



OPEN Enhanced nighttime nail detection using improved YOLOv5 for precision road safety

Haijian Wang¹, Ziliang Hu¹, Han Mo¹ & Xuemei Zhao²✉

To address the challenge of nail recognition and retrieval on roads at night, we present an enhanced nighttime nail detection system leveraging an improved YOLOv5 model. The proposed model integrates modified C3 modules, reparametrized feature pyramid networks (RepGFPN), and an optimal transport assignment loss (OTALoss), significantly boosting recognition accuracy while reducing parameters by 16%. Deployed on an NVIDIA Jetson Orin Nano device with a stereo matching algorithm, the system achieves synchronized recognition and localization of road nails within a 120° field of view, with localization errors maintained within 2.0 cm. Integrated with a binocular vision-based electromagnetic retrieval system and a ring marker system, the complete robot control system achieves retrieval and marking accuracies exceeding 98%. Experimental results demonstrate an average recognition accuracy of 91.5%, outperforming the original YOLOv5 model by 11.3%. This study paves the way for more efficient and accurate road nail removal, enhancing road traffic safety and demonstrating substantial practical value.

Keywords Object detection, Stereo matching, Localization, Retrieval/mark

Due to factors such as construction and dropping, small and inconspicuous nails are often present on roads, posing safety risks to pedestrians and vehicles. Traditional manual road cleaning methods are labor intensive and inefficient. During the day, with heavy traffic on the roads, the road sweeper designed for large, lightweight debris such as leaves, branches, and bottles are often ineffective at cleaning small, heavy nails, especially those partially embedded in the pavement due to vehicles pressure. To avoid disrupting traffic, we choose to conduct inspections on clear nights. Therefore, a robot system is designed to recognize nails on road surface and then attempts to locate and retrieval. If fails, the nails will be marked for further action. The core is to recognize road nails under uneven lighting conditions. Non-deep learning-based image recognition algorithms are sensitive to key features and sizes of objects; hence they struggle to accurately identify road nails in low quality images obtained at night. Deep learning-based image recognition algorithms learn the mapping from images to labels to improve the information utilization of low quality images. However, the end-to-end training process can not deal with occlusion, vague, and similarity between objects especially uneven lighting conditions. Therefore, this paper proposes an improved YOLOv5 algorithm, combined with improved C3 module, RepGFPN and OTALoss, which improves the detection accuracy of the network under uneven lighting conditions from the perspectives of feature extraction, multi-scale fusion and loss. To optimize the deployment of the improved network model on NVIDIA detection device, the network is lightweighted by reducing the number of parameters to enhancing detection speed.

Related work

With the rapid development from traditional technology to deep learning, object detection networks have become faster and achieved higher accuracy. Traditional manual feature extraction object detection algorithms such as Scale Invariant Feature Transform (SIFT)¹, Histogram of Oriented Gradients (HOG)², and Deformable Parts Model (DPM)³, heavily rely on manually designed image features and object contours. Fukushima⁴ first proposed the concept of neural machines, which was the initial implementation of Convolutional Neural Networks (CNN)⁵. Since Hinton first proposed AlexNet⁶ neural network in 2012, object detection has been elevated to a new level. D. Arora and K. Kulkarni used Faster R-CNN to address the difficulty of object recognition on Efficient Shelf⁷. However, two stage object detection models such as R-CNN⁸, Fast R-CNN⁹,

¹Guangxi Key Laboratory of Manufacturing System & Advanced Manufacturing Technology, School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin 541004, Guangxi, China. ²School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, Guangxi, China. ✉email: qingseyuji2010@163.com

Faster R-CNN¹⁰ and others exhibited larger model size and longer training times. Despite their incapacity for end-to-end training, the fundamental ideas, such as the Region Proposal Network (RPN)¹¹, continue to find widespread applications in contemporary object detection models. Nowadays, YOLO series has been widely used in object detection. The initial iterations of the YOLO family—specifically, YOLOv1¹², YOLOv2¹³, and YOLOv3¹⁴—incorporated prior frames, residual networks, feature fusion, multi-scale training techniques and outperform other object detection algorithms of the same period. R. Mi, Z. Hui et al. presents an improved YOLOv3-SPP algorithm with DIOU-NMS Loss and dilated convolution, achieving a 1.79% mAP improvement and reducing missed detections in dense vehicle scenarios¹⁵. The first three versions of YOLO series are relatively complex. YOLOv4¹⁶ optimized the training strategy and data enhancement approaches by using Mosaic data augmentation and introducing SPP¹⁷ structures. However, as the model structure is further expanded and optimized, the demand for computational resources in YOLOv4 increases. YOLOv5¹⁸ integrates Conv, C3, SPPF modules to enhance the network's learning ability. Moreover, it consists of various models' variants with different sizes, significantly expanding the applicability of YOLOv5. G. Ma, Y. Zhou et al.¹⁹ improved detection accuracy of lightweight models in complex backgrounds by adding lightweight convolution to the Neck and using CIOU Loss for regression. T. Jiang, Y. Xian et al. proposed an improved YOLOv5 algorithm for detecting traffic signs in complex environments by integrating SE modules, CoT modules and a small object detection layer²⁰. P. Singh, K. Gupta et al. applied machine vision and deep learning to unmanned aerial vehicles, evaluating the performance of YOLOv5, RetinaNet, and Faster R-CNN in challenging environments²¹. Following YOLOv5, YOLO family has subsequently introduced: YOLOv6²², YOLOv7²³ and YOLOv8²⁴, continuing to explore and innovate based on previous versions. Q. Wang, C. Li et al. integrated the improved YOLOv7 model into substation inspection robots²⁵. However, due to the large model size and the lack of a pre-trained backbone, there were certain difficulties in model deployment. As the latest version of YOLO, YOLOv9²⁶ introduced new concepts to address the various changes required for deep networks to achieve multiple objectives. However, due to the large model size, deployment remained challenging. Besides the YOLO series, many other algorithms have made significant contributions on object detection. RetinaNet²⁷ improves detection performance by addressing the class imbalance issue through the introduction of the focal loss function. EfficientDet²⁸ achieves efficient object detection by improving and extending EfficientNet²⁹.

Among all the above mentioned detection algorithms, balancing speed and accuracy is the most important. In terms of model parameters and FLOPs, YOLOv5 performs better than YOLOv6 and YOLOv8. Balancing detection accuracy and speed, YOLOv5n is currently the most suitable network model for deployment on constrained embedded systems. Despite this, the fixed size of the YOLOv5n model's detection frame poses challenges in its application to objects of varying sizes, it may not entirely fulfill the requirements for high precision detection of road nails.

Robot structure

As shown in Fig. 1, the robot's mechanical structure is divided into four parts: the walking mechanism, the visual inspection mechanism, the electromagnetic retrieval mechanism, and the ring marker mechanism. The first part is walking mechanism. It designed based on the traditional road sweepers, using a four wheel structure to support and drive the entire inspection robot system. This ensures that the inspection robot can operate in various complex road conditions, thereby reducing its maintenance costs. The second part is visual inspection mechanism. To meet the large field of view requirements during the inspection process, a spherical gimbal is used to mount a USB binocular industrial camera. The gimbal adjusts the angle and direction of the binocular camera, enabling full-range camera movement.

The third part is electromagnetic retrieval mechanism. As shown in Fig. 2a, the device consists of a guiding block, gears, timing pulleys, and a magnetic block. The gears and motor make the device move horizontally along the guide rail. The timing pulleys and guiding block combine to adjust the vertical height. The magnetic block can retrieval road nails by switching its power on and off. The fourth part is ring marker mechanism. As shown in Fig. 2b, the mechanism consists slider, sleeve, ferrule sleeve, and ejector wheel. It connected to the electromagnetic retrieval mechanism via an optical axis, enabling synchronized movement of ring marker and electromagnetic retrieval device. Fluorescent markers are placed in the sleeve. The ejector wheel ejects the

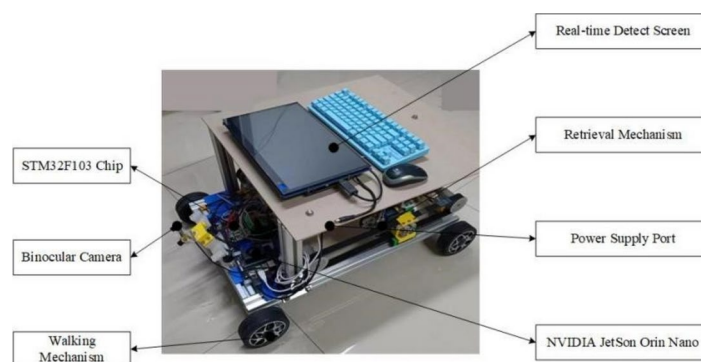


Fig. 1. Mechanical structure of the robot.

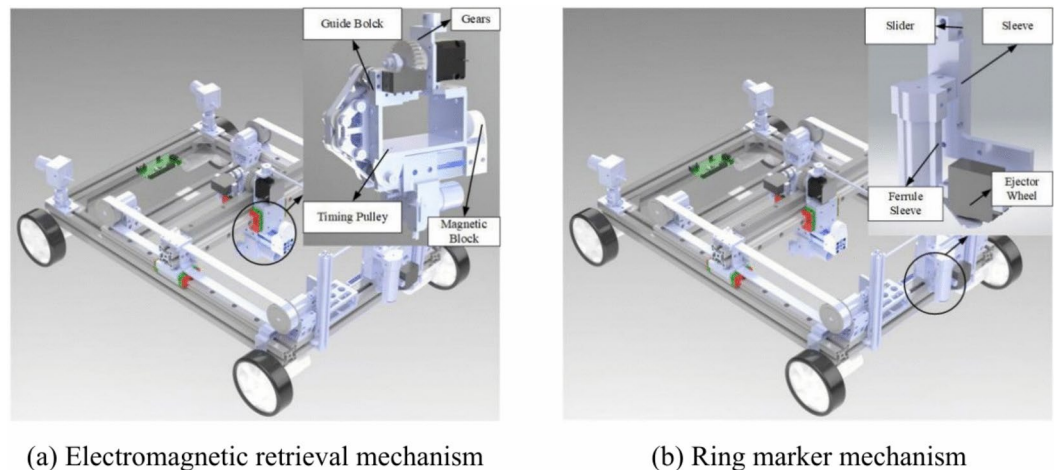


Fig. 2. Partial device mechanical structure.

bottom baffle of the sleeve, causing the fluorescent marker to drop onto the nails. Fluorescent markers are easier to detect in the dark with uneven lighting conditions, making them easier to process later.

The work process of the robot is shown in Fig. 3, we constructed a stereo vision system based on the robot. To ensure that the binocular camera field of view is maximized during the robot's movement, the camera was positioned in the center top of the robot with 45° diagonally downward. During the robot's inspection process, collected the images using stereo camera to construct a training and testing datasets for nighttime road nails. Considering the characteristics of the datasets, we designed the improved YOLOv5n network and carried out subsequent processing. The combined algorithm was deployed on NVIDIA Jetson Orin Nano device to provide real time localization of road nails to electromagnetic retrieval system. The robot was all controlled by STM32F103 chip, through the coordination of various components, it can complete the inspection of road nails.

Algorithms

Nails occupied a small proportion in the images captured by our designed robot. Due to the lighting equipment on the robot, uneven lighting appears in the images, which severely affects the recognition accuracy of existing algorithms. Additionally, in various complex road conditions, nails can easily be obscured by other objects, causing the low accuracy for nighttime detection. To enhance detailed information learning ability of the network, we improved the original model and lightweighted, willingly to enhance the recognition accuracy of road nails while reducing the model's parameters.

C3_improved

C3 module is a key component of the backbone, utilizing a hierarchical structure to extract features. It connects low to high level feature maps through a series of convolutional layers and bottleneck structures. However, with the continuous stacking of convolutional layers and bottleneck structures, the information of the road nails which occupies a relatively small part of the original image will be lost, especially in uneven lighting conditions. As shown in Fig. 4, we directly fuse low and high feature maps through residual connections to compensate for the lost detail information. We combined the C3_improved and C3 modules in the backbone to reduce the loss during the layer-by-layer transmission, making the network pays more attention to the detailed information of road studs during the training process.

RepGFPN

Although the backbone structure combining the C3_improved and C3 modules can enhance the learning ability of road nail details during the training process, the model's recognition capability is still lacking. To improve the fusion capability, we incorporated reparametrized feature pyramid network (RepGFPN) into the original structure. As shown in Fig. 4, the improved structure employed multiple sampling operations, using different channel dimensions for feature maps at different scales, integrate high semantic and low spatial features of road nails to assist C3_improved in learning details. Its re-parameterization mechanism enhances the network's feature representation capability by automatically adjusting features across different levels, effectively capturing the detailed information impacted by lighting changes. This improves the model's ability to recognize objects in uneven lighting environments and reduces information loss caused by inconsistent lighting conditions.

OTA loss

In uneven lighting images captured during the robot inspection process, road nails still occupied a small proportion. The stacking of convolutional layers and multiple sampling operations can result in the loss of detailed information, leading to a significant reduction in the usable road nail information for the network. Although newly added C3_improved and RepGFPN can significantly reduce information loss and improve feature fusion methods, they still cannot effectively recover more nail information from the limited data. We

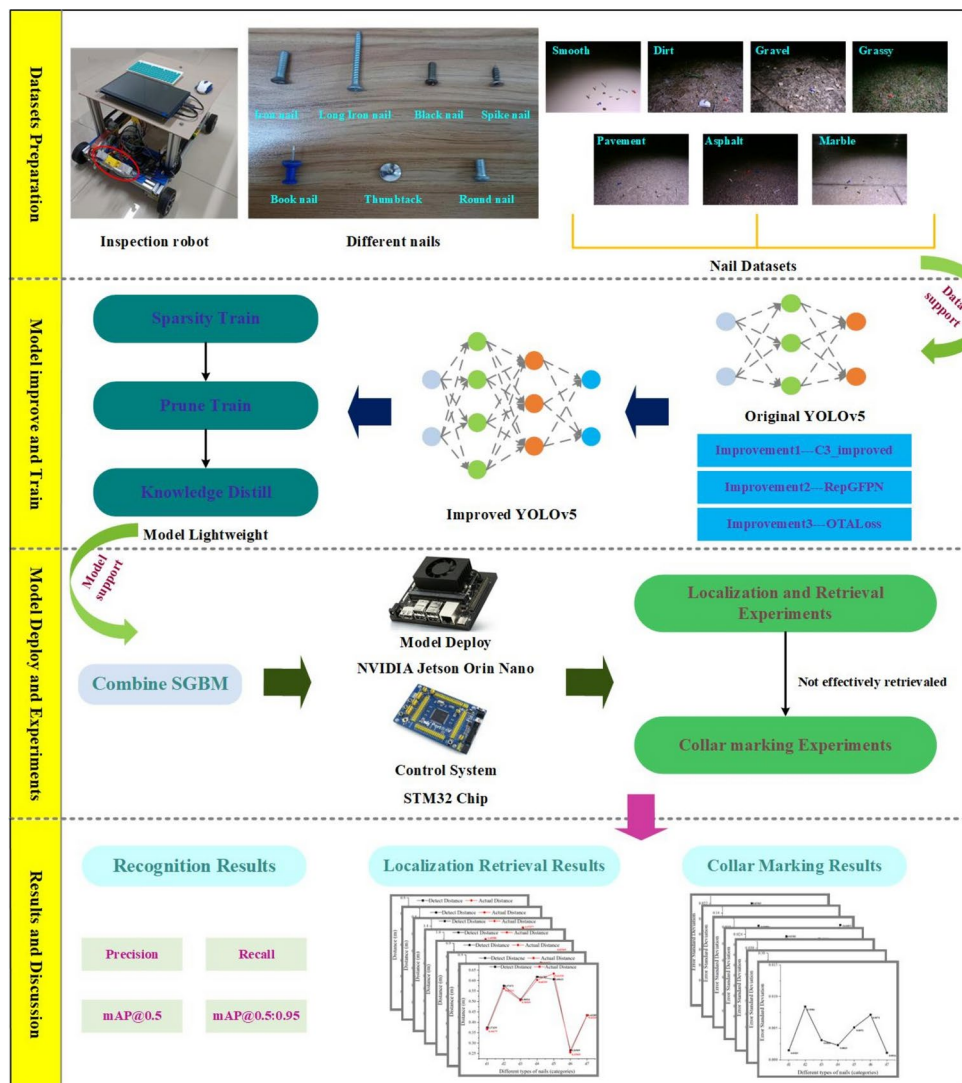


Fig. 3. The work process of the robot.

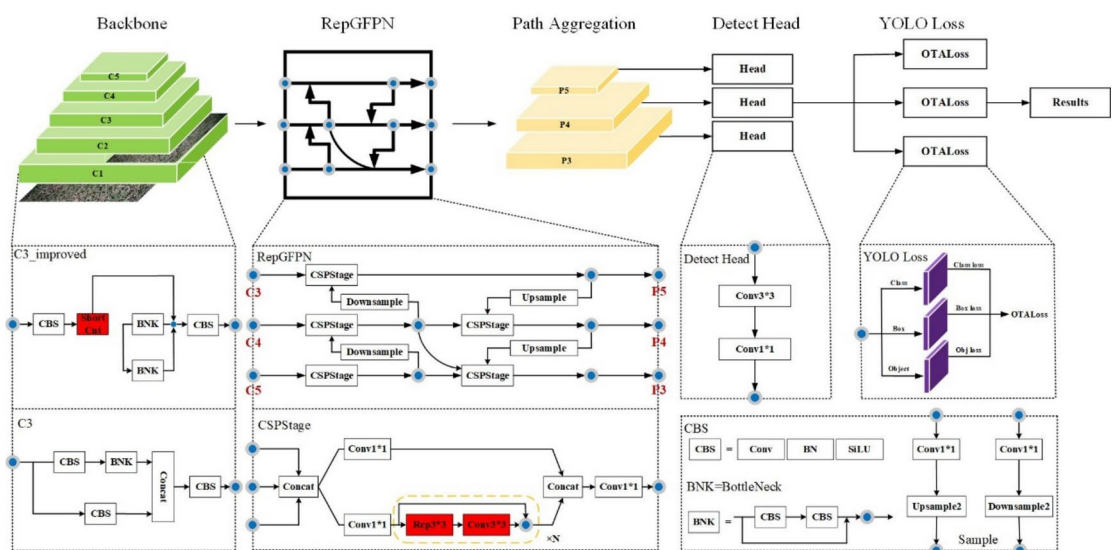


Fig. 4. The improved model structure.

proposed a dynamic label assignment method based on an optimization strategy (OTA) and replace the original loss as shown in Fig. 4. By weighting the classification loss of different nails or considering all of them as the cost of transmission, the network learns the optimal label assignment method by minimizing the cost. The pseudo code of the proposed method is listed in Algorithm 1.

Lightweight

Model lightweight acceleration is a crucial research area in deep learning, focusing on reduce model size and computational complexity to enhance the efficiency of network models on resource constrained devices. The improved object detection network model has more parameters compared to the original model. We achieve model lightweighting by sparsity training, model pruning and knowledge distillation training. These techniques reduce the model size and parameters, resulting in faster inference and improved detection accuracy for road nails.

BN layer is typically placed after the convolutional layer. Its primary role in the network model is to accelerate convergence and reduce the difficulty and complexity of training. The backbone of the network often contains a large number of parameters. In this paper, we lightweight the backbone as shown in Fig. 5.

We introduce a sparsity factor into BN layer and perform sparsity training. During training, the BN layer gradually becomes sparse. After sparsity training, we analyze weights of the BN layer, prune the layers with BN weights approaching zero. We then perform fine-tuning to restore the pruned model's original fit. The fine-tuned model serves as the student model, while the improved model acts as the teacher model. Using knowledge distillation to extract knowledge from the large teacher model and condense it into the smaller student model. By adjusting the distillation weight, we enhance the detection accuracy of the student model, gradually making it reach or exceed the teacher model's accuracy, thereby achieving a lightweight model. The pseudo code of the lightweight method is listed in Algorithm 2.

Experiments

Camera calibration

Camera calibration is an important technique in computer vision which used to determine the internal and external parameters of a camera. This process enables the accurate conversion of pixel coordinates in an image to world coordinates. For the road nails recognition application, the stereo camera is calibrated by Zhang's calibration method. The calibration summary are as follows:

- (1) A calibration board with 8×11 corner points and each square have a size of 19 mm is selected.
- (2) As shown in Fig. 6, maintain the stability of the stereo camera and take a few images of calibration board from different orientations and angles.
- (3) Use binocular calibration algorithm to detect the feature points of our calibration board.
- (4) Estimate and refine the intrinsic and the extrinsic parameters.

The intrinsic matrix, extrinsic matrix, and distortion coefficients can be calculated from each image. The extrinsic matrix includes the rotation matrix and translation vector. For a stereo camera, the distortion coefficients include radial distortion k and tangential distortion p . The parameters and results obtained from the calibration of the stereo camera are shown in Table 1, and Fig. 7a,b below.

Datasets

In this experiment, as shown in Fig. 8, we choose seven different types of nails named d1 to d7. These seven types of nails include long nails, iron nails, sharp nails, thumbtacks, and push pins. To simulate real road scenes, seven road surfaces: smooth road, marble road, pavement road, dirt road, gravel road, asphalt road and grassy road are selected, and several factors such as, shadow occlusion, puddles, obstructions, overlapping nails, similar objects influence are also considered. We collected a total of 2100 images of road nails by the binocular camera to construct a dataset for YOLOv5n network. This dataset was annotated and cross validated by multiple experts using LabelMe a to ensure the accuracy of the annotated data. 75% of the images were randomly selected as the training set and the remaining 25% were used for test. The datasets collection process and some samples of the dataset are shown in Fig. 9.

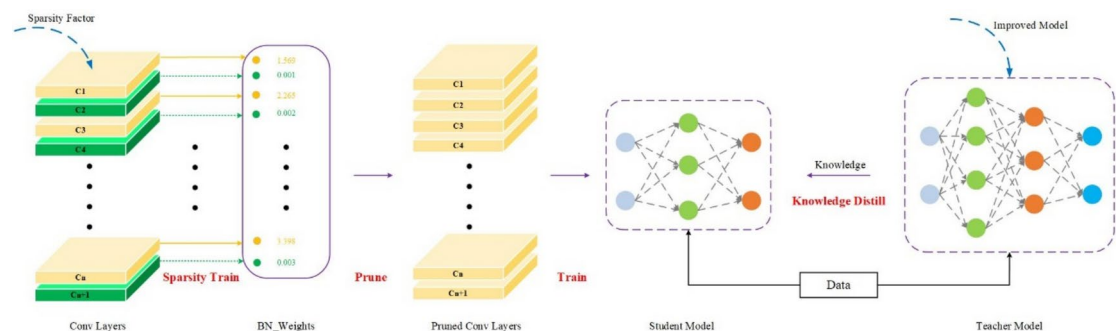


Fig. 5. Lightweight design.

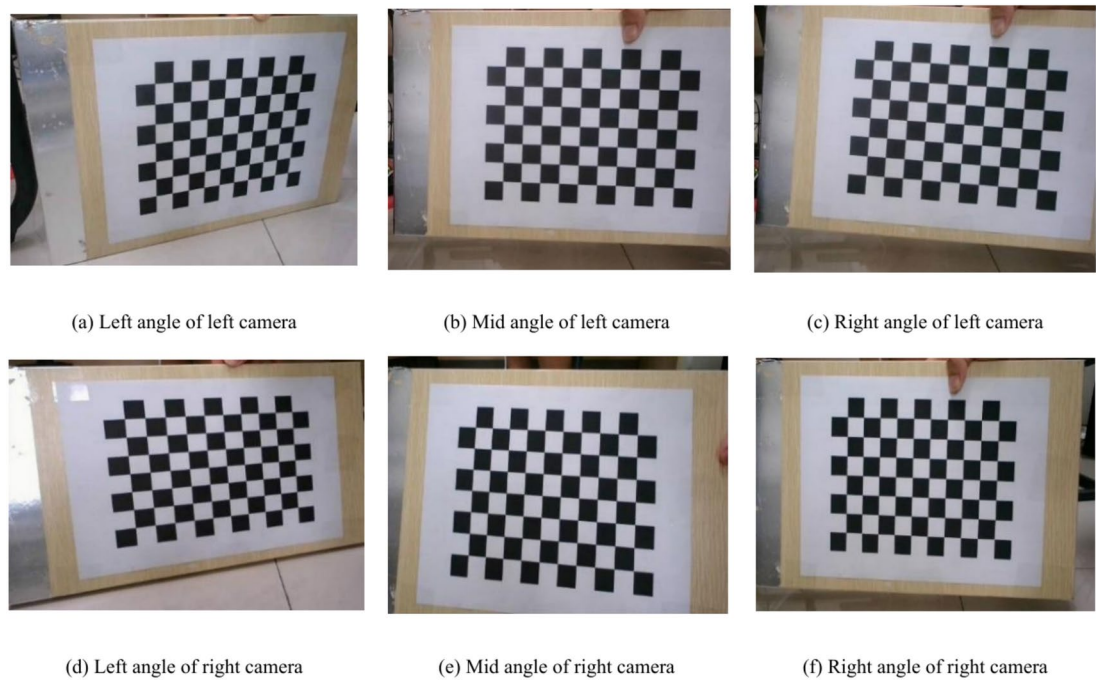


Fig. 6. The calibration process.

Accessories	Parameters
Intrinsic matrix	$\begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 5.1896e+02 & 1.3809 & 3.1464e+02 \\ 0 & 5.1593e+02 & 2.4813e+02 \\ 0 & 0 & 1 \end{bmatrix}$
	$\begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 5.2218e+02 & 0.4667 & 3.1215e+02 \\ 0 & 5.1849e+02 & 2.4516e+02 \\ 0 & 0 & 1 \end{bmatrix}$
Extrinsic matrix	$R = \begin{bmatrix} 0.9999 & -0.4041 & 0.0028 \\ 4.0725 & 0.9999 & -0.0011 \\ -0.0028 & 0.0011 & 0.9999 \end{bmatrix}$
	$T = \begin{bmatrix} -60.5415 & -0.0814 & 0.8879 \end{bmatrix}$
Distortion coefficients	$\begin{bmatrix} k_1 & k_2 & k_3 & p_1 & p_2 \end{bmatrix} = \begin{bmatrix} -0.0289 & 0.0817 & -8.3891e-04 & 8.4552e-04 & 0 \end{bmatrix}$
	$\begin{bmatrix} k_1 & k_2 & k_3 & p_1 & p_2 \end{bmatrix} = \begin{bmatrix} -0.0332 & 0.0874 & 3.0961e-04 & 5.8494e-04 & 0 \end{bmatrix}$

Table 1. Binocular camera parameters.

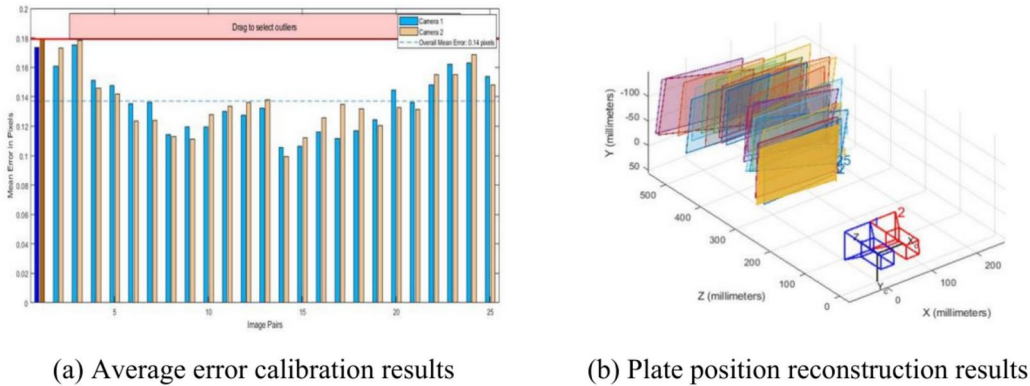


Fig. 7. Calibration results.

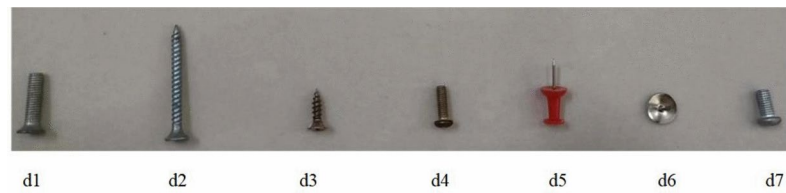


Fig. 8. Different types of nails.



Fig. 9. Datasets collection process and a partial sample of the datasets.

Accessories	Parameters
Operate system	64-bit Linux
CPU	11th Gen Intel(R) Core(TM) i9-11900K
GPU	NVIDIA GeForce RTX 3060
CUDA/CUDNN	12.2/8.0.4
Epochs	400
Batch size	16
Work process	8
Baseline model	YOLOv5n
Hardware platform	NVIDIA Jetson Orin Nano

Table 2. Experimental environment.



Fig. 10. A partial of the cropped datasets.

Results analysis

To verify the effectiveness of the model designed in this paper, we established a training platform and deployed the trained model for validation. The experimental environment settings for model training and deployment are detailed in Table 2. Unless otherwise specified, the baseline model remains unchanged.

Improved experiments results

To avoid ineffective recognition caused by large black areas in the original images collected at night, we cropped the original images from 640 × 480 to 640 × 300. The cropped image datasets are shown in Fig. 10. The comparative experimental results are shown in Fig. 11.

From the results, we found that due to the small proportion of uneven lighting in the cropped images, the precision, recall, and mAP of the network model have been significantly improved, with the improvement generally ranging from 6 to 11%. From the perspective of curve trends, after 100 epochs, the curve is higher than the original image, further proving the effectiveness of image cropping in reducing the impact of uneven lighting on model detect accuracy. To further evaluate the contribution of our improvement to reducing the impact of

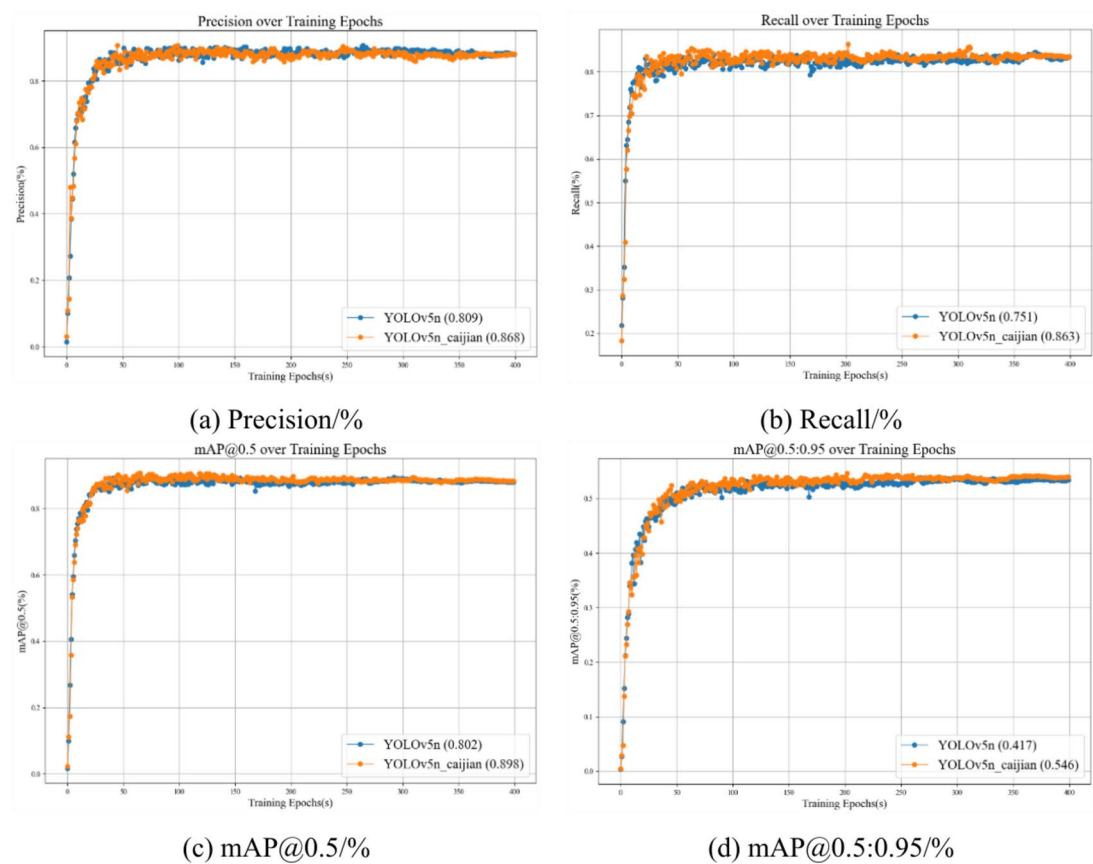


Fig. 11. Comparison experimental results.

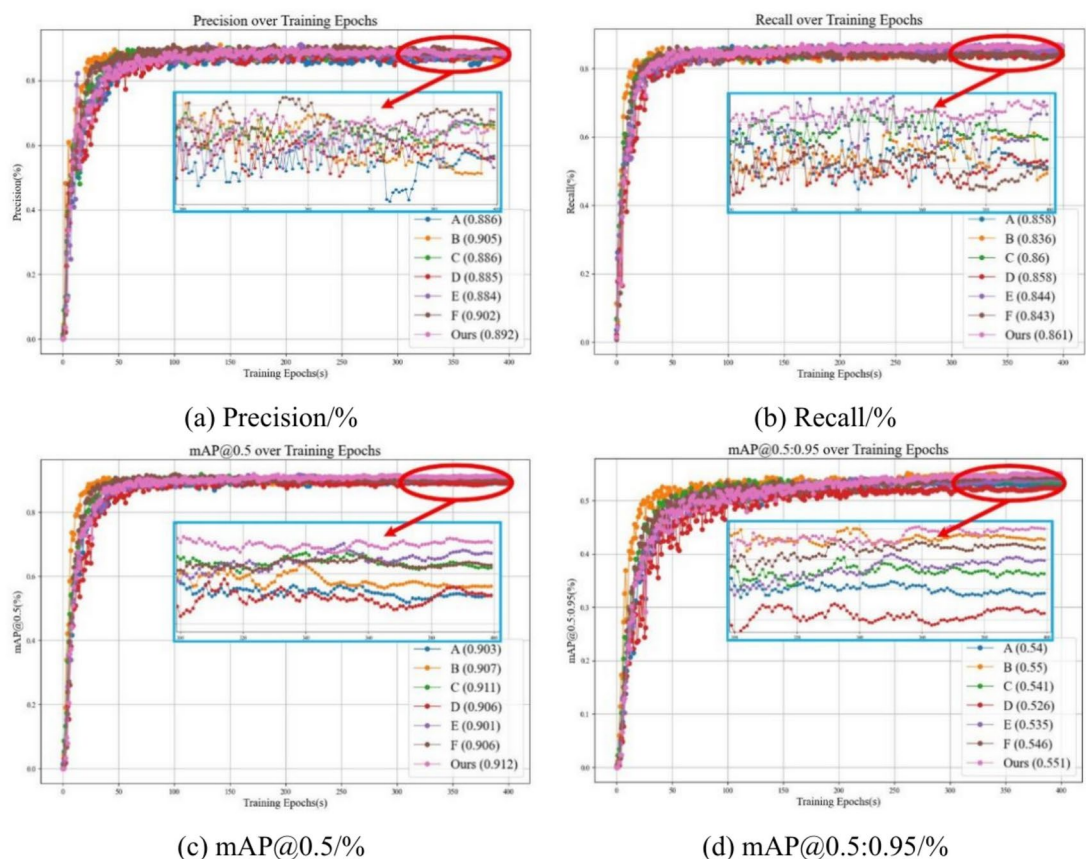
Different improved			Name
C3_improved	OTA Loss	RepGFPN	
√	×	×	A
×	√	×	B
×	×	√	C
√	√	×	D
√	×	√	E
×	√	√	F
√	√	√	Ours

Table 3. Different improved name.

the dark environment on the model effect, we designed ablation experiments. In the subsequent training and evaluation, we used the cropped datasets. Different improved names are shown in Table 3, and the results are shown in Table 4, Figs. 11 and 12.

As the results shown in Table 4, the improved C3 module, in combination with the RepGFPN through residual connections, enables the network to better fuse and share information from multi-level feature maps in uneven lighting conditions. This allows the network to capture features of small objects such as nails across multiple scales, enhancing detection accuracy. Meanwhile, the proposed loss function, with its object aware mechanism, dynamically adjusts the loss for different objects during training, allowing the model to focus more on small objects. Further improved the network's Recall and mAP. To further analyze the detection performance of the proposed model, we enlarged the region between 300 and 400 epochs in Fig. 12. By examining the enlarged region, we found that the recall and mAP curves of our proposed model are significantly higher than those of the other models. However, the simultaneous integration of three distinct improvements in the network leads to an overemphasis on detailed image features under uneven lighting conditions, as well as adaptation inaccuracies during multi-level feature fusion, resulting in errors between the predicted boxes and true boxes. Consequently, the model's detection accuracy is slightly lower than that of Model B and F, and the box_loss is higher than Model B, but the slight box_loss has little effect on the recognition accuracy. Training loss is also a

Models	Precision%	Recall%	mAP@0.5%	mAP@0.5:0.95%	box_loss	Params
A	0.886	0.858	0.903	0.540	0.023	2.33M
B	0.905	0.836	0.907	0.550	0.021	1.78M
C	0.886	0.86	0.911	0.541	0.023	2.33M
D	0.885	0.858	0.906	0.526	0.024	2.34M
E	0.884	0.844	0.901	0.535	0.023	2.63M
F	0.902	0.843	0.906	0.546	0.024	2.32M
Ours	0.892	0.861	0.912	0.551	0.022	2.32M

Table 4. Ablation experiments results.**Fig. 12.** Ablation experiments results.

direct criterion for evaluating the performance of a model. The training loss of the model is composed of three weighted components: classification loss, object loss, and bounding box loss. As shown in Fig. 13, even though the network complexity increases after adding improved C3 module and RepGFPN, the loss is still significantly lower than other models, which is only about 0.05.

Lightweight experiments results

In this subsection, we called the model CRO-YOLOv5n. First, we sparsified and added the regularized loss of BN parameters to the model loss function forces the BN parameter to converge towards 0 during training. After sparsify trained, we cut off the channel corresponding the BN parameters closest to 0. Parameters we set during the training process are shown in Table 5, the results obtained are shown in Fig. 14.

BN_weights histogram can visualize the process of sparsification training, the vertical axis of the histogram represents the number of training sessions, the number of training sessions increases from top to bottom. As shown in Fig. 14, with the training rounds increases, horizontal axis of histogram's training peak is constantly approaching 0, represents that most of the bn have become sparse. The purple curve indicates the peak is approaching the X-axis with a smoother process. In the process of BN sparsity, both mAP@0.5 and mAP@0.5:0.95 shows improvement compared to CRO-YOLOv5n, indicates the promising training results.

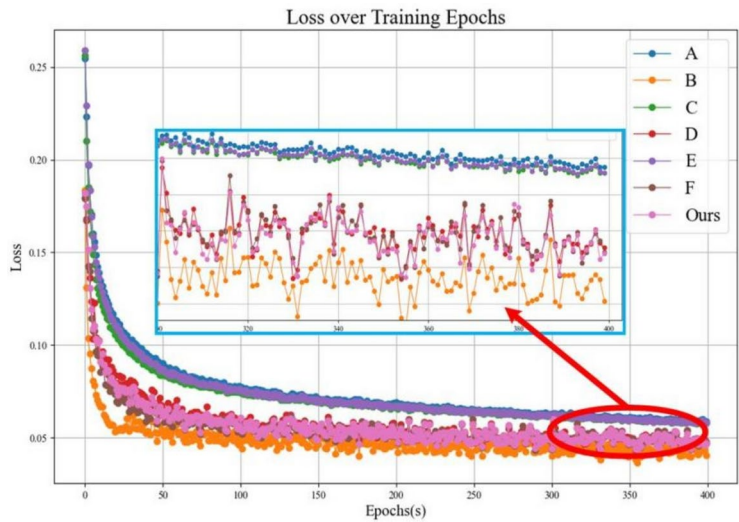


Fig. 13. Comparison of training loss.

Parameters	Value
St	True
Sr	0.0002
Epochs	200
Work process	16
Batch size	8
Adam	True

Table 5. Parameters set during sparsify train.

Then, we pruned the backbone. In this paper, we designed different pruned rates for comparative experiments. Parameters we set during training process are shown in Table 6, the prune trained results are shown in Table 7 and Fig. 15.

From Fig. 15 we can see that, when pruned rate is 0.2, the detect precision, recall, and mAP all achieved optimal performance. At 0.4 and 0.6, the trained results are similar. However, when the pruned rate is 0.8, the trained results are significantly lower. For model inference, we set batch size to 1, the inference results are shown in Table 7. FPS indicates the number of images processed per-second, while inference time refers to the duration needed to process a single image. With reductions in both model parameters and size, when pruned rate is 0.2, the inference time is 9.4 ms and FPS is 92.611, which are both better than those of the unpruned model. Although there is a slight decrease compared to the optimal pruned rate 0.6, considered the performance of the pruned model, we conclude that pruned rate 0.2 offers the best overall detect performance.

At last, we used spartied.pt as teacher network weight and 0.2pruned.pt as student network weight. Distilled the knowledge from teacher network model and transferred to student network model. Knowledge distillation weights ranging was set from 10 to 100, the results obtained through training are shown in Fig. 16.

As the results shown in Fig. 16, it can be seen that different knowledge distillation weights have a certain impact on results. During training process, the trained model achieved the highest precision as the weight set to 40%, the recall and mAP slightly decrease compared to the others, but remain generally consistent. We choose the 40% distillation weight trained model compared with the original model and the improved model, the comparison results are shown in Fig. 17. We also set batch size to 1, the inference results are shown in Table 8.

As the results shown in Fig. 17, the model got the best precision and mAP after sparse train, prune train and knowledge distillation. As can be seen from Table 8, the lightweight model has 16% fewer parameters compared to the improved model, only 0.18 M more parameters than the original model. Additionally, the computational complexity of the model decreased from 5.2 GFLOPs to 4.5 GFLOPs, resulting an increase in the inference speed. Overall, the lightweight model shows significant advantages in recognizing road nails at nighttime.

Localization retrieval and ring mark experiments

We combined the lightweight model with SGBM and deployed it on NVIDIA Jetson Orin Nano device. This allows for accurate recognition and localization of road nails during robotic inspections. We designed multiple localization experiments for different nails under various conditions, including different times and road surfaces. The experimental results are shown in Fig. 18. In result figures, the horizontal axis indicates the group number,

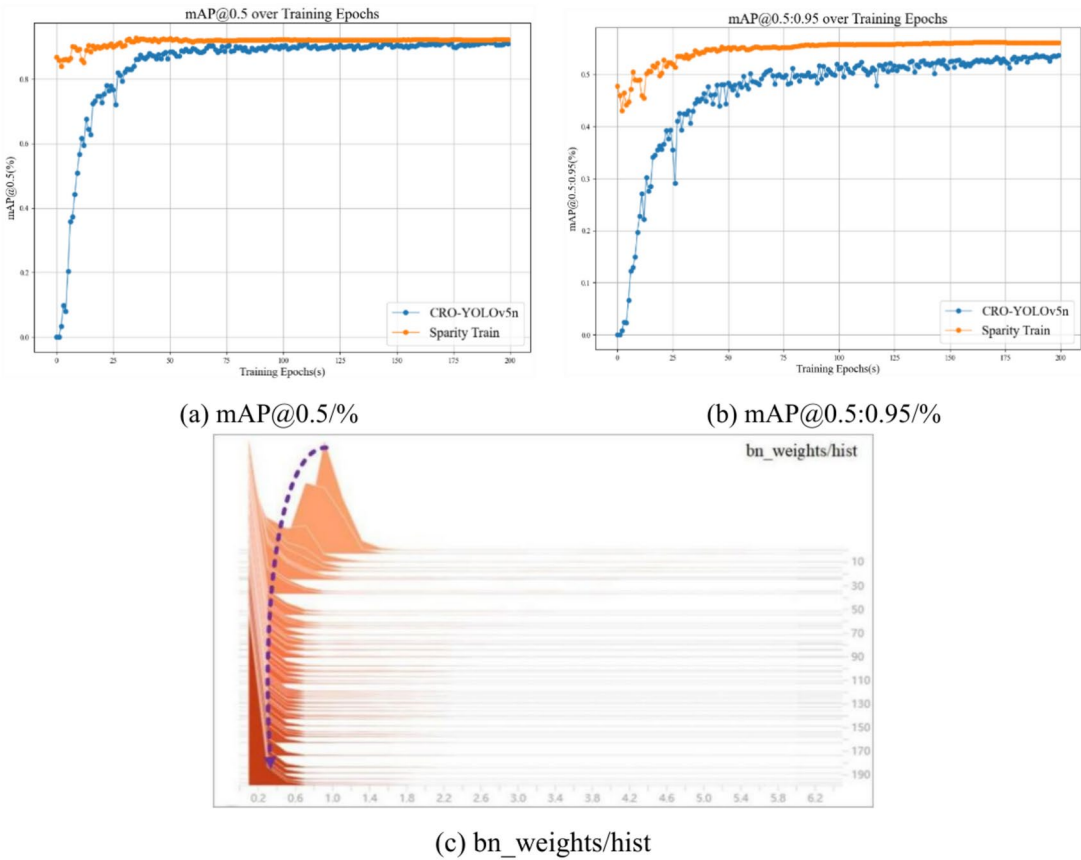


Fig. 14. Sparsify trained results.

Parameters	Value
Weights	Spartied.pt
Percent	0.2/0.4/0.6/0.8
Epochs	300
Work process	4
Batch size	1(Val)/16(Train)
Adam	True

Table 6. Parameters set during pruned train.

Different pruned rates	Params/M	GFLOPs	Model size/MB	Inference/ms	FPS
0.0pruned	2.33	5.2	4.74	9.6	91.867
0.2pruned	1.96	4.5	4.04	9.4	92.611
0.4pruned	1.65	3.8	3.44	9.6	92.241
0.6pruned	1.40	3.2	2.96	9.2	93.821
0.8pruned	1.20	2.7	2.57	9.5	90.976

Table 7. Inference results of the model pruned.

while the vertical axis represents the actual distance to the road nails. The colored dashed lines indicate the range of ± 2 cm. From the above results, it can be seen that the combined algorithm achieves high localization accuracy for different nails across various road surfaces, with errors generally staying within ± 2 cm. Building on the high precision localization, we designed retrieval and ring mark experiments for road nails. We tested the robot's retrieval and ring mark results over 200 inspection cycles. Table 9 lists the nail retrieval results on different road

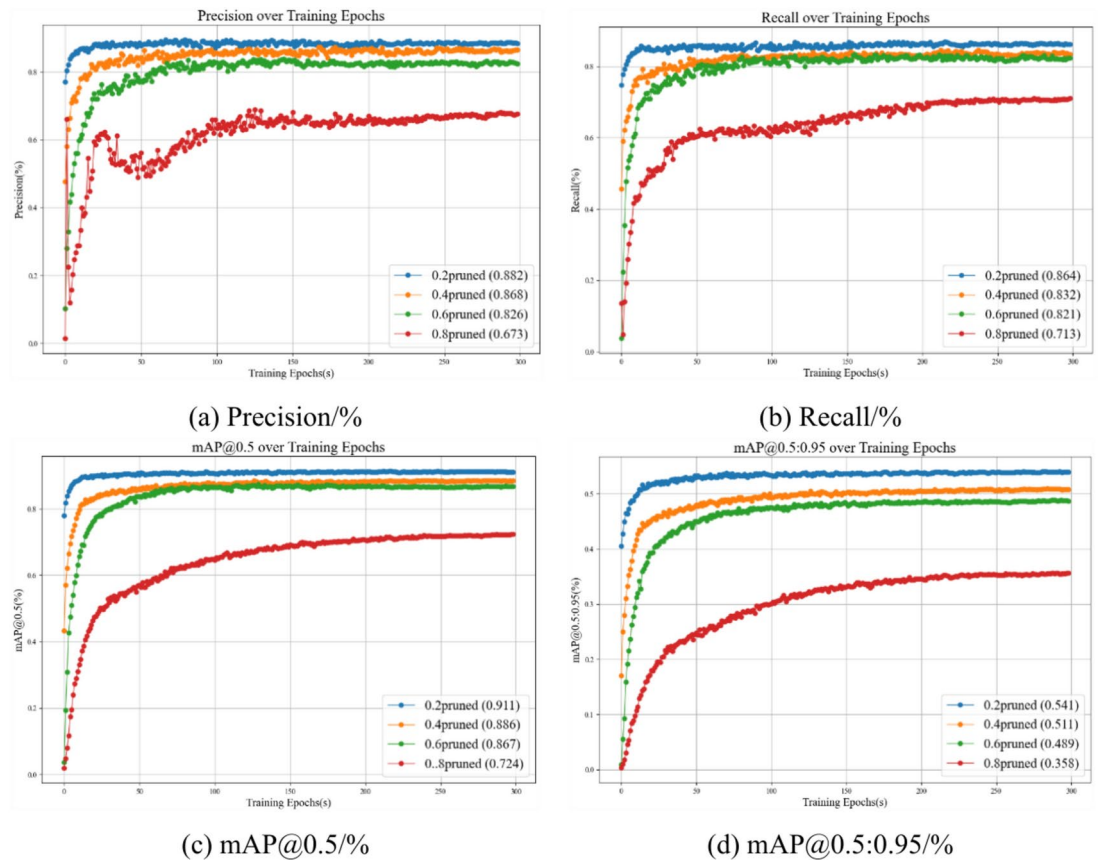


Fig. 15. Prune trained results.

surfaces. Table 10 lists the ring mark results on different road surfaces. Figure 19 shows the retrieval rate for different road surfaces and different road nails.

Combining the retrieval results from Tables 9, 10, we found that the inspection robot achieved a high nail retrieval rate and a high success rate in ring mark on each type of road surface. Similarly, sand and weeds on dirt and grass road reduce the magnetic mechanism's effectiveness, also leading to lower rate. Other road types have minimal impact on the magnetic retrieval system, resulting in higher retrieval success rate. For the same reasons, gravel road, dirt road and grass road also exhibit some instances where nails are not effectively marked with rings. In contrast, other types of road surfaces allow for effective mark. Based on the above results, we can calculate that the retrieval rate for each type of nail and the ring mark rate for each condition is both maintained above 99%. The overall retrieval rate for each condition remains above 98%. Compared to other types of nails, the d7 nail is relatively short and wide, which leads to a higher concentration of iron, making it easier to retrieval. On the other hand, the d3 nail is relatively small in size, making it more challenging. The overall retrieval rates for the remaining types of nails are similar. Thus, the experimental results indicate that the shape of the nails has a certain impact on the retrieval accuracy.

Conclusion

This paper presents a robotic system designed for the localization, retrieval, and ring marking of road nails on nighttime road surfaces. In order to improve the accuracy of road nail recognition at night, we proposed a YOLOv5 object detection algorithm which integrates improved C3, RepGFPN, and OTALoss. By integrating the improved C3 and RepGFPN modules, the network's ability to capture multi-scale nail details in uneven lighting conditions is significantly improved. Additionally, OTALoss is introduced to dynamically adjust the loss for different objects, enabling the network to focus more effectively on small objects. To optimize the deployment of the network, we applied techniques such as sparsification, fine-tuning, and distillation training, reducing the network's parameters. Experimental results demonstrate that, with a 16% reduction in parameters, the network achieves a mAP of 91.5%, marking an 11.3% improvement over the original network. Experiments on seven different road surface show that the nail localization error remains within 2 cm, with the retrieval and ring marking success rate for each type of nails exceeding 99%, and the retrieval success rate for each road type surpassing 98%.

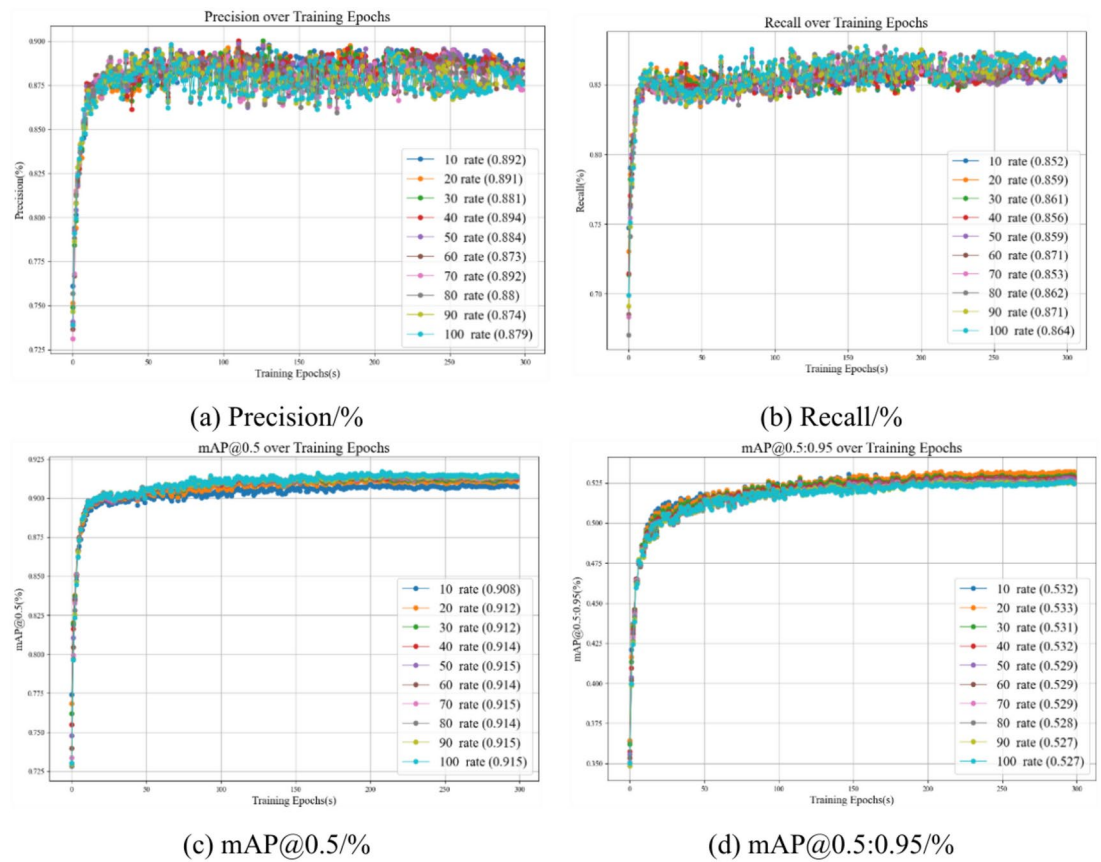


Fig. 16. Distilled train results.

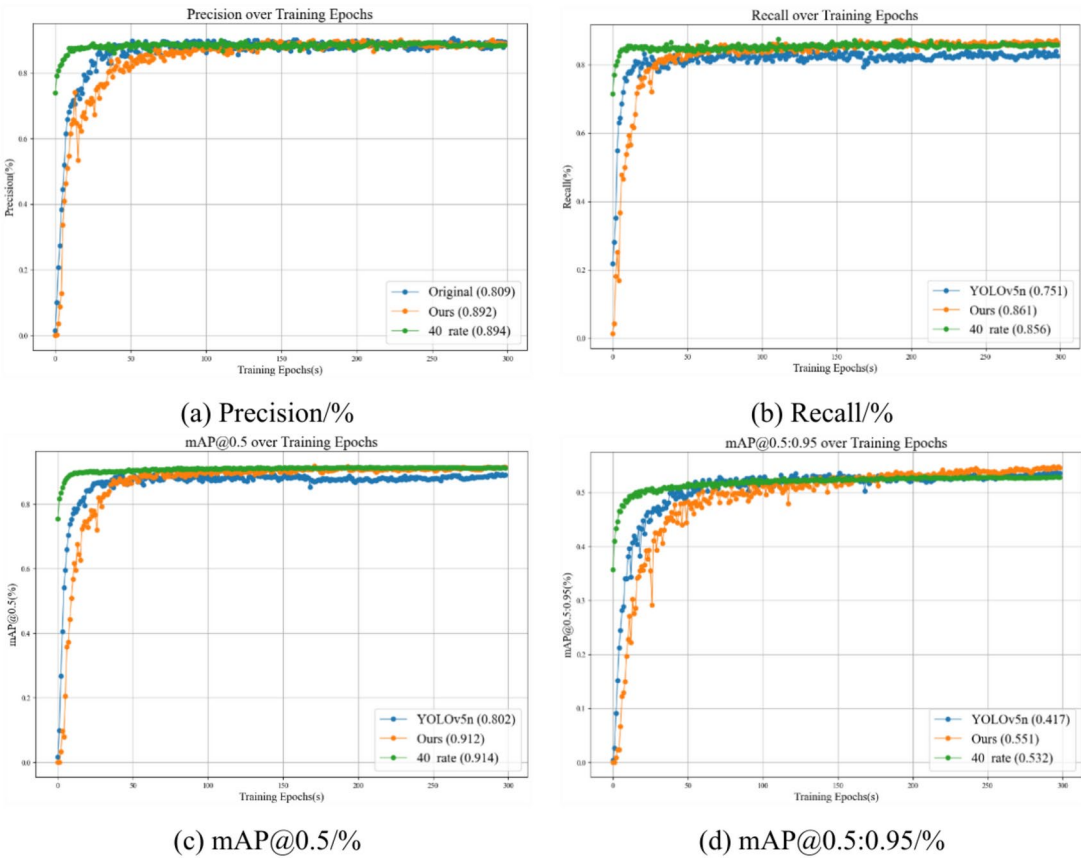


Fig. 17. Trained results of the model comparison.

Models	Params/M	GFLOPs	Model Size/MB	Inference/ms	FPS
Original	1.77	4.2	3.73	6.9	120.002
Ours	2.32	5.2	4.74	9.6	91.867
40 rate	1.95	4.5	4.04	9.4	99.789

Table 8. Inference results of the model comparison.

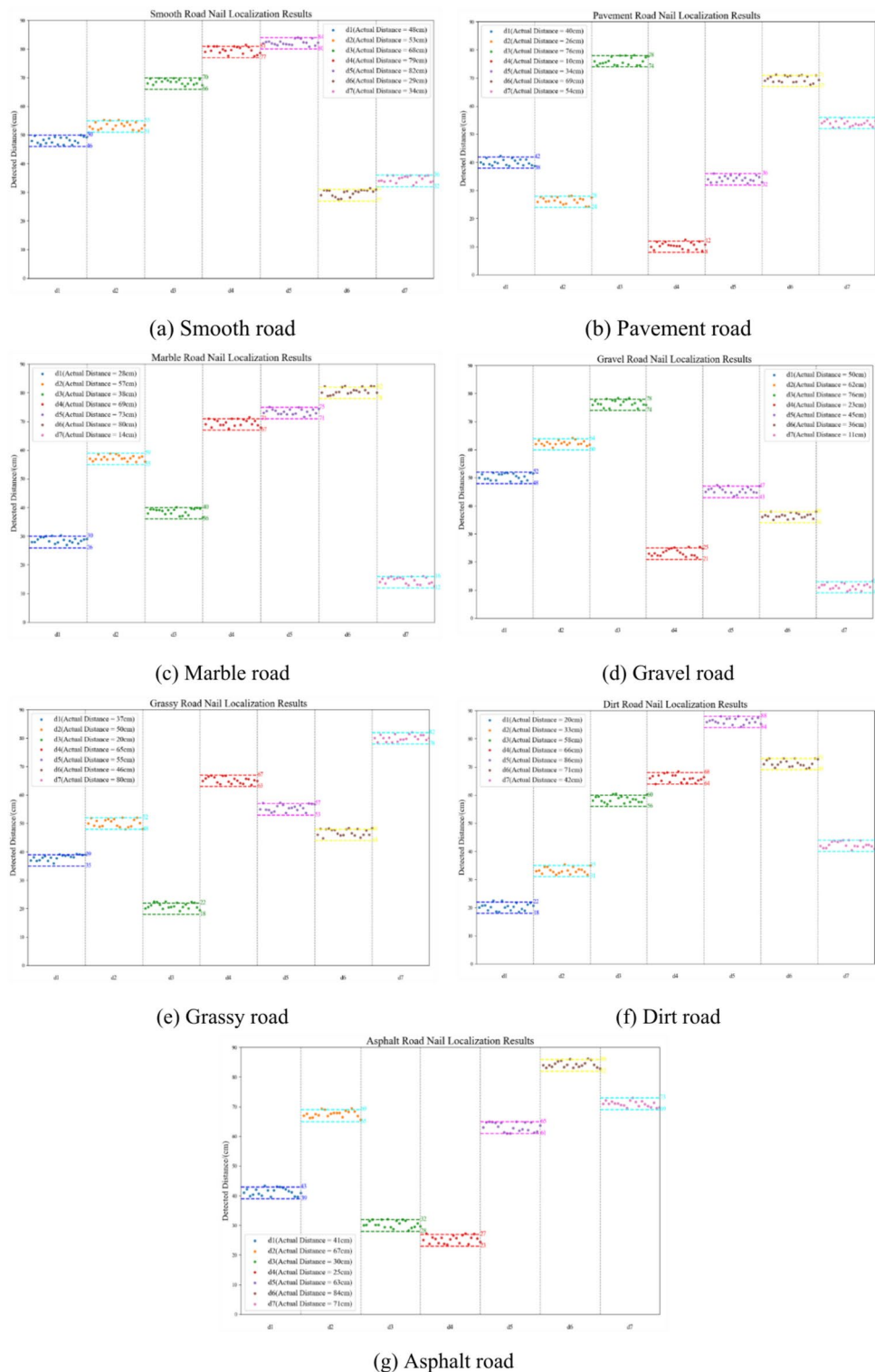


Fig. 18. Nighttime road nail localization results.

Successful retrieval frequency							
Nail type	Smooth road	Marble road	Pavement road	Dirt road	Gravel road	Asphalt road	Grassy road
d1	200	199	200	198	196	200	199
d2	200	200	200	197	197	199	198
d3	200	200	199	196	195	199	197
d4	200	198	200	198	196	200	198
d5	200	200	200	198	200	197	196
d6	200	200	200	197	197	200	197
d7	200	200	200	200	198	200	200

Table 9. Electromagnetic retrieval results.

Successful ring mark frequency							
Nail type	Smooth road	Marble road	Pavement road	Dirt road	Gravel road	Asphalt road	Grassy road
d1	0	1	0	2	4	0	1
d2	0	0	0	2	3	1	2
d3	0	0	1	4	3	1	3
d4	0	2	0	2	3	0	2
d5	0	0	0	2	0	3	3
d6	0	0	0	3	3	0	3
d7	0	0	0	0	2	0	0

Table 10. Ring mark results.

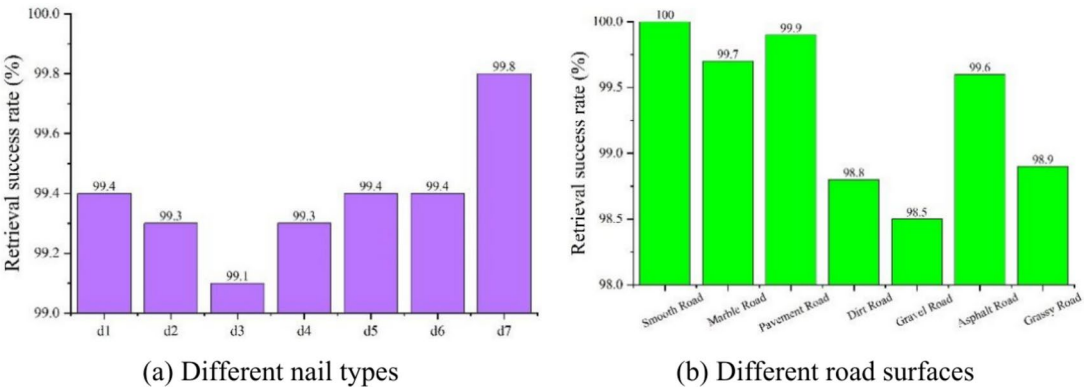


Fig. 19. Successful retrieval results.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available. We sincerely appreciate the editor’s attention and recognition of our research work. We fully understand journal’s requirements for data sharing. However, due to our laboratory’s policies or confidentiality agreements, we are unable to provide the raw data. We have comprehensively described the experimental design, analysis, and results, as well as the processes of data analysis and handling. If the editor or reviewers have any specific questions regarding the data, we will do our best to provide more detailed explanations and clarifications. In the study, Ziliang Hu should be contacted in case of any queries or data requirements. E-mail: huziliang@mails.guet.edu.cn.

Received: 8 December 2024; Accepted: 4 February 2025
Published online: 12 February 2025

References

1. Zhang, D. & Lu, G. Shape-based image retrieval using generic Fourier descriptor. *Signal Process. Image Commun.* 17(10), 825–848 (2002).
2. Dalal, N., Triggs, B. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, 886–893 (2005).

3. Jie, G., Honggang, Z., Daiwu, C., Nannan, Z. Object detection algorithm based on deformable part models. In *2014 4th IEEE International Conference on Network Infrastructure and Digital Content, Beijing, China*, 90–94 (2014).
4. Yang, S. Research on image recognition method based on improved neural network. In *2022 4th International Conference on Communications, Information System and Computer Engineering (CISCE), Shenzhen, China*, 125–129 (2022).
5. Zhang, X., Xu, C., Shen, M., et al. Survey of convolutional neural network. In *2018 International Conference on Network, Communication, Computer Engineering (NCCE 2018)* 93–97 (Atlantis Press, 2018).
6. Alex, K., Ilya, S., Geoffrey, E. H. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* 1097–1105 (2012).
7. Arora, D., Kulkarni, K. Efficient shelf monitoring system using faster-RCNN. In *2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT), Bengaluru, India*, 1–6 (2024).
8. Girshick, R., Donahue, J., Darrell, T., et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 580–587 (IEEE, 2014).
9. Girshick, R. Fast R-Cnn. *Comput. Sci.* **4** (2015).
10. Ren, S., He, K., Girshick, R., Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, 1137–1149.
11. Cho, M., Chung, T. -y., Lee, H., Lee, S. N-RPN: Hard Example Learning For Region Proposal Networks. In *2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan*, 3955–3959 (2019).
12. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–788 (2016).
13. Redmon, J., Farhadi, A. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7263–7271 (2017).
14. Redmon, J., Farhadi, A. Yolo3: An incremental improvement. arXiv preprint [arXiv:1804.02767](https://arxiv.org/abs/1804.02767) (2018).
15. Mi, R., Hui, Z., Li, L., Zhao, H., Li, W., Chen, Z. Vehicle target detection based on improved YOLOv3 algorithm in urban transportation. In *2024 3rd International Conference on Big Data, Information and Computer Network (BDICN), Sanya, China*, 21–27 (2024).
16. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y. M. Yolo4: Optimal speed and accuracy of object detection. arXiv preprint [arXiv:2004.10934](https://arxiv.org/abs/2004.10934) (2020).
17. He, K. et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(9), 1904–1916 (2015).
18. Jocher, G. YOLOv5 by Ultralytics. <https://github.com/ultralytics/yolov5>, 2020. (accessed 30 Feb 2023).
19. Ma, G., Zhou, Y., Huang, X., Zhen, S. A traffic flow detection system based on improved YOLOv5 in complex weather conditions. In *2023 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), Dalian, China*, 321–326 (2023).
20. Jiang, T., Xian, Y. Detection of traffic signs in complex weather conditions based on YOLOv5. In *2023 4th International Conference on Computer Engineering and Intelligent Control (ICCEIC), Guangzhou, China*, 534–538 (2023).
21. Singh, P., Gupta, K., Jain, A. K., Vishakha, Jain, A., Jain, A. Vision-based UAV detection in complex backgrounds and rainy conditions. In *2024 2nd International Conference on Disruptive Technologies (ICDT), Greater Noida, India*, 1097–1102 (2024).
22. Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., et al. Yolo6: A single-stage object detection framework for industrial applications. arXiv preprint [arXiv:2209.02976](https://arxiv.org/abs/2209.02976) (2022).
23. Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y. M. Yolo7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint [arXiv:2207.02696](https://arxiv.org/abs/2207.02696) (2022).
24. Jocher, G., Chaurasia, A., Qiu, J. YOLO by Ultralytics. <https://github.com/ultralytics/ultralytics>, 2023. (accessed 30 Feb 2023).
25. Wang, Q., Li, C., Liu, C., Wang, S., Tian, Y. YOLO-substation: Inspection target detection in complex environment based on improved YOLOv7. In *2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML), Chengdu, China*, 1142–1147 (2023).
26. Wang, C.-Y., Yeh, I.-H., and Liao, H.-Y. M. YOLOv9: Learning what you want to learn using programmable gradient information. arXiv e-prints (2024).
27. Lin, T. Y., Goyal, P., Girshick, R., et al. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **99**, 2999–3007 (2017).
28. Tan, M., Pang, R., Le, Q. V. EfficientDet: Scalable and efficient object detection. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2020).
29. Tan, M., Le, Q. V. EfficientNet: rethinking model scaling for convolutional neural networks. (2019).

Acknowledgements

This study was financially supported by the National Natural Science Foundation of China (Grant No. 52204130), the Guangxi Key Research and Development Program (Grant No. 2021AB04008), the Guangxi Key Laboratory of Manufacturing Systems and Advanced Manufacturing Technology (Grant numbers 22-35-4-S005), and the Innovation Project of GUET Graduate Education (Grant No. 2024YCX015).

Author contributions

Z.H. and X.Z. wrote the main manuscript text and prepared Figs. 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10. H.M. and H.W. prepared Figs. 11, 12, 13, 14, 15, 16, 17, 18 and 19 and Tables 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to X.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025