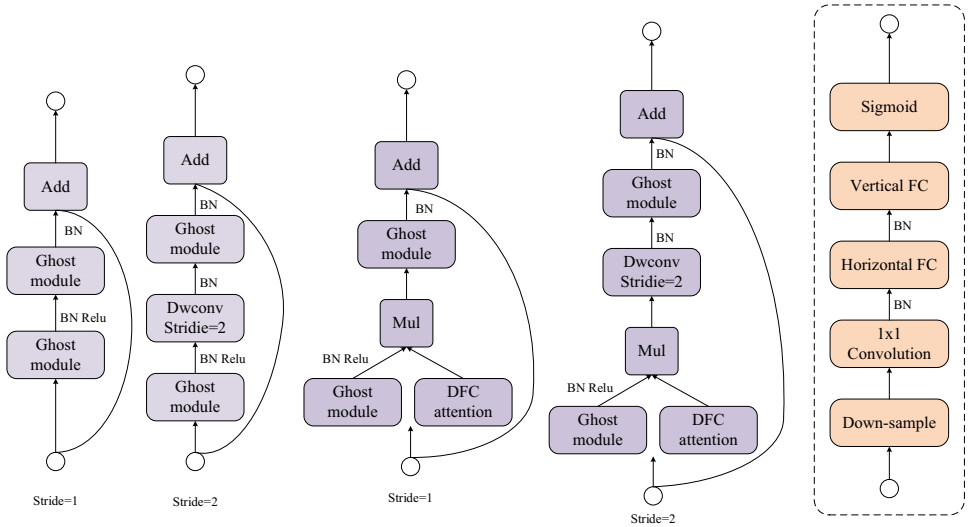


Fig. 5 Improvement of the network structure of the Neck part



(a) GhostNet (b) GhostNetV2 (c) DFC attention

Fig. 6 Principle of the GhostNetV2 module

ability and capacity. Parallelizing DFC attention with the first Ghost module enhances the features of the extension. The enhanced features are input into the Ghost module to generate output features, and the long-distance dependence between pixels in different spatial locations is captured at the same time, which greatly reduces the computational complexity and greatly enhances the feature fusion ability and the model expression ability. GhostNetV2 parallelizes network computing by grouping channels, which can accommodate input data of different sizes with less computing overhead. In addition, the low-rank decomposition technique is used to reduce the number of redundant parameters and ensure the accuracy of the model.

Incorporating the Cross Stage Partial Networks (C3, CSPNet) ahead of the large feature size detection head, Detect3, leverages advanced features provided by the C3 module for complex object detection. C3 splits the input feature map into two, processes one part through a bottleneck layer, then merges them back, enhancing gradient flow and reducing information loss, thereby increasing efficiency. This structure balances performance and speed, enhances feature representation, and when combined with GhostNetV2, enriches feature extraction while maintaining low resource use. This integration is ideal for deployment in resource-limited environments, making the model more efficient and powerful for real-time object detection.

Considering that the GhostNetV2 module met the model's improvement needs and performed excellently when combined with the C3 module, we conducted over 100 experiments, ultimately determining the deployment location of the GhostNetV2 module. Figure 7 shows the network structure of the improved GhostNetV2 module.

3.4 Improved network structure

This paper proposes an algorithm named SBG-YOLO, whose network structure is shown in Fig. 8. The algorithm integrates several techniques and methods, including SimAM, BiFPN, and GhostNetV2, resulting in more accurate and stable detection outcomes. The inclusion of the SimAM attention mechanism in the feature fusion network before the detection head increases the weight of critical railway foreign object targets. Inspired by the BiFPN structure, the 8th layer and the 22nd layer are concatenated to achieve cross-layer feature fusion. Additionally, the ELAN-S modules in the 6th and 19th layers of the overall network structure are replaced with GhostNetV2 modules, significantly reducing the computational load and effectively enhancing algorithm performance. Replacing the ELAN-S module with the C3 module before Detect3 enhances feature representation and gradient flow. The improved network structure not only meets the lightweight deployment requirements [23, 24] but also enhances target detection performance, fulfilling the needs of practical engineering applications.

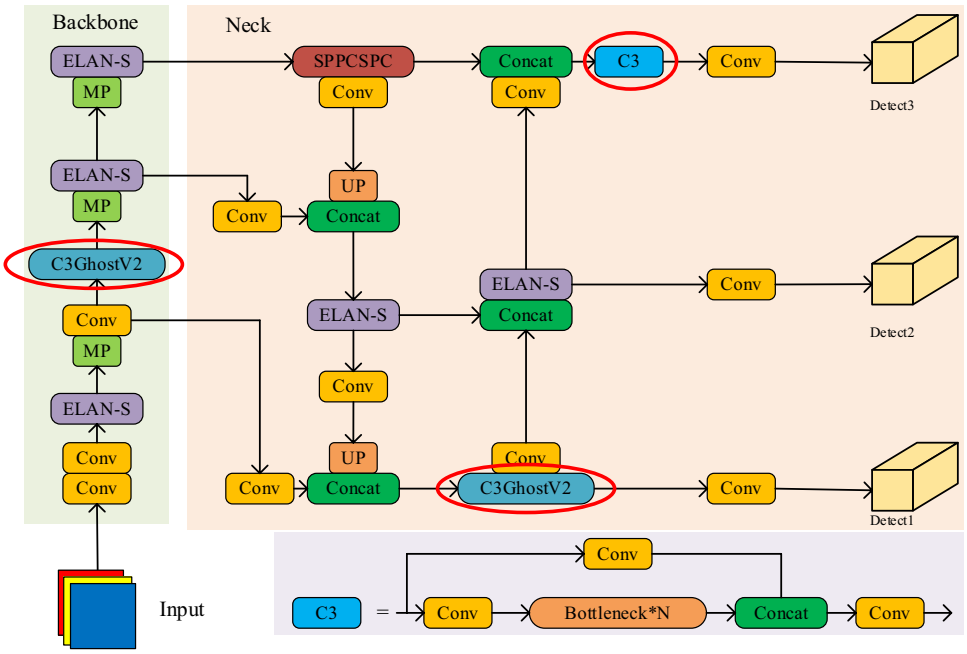


Fig. 7 GhostNetV2 with replacement of ELAN-S

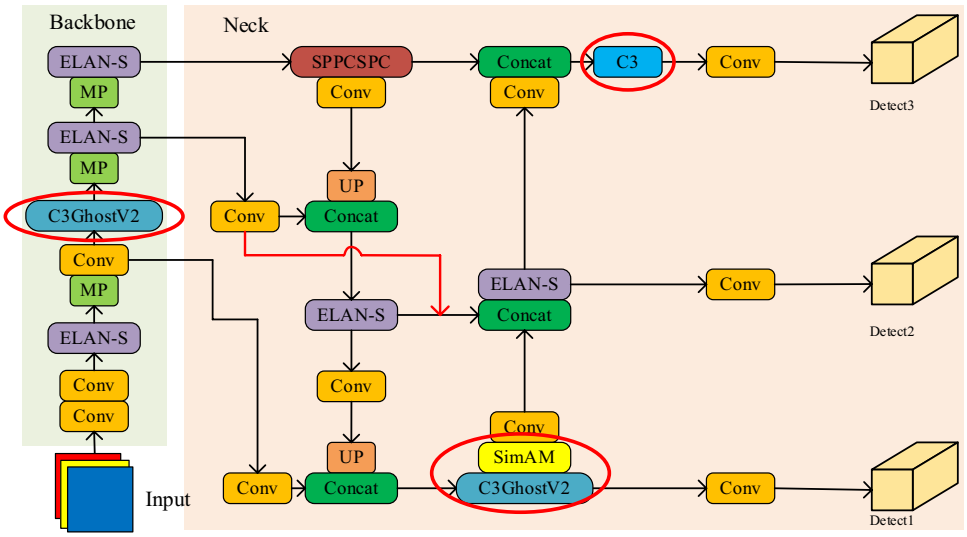


Fig. 8 Network structure of SBG-YOLO

Table 1 Training platform setup

Development environment	Version
Batch size	128
Image size	640*640*3
Epoch	150
Weight	0
Lr0	0.01
Learning rate decay strategy	Cos

4 Experimental platform

4.1 Training platform setup

All experiments use pytorch2.0.0 deep learning framework to train the model under Win11 system. The CPU and GPU of the training platform are Inter i9-13900HX processors, NVIDIA GeForce RTX4090 24 g Gpus, Python version 3.8, Cuda version 11.8. The experiment did not use the officially provided pre-training weights for training on the COCO dataset. Before training the model, the parameters in the Train module need to be adjusted to ensure that the parameters of all network models are the same. Table 1 shows the specific parameters after adjustment.

4.2 Embedded platform setup

To verify the detection effect of the improved algorithm in the actual railway environment, improved YOLOv7-tiny model was deployed to the Jetson Xavier NX embedded terminal for testing. The embedded platform used in this paper is the Jeston Xavier NX, which has two DLAs (Deep Learning Accelerator) and can reach a maximum arithmetic power of 21 TOPs to accelerate the inference of the model. This device has applications in many areas of deep learning. The system is configured with Ubuntu 16.04 for ARM and the model running environment is configured with JetPack 4.6, Python 3.6, Pytorch 1.8 and Cuda 10.2. Its structure is shown in Fig. 9.

4.3 Data set construction

Open railways are often set up in unattended suburbs, and the railway perimeter environment is very complicated. In order to ensure the high precision detection of the intrusion target in the complex railway environment, the experiment needs to collect a sufficient number of railway foreign object data sets for training. Since there is no public railway foreign body data set, this project constructs a self-built railway foreign body data set based on Pscal VOC2007 data set.

In the open railway environment [25], the scene of railway foreign body encroachment is simulated. Different kinds of foreign objects are placed on the railway track and video is captured by high-definition cameras. A total of 2152 pictures of railway foreign body intrusion were obtained by video collection at 3 s interval. In order to avoid insufficient data sets and single scenes affecting the training effect, this paper captures different types of railway foreign body intrusion images through the network, and extends the self-built data set by combining part of the Pscal VOC2007 data set. By combining the images collected in the suburban railway environment with the images collected on the Internet, a self-built railway foreign body data set consisting of 18,323 images was constructed, including images under different weather conditions, jitter and blur images, and images of different angles and sizes. The self-built data set was manually annotated using the Labelimg tool, which contained 12 types

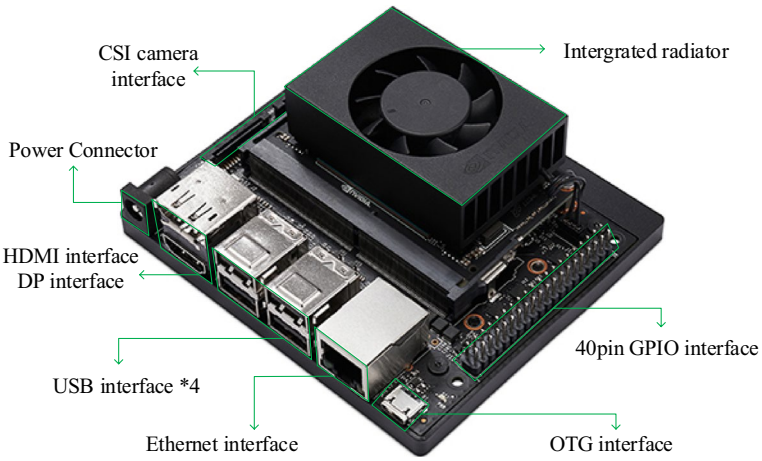


Fig. 9 Structure of Jeston Xavier NX

Table 2 Number of foreign objects in training set

Type	Number
Person	6124
Bus	2528
Bicycle	2320
Motorbike	1017
Train	2427
Stone	2227
Dog	2632
Cat	2152
Sheep	1039
Cow	1257
Horse	1052
Car	6728

Table 3 Number of foreign objects in training set

Size	Number
Small target	12,851
Middle target	10,570
Big target	8482

of intruders, including people, vehicles, animals, etc. The data set is divided into training set, test set and verification set according to the ratio of 7:1:2. The number of intrusions in the training set is shown in Table 2.

In order to avoid overfitting caused by slight changes in the target size of the dataset, images of railway intrusion objects of different sizes were selected as the dataset, and the target size of the railway intrusion test set was statistically analyzed according to the definition standard of target size in the MS COCO dataset. Specific data are shown in Table 3. The bounding box size less than 32×32 is defined as a small target, the bounding box size between 32×32 and 96×96 is defined as a medium target, and the bounding box size of other sizes is defined as a large target.

4.4 Model evaluation index

In the experiments of this paper, Precision, Recall, Average Precision (AP), mAP@0.5, Floating point Operations Per Second (FLOPS) and detection speed are used to comprehensively evaluate the precision performance and deployment performance of the object detection algorithm. When the model performs target detection, four detection results are obtained: TP, FP, TN and FN. Where TP is the true value, representing the number of correctly detected objects. FP represents the false-positive value of the number of incorrectly detected objects, and FN represents the false-negative value of the number of undetected objects. Therefore, Precision and Recall can be expressed as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

AP can be obtained by calculating the area under the PR curve formed by Precision and Recall. The AP values for all categories are averaged to get the mAP. The AP and mAP of a class of objects can be represented as:

$$AP = \int_0^1 P(R)dr \tag{6}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{7}$$

By combining Precision and Recall, $F1$ can more effectively reflect the accuracy performance of the network model. $F1$ score can be expressed as:

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{8}$$

In addition, according to the application requirements of railway perimeter foreign body intrusion detection, FLOPS and detection speed are used to verify the deployment performance of the network model. Parameters are the total number of parameters that need to be trained in model training. FLOPS refers to model computation, which is a parameter to measure algorithm complexity. The detection speed represents the number of frames of image or video detected.

In this study, the deployment performance of the model is evaluated based on three indicators: parameters, Floating point Operations (FLOPs), and detection speed. The parameters in the model represent the number of parameters involved in the model, which is closely related to the depth and width of the network. Reducing the number of parameters can improve the deployment performance of the model, especially in resource-constrained environments such as embedded systems and mobile devices. FLOPs indicate the computational load of the model and are a parameter for measuring algorithm complexity. Lower FLOPs mean less computational resource and memory usage, making the model more suitable for deployment on embedded devices with limited arithmetic capabilities. Detection speed is measured in FPS, with a higher frame rate indicating that the model can process image frames faster, reflecting faster detection speed. A faster detection speed demonstrates the superior detection performance of the algorithm model under the same hardware conditions.

5 Experimental results and analysis

5.1 Analysis of experimental results

The preprocessed images are input into the YOLOv7-tiny network model, and the model is trained according to preset parameters. The loss function is one of the most critical evaluation metrics in machine learning. The faster the loss function converges, the stronger the model's feature extraction ability and the better the model's performance. To demonstrate the effectiveness of the NWD loss function [26], the loss function of the improved YOLOv7-tiny algorithm is compared with that of the original algorithm, as shown in Fig. 10a. The training includes 150 iterations. During the first 20 rounds of training, the loss function value drops sharply and gradually stabilizes in subsequent training, achieving convergence by the 140th iteration. Compared to the original network, the improved YOLOv7-tiny's loss function converges faster and more smoothly, with lower loss function values, bringing the predicted frames closer to the actual targets and enhancing the model's target localization performance and detection accuracy. The model detection results are shown in Fig. 10b. The mAP@0.5 value rises rapidly in the initial stage and stabilizes after approximately 100 iterations of training. Compared to the original network, the improved YOLOv7-tiny algorithm converges faster and more smoothly, indicating more stable detection performance. Additionally, the detection accuracy of the model is improved by 3.3%, ultimately maintaining at 77.5%. This result underscores the model's excellent performance in terms of accuracy.

The results are shown in Fig. 11. Figure 11a illustrates the relationship between this model accuracy and the confidence

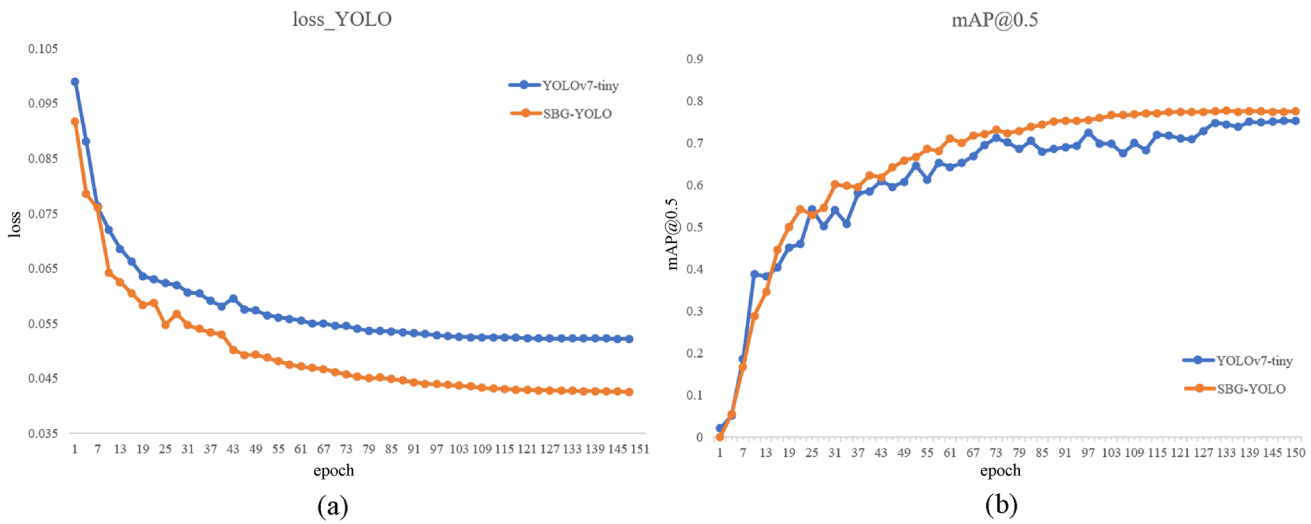


Fig. 10 Improved YOLOv7-tiny training results

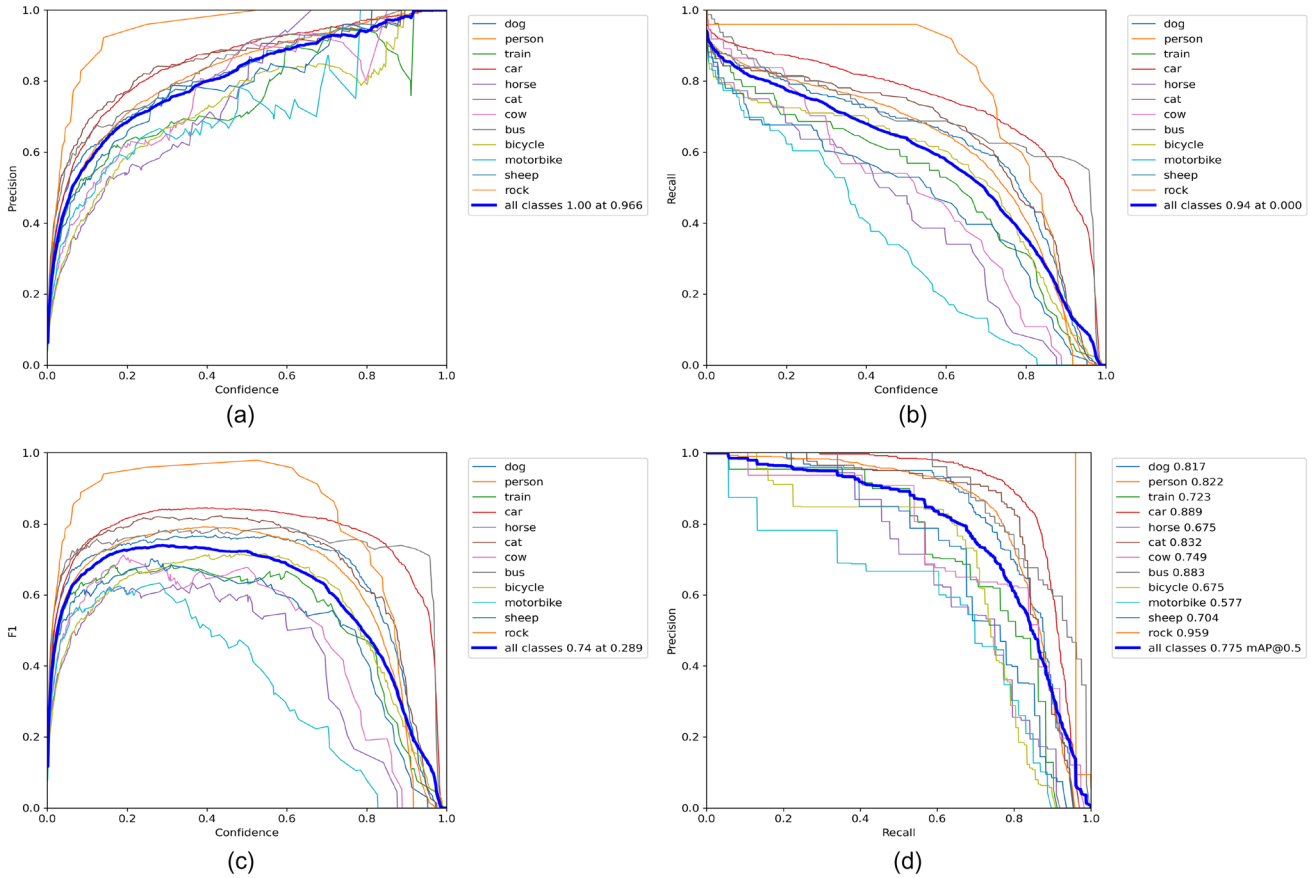


Fig. 11 Improved YOLOv7-tiny training result line graph

threshold. The results show that the accuracy of the fitting results increases with the increase of the confidence threshold, and the recovery rate is up to 96.6%. This phenomenon indicates that this model is able to achieve a low false alarm detection rate of a higher confidence threshold. The curve in Fig. 11b reflects that the recall rate decreases with the increase of confidence threshold and the recall rate reaches 94.0%. These data show that the model has higher sensitivity in terms of lower confidence in the threshold and is able to detect more positive samples. The variation curve of F1 value is shown in Fig. 11c for the performance of Precision and Recall, and F1 value, as an overall performance indicator, provides a comprehensive evaluation of the detection accuracy of the model. In this study, the highest F1 value was obtained, the value of YOLOv7-tiny model is 0.74, which proves that the model has a good balance between the precision rate and the recall rate, ensuring the accuracy and reliability of accurate detection. Figure 11d shows the AP value for each foreign object category, which is composed of the surrounded area accuracy and recall rate, from which it can be seen that YOLOv7-tiny has good performance results for all kinds of foreign objects [27] invading the railway. They are sorted according to the size of the enclosed area, i.e. dogs, people, trains, cars, horses, cats, cows, buses, bicycles, motorcycles, sheep and stone cats in order.

5.2 Comparison of network models

In order to better verify the superiority of the improved model in the performance of foreign body intrusion detection under the background of open railway, the international mainstream lightweight network was trained under the same conditions, and different models were analyzed and compared. To better demonstrate the benefits of the improved model, this experiment compared SBG-YOLO with the models of YOLOv5s, YOLOX-tiny, YOLOv8s, and YOLOv7-tiny. All these models were trained and validated on self-built data sets, and the test results are shown in Table 4. From the improvement of the YOLO version, it can be seen that with the upgrade of the version, the image recognition accuracy has been greatly improved. For example, from YOLOX-tiny to YOLOv5s, the recognition accuracy has improved by 4.6%; From YOLOv5s to YOLOv7-tiny, the recognition accuracy has been improved by 2.9%. Compared to the YOLOv7-tiny, our improved SBG-YOLO improves the recognition

Table 4 Comparison of performance metrics of different target detection algorithms

Algorithm	mAP@0.5	Parameters/MB	FLOPS	FPS
YOLOv7-tiny	74.2%	6.044	13.3G	117.6
YOLOv5s	71.6%	7.084	16.2G	86.2
YOLOX-tiny	67.0%	6.056	13.3G	109.9
YOLOv8s	74.9%	11.132	28.5G	99.0
SBG-YOLO	**77.5%**	**6.208**	**5.4G**	**122.0**

accuracy by 3.3% and greatly reduces the computational effort. In the subsequent table, the bolded data points represent key metrics that demonstrate the superior performance of the SBG-YOLO algorithm model, particularly highlighting its minimal FLOPS and optimal mAP@0.5.

As shown in Table 4, the comparison experiment selected network models from various YOLO series that exhibit significantly lightweight characteristics in terms of parameter count, computational load, and model weights. However, each detection model has certain drawbacks. For instance, YOLOv5s sacrifices considerable accuracy in mAP to achieve higher detection speed, which increases the risk of false positives and missed detections in practical applications. Although YOLOv8s improves detection accuracy, it significantly increases the parameter count and computational load of its network structure. Therefore, we chose to improve the YOLOv7-tiny network, which performs well in all aspects. The improved YOLOv7-tiny achieves a maximum mAP of 77.5%, delivering high precision in railway perimeter intrusion detection. By integrating the GhostNetv2 network, it significantly reduces the computational load of the model, greatly enhancing the detection speed. A comprehensive evaluation of various performance metrics shows that the improved YOLOv7-tiny strikes a good balance between high-precision target detection and speed. This algorithm maintains fast detection speeds while ensuring high precision, making it highly suitable for railway perimeter intrusion detection scenarios where real-time performance and detection accuracy are crucial.

5.3 Ablation experiment

The four improvement methods proposed in this paper are S(SimAM), B(BiFPN), G(GhostNetV2). In order to verify the effectiveness of different improvement methods, four improvement methods are used to improve YOLOv7-tiny, and the four improvement methods are added together on YOLOv7-tiny for comparison, where $\surd$ indicates the use of methods. As can be seen from Table 5, all the four methods have improved the accuracy of YOLOv7-tiny to varying degrees. Although SimAM and BiFPN have slightly improved the original detection accuracy, the improved algorithm jointly applied to YOLOv7-tiny by the four improvements has the most significant effect, increasing by 3.3 percentage points. In the table below, we have prominently displayed the exemplary performance data of SBG-YOLO, which has been progressively refined through continuous improvements, achieving high accuracy and low computational cost.

In the process of implementing model lightweighting, reducing the number of parameters and computational load of the YOLOv7-tiny network with the integration of GhostNetV2 modules inevitably sacrifices some accuracy. However, by simultaneously introducing GhostNetV2 modules in both the Backbone and Neck, the computational load is significantly reduced without changing the number of parameters, leading to a notable increase in accuracy by 2.2%. Additionally, to further enhance the detection accuracy, the BiFPN cross-regional feature fusion module is incorporated in the feature fusion layer, and the SimAM attention mechanism is added before the detection head. This combination increases the weight of neurons corresponding to key targets, resulting in an additional 1.1% improvement in detection accuracy. Compared with the original network, the improved network reduces the calculation amount by 61.2%, and the mAP value increases by 3.3%, indicating that the network model greatly increases the detection accuracy of railway foreign objects while being lightweight.

5.4 Embedded platform test results

After training, we obtained the required weight, loaded the weight into improved YOLOv7-tiny network, and migrated the entire network model to the embedded platform for testing

Table 5 Comparison of evaluation indicators of various improved algorithms

Methods	S	B	G	Parameters/MB	mAP@0.5	FLOPS	FPS
YOLOv7-tiny				6.044	74.2%	13.3G	117.6
S	$\surd$			6.044	74.2%	13.3G	102.0
G		$\surd$		6.619	76.0%	5.4G	126.2
BG		$\surd$	$\surd$	6.208	76.4%	5.4G	126.1
SBG-YOLO	$\surd$	$\surd$	$\surd$	**6.208**	**77.5%**	**5.4G**	**122.0**

Table 6 Comparison of deployment performance

Model	Parameters	FLOPS	Inference time	Image	Video
YOLOv7-tiny	6.044M	13.3G	37.4s	26.7	24.3
SBG-YOLO	**6.208M**	**5.4G**	**35.8s**	**28.1**	**25.8**

applications. Parameters, FLOPs, FPS, inference time, and video detection speed of the model were compared. The specific data are shown in Table 6, where the bolded data points highlight the exceptional performance of SBG-YOLO when deployed on the embedded platform, showcasing its ability to process frames at a high rate, making it well-suited for real-time applications. As shown in Table 6, compared with YOLOv7-tiny, improved YOLOv7-tiny model reduces the amount of FLOPS by 61.2%, which reduces the difficulty of deploying the model on the embedded side with limited computing power, and the video detection reaches 25.8 FPS, which meets the needs of real-time target detection.

5.5 Example of railway foreign object intrusion detection result

To validate the detection performance of the improved YOLOv7-tiny model in the field of railway perimeter intrusion

Fig. 12 Example of railway foreign object intrusion detection results

detection, this study selected some representative images of railway perimeter intrusion from the test set as input data to be sent to the improved YOLOv7-tiny network model for testing. Figure 12 shows the results of this comparison, with the left side displaying the results of the original YOLOv7-tiny model and the right side displaying the results of the improved YOLOv7-tiny model. As shown in Fig. 12, the improved YOLOv7-tiny model has achieved an increase in detection accuracy, with the confidence of almost all detected objects being improved. Furthermore, the improved model effectively addressed the issue of missing detection of some people and vehicles in the first and third images of the original network, ensuring the completeness of the detection results. Additionally, the improved model demonstrates higher precision in locating intrusion targets, and the alignment between detected targets and actual situations is further improved, reflecting the enhanced adaptability and reliability of the improved model to real-world application scenarios.

5.6 Recognition results of foreign bodies in railway tracks

The enhanced YOLOv7-tiny model demonstrates significant improvements in detecting a variety of foreign objects on railway tracks. As illustrated in Fig. 13, the model showcases its recognition accuracy and the types of objects it can detect. In Fig. 13a, the model accurately identifies foreign objects such as sheep and humans in a complex and chaotic environment. Figure 13b, g highlights the model's ability to recognize cars encroaching on open railway tracks, a common occurrence in such settings. Figure 13c, h, i demonstrates that the model can detect individuals riding motorcycles or bicycles, showcasing its versatility.

Figure 13d, e shows the model's detection of common domestic pets like cats and dogs, which frequently appear on residential railway tracks. Figure 13f emphasizes the model's critical role in identifying fallen rocks on the

Fig. 13 Recognition results of foreign bodies in railway tracks

tracks, a significant cause of train derailments. Overall, the results illustrate the powerful performance of the improved YOLOv7-tiny model, capable of accurately detecting a wide range of foreign objects, including cars, humans, fallen rocks, cats, dogs, sheep, bicycles, and motorcycles. The model can make precise identifications even in complex environments with multiple encroachments, proving its value in maintaining railway safety.

5.7 Analysis of testing results based on improved YOLOv7-tiny

Given the complexity of the railway operating environment, this section primarily examines the reliability of the improved YOLOv7-tiny model under variable weather conditions and complex railway backgrounds. By testing intrusion images in various railway perimeter scenarios, the detection results are shown in Fig. 14. The improved YOLOv7-tiny demonstrates excellent performance in adapting to different environmental changes. Whether under poor night visibility, adverse rain, snow, foggy weather, or in complex railway crossings and bridge environments, the improved model maintains stable detection performance, with detection confidence generally above 0.85. Additionally, the improved algorithm shows high accuracy and low false alarm rates when handling multi-object complex railway scenarios, without missing or falsely detecting targets. In summary, the improved YOLOv7-tiny algorithm, due to its strong environmental adaptability, becomes an ideal choice for foreign object intrusion detection on railway perimeters.

6 Conclusion

In the current railway environments, such as subways and high-speed railways, due to the limited reaction time of drivers during high-speed operation, the entire line is

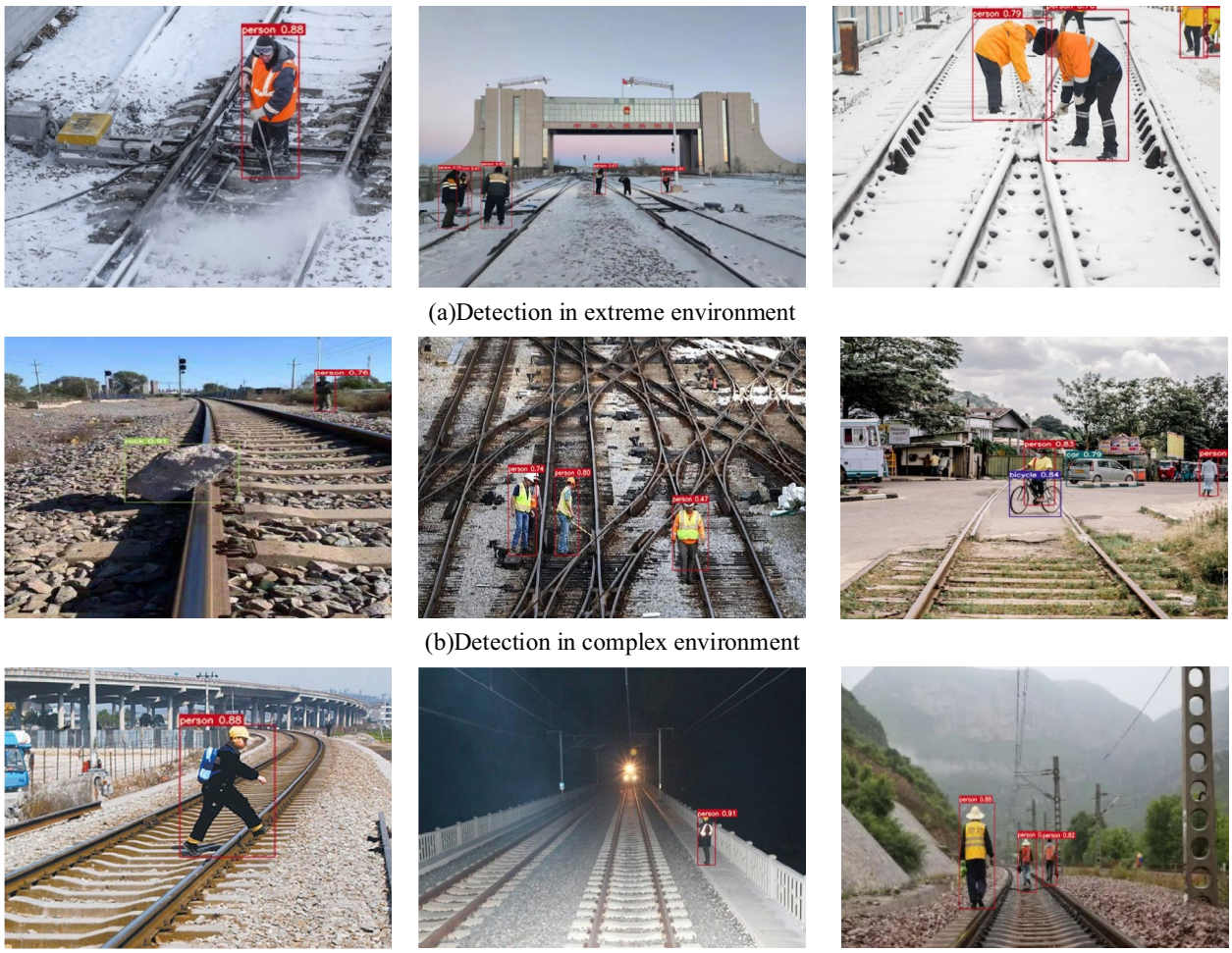


Fig. 14 Target detection of railway foreign bodies in various environments

planned as a closed railway. However, since coal transport railways are long and operate at relatively low speeds, locomotive drivers have enough reaction time. To reduce unnecessary costs, most of the lines are in an open environment with complex and diverse backgrounds, often accompanied by extreme weather conditions. To address these issues and ensure safe and stable railway locomotive operation, a lightweight and high-precision recognition algorithm is proposed to tackle the low detection accuracy of foreign objects invading the clearance in complex open railway environments, especially under extreme weather conditions, and to improve mobile deployment performance. Firstly, the SimAM attention mechanism is introduced into the model to provide higher neuron weights for foreign object targets in complex environments within the detection head module, thereby improving detection performance. Secondly, inspired by the BiFPN network structure, an original feature path is introduced to achieve cross-region feature fusion, further enhancing detection accuracy. Finally, to meet mobile deployment requirements, the ELAN-S module is improved to the GhostNetV2 module, reducing model computation and enhancing deployment performance. Moreover, the combination of Focal_loss and CIoU_loss is replaced with the NWD loss function, which improves the network model's detection of distant, blurry small targets.

Experimental results show that our improved algorithm increases detection accuracy to 77.5% with a slight increase in the number of parameters and a significant reduction in computation. The loss function converges faster, robustness is improved, and it meets real-time requirements at a lower computational cost. Compared to other mainstream models, our improved algorithm is more suitable for railway foreign object detection in complex backgrounds.

In future work, binocular vision technology [28] can become the exploration direction to realize three-dimensional monitoring of locomotive front and rear, further improve the efficiency of model detection, and ensure the safety of locomotive running. Additionally, continuous learning technology [29] can be added to accumulate large datasets during locomotive operation to improve detection performance. Simultaneously, we can explore lightweight optimization of network structure parameters [30] while reducing computational load, aiming to achieve better performance on embedded systems. Our goal is to ensure the model remains lightweight while accelerating inference speed and enhancing detection accuracy, maximizing performance under limited resources. Ultimately, we will further optimize these technologies to ensure the safety of locomotive operations in open railway environments.

Author contributions B. Z.; Data analysis and Writing, Conceptualization. Q. Y.; Conceptualization, Resources. F. C.; Validation, writing-review and editing. D. G.; Methodology, Visualization. All authors have read and agreed to the published version of the manuscript.

Funding The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China (U1806201), the Shandong Province Natural Science Foundation of China (ZR2022ME194) and the Major Basic Research Project of Shandong Province Natural Science Foundation (ZR2021ZD12).

Availability of data and materials The data that support the findings of this study are available on request from the corresponding author upon reasonable request. No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval The type of research in this paper does not involve ethical issues.

Informed consent Informed consent was obtained from all authors for the publication of this article.

References

1. Bešinović, N.: Resilience in railway transport systems: a literature review and research agenda. Transp. Rev. **40**, 457–478 (2020)
2. Liljenström, C., Björklund, A., Toller, S.: Including maintenance in life cycle assessment of road and rail infrastructure—a literature review. Int. J. Life Cycle Assess. **27**, 316–341 (2022)
3. Liu, W., Anguelov, D., Erhan, D.: Ssd: Single shot multibox detector. arXiv preprint arXiv:1512.02325 (2016)
4. Lin, T., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. IEEE Trans. **42**, 318–327 (2020)
5. Redmon, J., Farhadi, A.: YOLO9000: Better, Faster, Stronger. IEEE Conference on Computer Vision and Pattern Recognition. 6517–6525 (2017)
6. Redmon, J., Farhadi, A.: YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
7. Xu, Y., Tao, H., Hu, L.: Railway foreign body intrusion detection based on faster R-CNN network model. J. China Railw. Soc. **42**, 91–98 (2020)
8. Wang, H., Jiang, Z., Wu, Y.: Fast detection algorithm of railway clearance based on deep learning. J. Railw. Sci. Eng. **21**, 2086–2098 (2024)
9. Meng, C., Wang, Z., Shi, L., Gao, Y., Tao, Y., Wei, L.: SDRC-YOLO: a novel foreign object intrusion detection algorithm in railway scenarios. Electronics **12**, 1256 (2023)
10. Wang, S., Wang, Y., Chang, Y., Zhao, R., She, Y.: EBSE-YOLO: High precision recognition algorithm for small target foreign object detection. IEEE Access. **11**, 57951–57964 (2023)
11. Wang, C., Bochkovskiy, A., Liao, H.: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 7464–7475 (2023)

12. Zhao, H., Zhang, H., Zhao, Y.: YOLOv7-sea: Object detection of maritime uav images based on improved YOLOv7. Proceedings of the IEEE/CVF winter conference on applications of computer vision. 233–238 (2023)
13. Wei, G., Wan, F., Zhou, W.: BFD-YOLO: a YOLOv7-based detection method for building façade defects. Electronics **12**, 3612 (2023)
14. Chen, Z., Liu, C., Filaretov, V.F., Yukhimets, D.A.: Multi-scale ship detection algorithm based on YOLOv7 for complex scene SAR images. Remote Sensing. **15**(8), 2071 (2023)
15. Chen, J., Liu, H., Zhang, Y., Zhang, D., Ouyang, H., Chen, X.: A multiscale lightweight and efficient model based on YOLOv7: applied to citrus orchard. Plants. **11**(23), 3260 (2022)
16. Mandel, N., Milford, M., Gonzalez, F.: A method for evaluating and selecting suitable hardware for deployment of embedded system on UAVs. Sensors. **20**, 4420 (2020)
17. Yang, L., Zhang, R. Y., Li, L., Xie, X.: Simam: A simple, parameter-free attention module for convolutional neural networks. In: International conference on machine learning, PMLR, pp. 11863–11874 (2021)
18. Shen, L., Dong, Y., Pei, Y., Yang, H., Zheng, L., Ma, J.: One-dimensional feature supervision network for object detection. International Conference on Intelligent Computing. 147–156 (2023)
19. Chen, Z., Tian, S., Yu, L., Zhang, L., Zhang, X.: An object detection network based on YOLOv4 and improved spatial attention mechanism. J. Intell. Fuzzy Syst. **42**(3), 2359–2368 (2022)
20. Yu, C., Shin, Y.: SAR ship detection based on improved YOLOv5 and BiFPN. ICT Express. **10**(1), 28–33 (2024)
21. Tang, Y., Han, K., Guo, J., Xu, C., Xu, C., Wang, Y.: GhostNetv2: enhance cheap operation with long-range attention. Adv. Neural. Inf. Process. Syst. **35**, 9969–9982 (2022)
22. Zhou, J., Jampani, V., Pi, Z., Liu, Q., Yang, M.: Decoupled dynamic filter networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 6647–6656 (2021)
23. Zhang, Y., Lee, C., Hsieh, J., Fan, K.: CSL-YOLO: A new lightweight object detection system for edge computing. arXiv preprint arXiv:2107.04829 (2021)
24. Liu, Y., Xue, J., Li, D., Zhang, W., Chiew, T., Xu, Z.: Image recognition based on lightweight convolutional neural network: recent advances. Image Vis. Comput. (2024). https://doi.org/10.1016/j.imavis.2024.105037
25. Stanisavljević N, Stojanović D, Petrović L.: Open innovation and crowdsourcing: challenges and opportunities for Serbian railways. (2022)
26. Wang, J., Xu, C., Yang, W., Yu, L.: A normalized Gaussian Wasserstein distance for tiny object detection. arXiv preprint arXiv:2110.13389 (2021)
27. Dai, Y., Hu, Z., Zhang, S., Liu, L.: A survey of detection-based video multi-object tracking. Displays (2022). https://doi.org/10.1016/j.displa.2022.102317
28. Ding, J., Yan, Z., We, X.: High-accuracy recognition and localization of moving targets in an indoor environment using binocular stereo vision. ISPRS Int. J. Geo-Inf. **10**(4), 234 (2021)
29. Menezes, A., Moura, G., Alves, C., Carvalho, A.: Continual object detection: a review of definitions, strategies, and challenges. Neural Netw. (2023). https://doi.org/10.1016/j.neunet.2023.01.041
30. Ruan, D., Han, J., Yan, J.: Light convolutional neural network by neural architecture search and model pruning for bearing fault diagnosis and remaining useful life prediction. Sci. Rep. **13**, 5484 (2023). https://doi.org/10.1038/s41598-023-31532-9