China Instrument and Control Society
Instrumentation
INSTRUMENTATION, Vol. 12, No. 2, June 2025 https://doi.org/10.15878/j.instr.202500252

Article

An Efficient Improved Yolov5-Based Method for Detecting Iron Waste in Ores

Kaiyu Yan[1,2], Juan Wang[1,2*], Jia Wang[1,2], Dawei Tian[1,2], Shu Peng[3], Yunhua Xu[1,2*]

[1] Shaanxi Key Laboratory of Nanomaterials and Nanotechnology, College of Mechanical and Electrical Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China

[2] Xi'an Key Laboratory of Clean Energy, School of Mechanical and Electrical Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China

[3] Xi'an Aerospace Yuanzheng Fluid Control Co., Ltd.710055, China

[*] Corresponding author email: juanwang618@126.com (J. Wang), xuyunhua2019@163.com9 (Y. Xun)

Abstract: Detection of ore waste is crucial for achieving automation in mineral metallurgy production. However, deep learning-based target detection algorithms still face several challenges in iron waste screening, including poor lighting conditions in underground mining environments, dust disturbances, platform vibrations during operation, and limited resources for large-scale computing equipment. These factors contribute to extended computation times and unsatisfactory detection accuracy. To address these challenges, this paper proposes an ore waste detection algorithm based on an improved version of YOLOv5. To enhance feature extraction capabilities, the RepLKNet module is incorporated into the YOLOv5 backbone and neck networks. This module enhances the deformation information of feature extraction with the maximum effective Receptive Field to increase the model's accuracy. The Normalization-based Attention Module (NAM) was introduced to enhance the attention mechanism by focusing on the most relevant features. This improves accuracy in detecting objects against noisy or unclear backgrounds, thereby further enhancing detection performance while reducing model parameters. Additionally, the loss function is optimized to constrain angular deviation using the SIOU loss function, which prevents the training frame from drifting during training and enhances convergence speed. To validate the performance of the proposed method, we tested it using a self-constructed dataset comprising 1,328 images obtained from the crushing station at Jinchuan Group's No. 2 mine. The results indicate that, compared to YOLOv5s on the self-constructed dataset, the proposed algorithm achieves an 18.3% improvement in mAP (0.5), a 54% reduction in FLOPs, and a 52.53% decrease in model parameters. The effectiveness and superiority of the proposed algorithm are demonstrated through case studies and comparative analyses.

Keywords: YOLOv5; ore waste detection; RepLKNet; NAM; SIOU

Citation: Kaiyu Yan, Juan Wang, Jia Wang, Dawei Tian, Shu Peng, Yunhua Xu."An Efficient Improved Yolov5-Based Method for Detecting Iron Waste in Ores." Instrumentation 12, no.2 (June 2025). https://doi.org/10.15878/j.instr.202500252

1 Introduction

In metallurgical and mining operations, crushing ore is often necessary for transportation and testing. However, iron wires, bolts, and other mixed iron waste can damage transmission and crushing equipment, leading to safety accidents during production[1]. Therefore, the accurate and efficient identification of iron

waste is crucial for enhancing production efficiency, reducing operational costs, and ensuring safety. Driven by advancements in computer technology, sensing technology, and artificial intelligence algorithms, traditional industrial production has gradually transitioned to automation, intelligence, and information technology modes in recent years. This transition effectively enhances production efficiency, reduces costs, minimizes safety accidents, and offers viable solutions for the industrial upgrading of mineral and metallurgy production sectors[2]. Currently, iron waste removal primarily depends on manual identification and screening processes. Image recognition and object detection technologies for iron waste detection offer convenient and efficient technical support.

Feature extraction strategies significantly influence the accuracy and efficiency of detection algorithms. Object detection algorithms can be categorized into traditional and deep learning-based approaches, depending on their feature extraction strategies[3]. Traditional object detection algorithms rely on manually designed features, such as Histogram of Oriented Gradient (HOG) and Scale Invariant Feature Transform (SIFT), for target recognition. Their performance is largely dependent on the quality of feature design and selection. However, feature design and optimization are costly and suffer from poor adaptability to complex scenes and inadequate processing capabilities for large-scale datasets. Deep learning object detection algorithms replace traditional manual feature design and extraction with deep network learning methods. This approach enhances the accuracy of feature extraction, increases algorithm precision and efficiency, and eliminates the need for distinct feature extraction solutions for varying scenarios. Consequently, it becomes more applicable in diverse contexts.

Currently, mainstream deep learning-based object detection algorithms are classified into two categories: one-stage and two-stage detection. One-stage detection algorithms directly identify targets based on features within the image's grid division, while two-stage algorithms first generate preselected frames to narrow the detection range, subsequently detecting targets within these frames[4]. Although two-stage object detection generally achieves higher accuracy, its complexity and longer operation times result in slower response rates, making it less adaptable to most practical applications. While improved algorithms, such as Fast R-CNN[5], have been proposed to address this issue, they still struggle to meet existing requirements. In contrast to two-stage object detection, one-stage methods like Single Shot Multibox Detector (SSD), Deconvolutional Single Shot Detector (DSSD), RetinaNet, and You Only Look Once (YOLO) offer distinct advantages. Although they may slightly compromise detection accuracy, their faster response times make them more suitable for practical engineering applications, with YOLO being the most representative example.

The YOLO algorithm, first proposed by Joseph et al. [6], employs a single Convolutional Neural Network to simultaneously perform target detection and classification. It transforms the target detection task into a regression problem, significantly simplifying the target localization process. This algorithm is characterized by rapid response and multi-target detection capabilities. Through iterative optimization, researchers have enhanced its learning framework[7], backbone structure[8], and anchor decision-making[7,9], leading to the development of several versions of the YOLO algorithm. In the evolution of the YOLO series, the relationship between versions is not characterized by a simple hierarchical substitution but rather by a trend toward diversification. Each version emphasizes different aspects to meet a wide range of application scenarios and requirements. For scenarios that demand high detection accuracy, YOLOv8 may be the better choice. In contrast, YOLOv5's flexibility and scalability may offer advantages in typical target detection situations. Based on the specific test data from this study, as shown in Table 1. It is worth mentioning that YOLOv5 is highly favored in practical engineering applications due to its high detection accuracy, fast processing speed, low dependency on device performance, and ease of use of its code[10].

Table 1 YOLO Model Comparison

Model	Precision	Recall	mAP (0.5)	mAP (0.5:0.95)	Parameters(M)	FLOPs(G)
YOLOv5s	90.4%	78.0%	80.2%	49.3%	7.03	16.0
YOLOv7-tiny	87.0%	75.1%	76.2%	37.0%	6.02	13.2
YOLOv8s	78.0%	68.5%	72.2%	43.5%	11.13	28.4
YOLOv11s	71.9%	67.9%	71.4%	42.4%	9.43	21.6

This article includes a side-by-side comparison of the widely used YOLOv8 and YOLOv5. The results indicate that the mean Average Precision (mAP) at 0.5 for YOLOv5 improved by 8.0%, and the model's Parameters were reduced by 40.18%. Experiments show that YOLOv5 outperforms YOLOv8 in terms of detection accuracy and deployment challenges. Liu and Luo[11] replaced the YOLOv5 backbone network with EfficientLite and incorporated an adaptive feature fusion module at the head of the base network, significantly enhancing

YOLOv5's detection accuracy for multicopters. Liu et al. [12] introduced the C3Ghost and GhostConv modules into the YOLOv5 backbone network to further enhance detection efficiency. They also incorporated the Coordinated Attention Mechanism module, effectively reducing the number of model parameters and minimizing the resource requirements for deploying robot object detection systems. Nevertheless, the YOLO object detection algorithm has yet to achieve a groundbreaking application in the field of ore waste detection.

The actual working conditions for ore waste detection are complex. Challenges affecting detection include relatively small objects, obscured targets, dust from crushing stations disturbing pixels, and the rapid transmission speed of ore. Effective analysis of image features and the reduction of network parameters are key challenges in ore waste detection. To address these issues, Dong et al. employed median filtering to pre-treat foreign objects, mitigating dust effects, and integrated the GECA attention module into the YOLO algorithm to further enhance speed and accuracy in object detection[13]. However, the average accuracy and detection time of these YOLO models remains insufficient for practical applications.

To enhance the speed and accuracy of ore waste object detection, this paper proposes an improved YOLOv5 detection model. Firstly, the accuracy of object detection is enhanced by replacing the C3 module in the YOLOv5 network with RepLKNet, resulting in a larger receptive field and more efficient feature information. Secondly, a lightweight Normalization-based Attention Module (NAM) is added to the head of the YOLOv5 network, reducing computational overhead and improving detection accuracy. Finally, we modify the loss function in YOLOv5 from CIOU to SIOU. This adjustment introduces an angular cost criterion, enhancing the model's convergence speed. Experimental results demonstrate that the improved YOLOv5 model proposed in this paper efficiently and accurately identifies ore waste, aligning with practical operational requirements.

2 Research Methodology

2.1 YOLOv5 Network

In YOLOv5 (v6.1), there are five versions: YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. Among these, YOLOv5s is the most widely adopted model for engineering applications within the YOLOv5 series.

It offers high detection speed, superior detection accuracy, and greater ease of model modification. It is suitable for deployment in application environments with limited computational resources, such as mining production operations[12]. Therefore, this paper utilizes YOLOv5s as the base network and implements improvements to achieve efficient detection of ore waste. The structure of YOLOv5s is divided into four components: the input module, the backbone network, the neck network, and the detection head. This structure is illustrated schematically in Fig. 1.

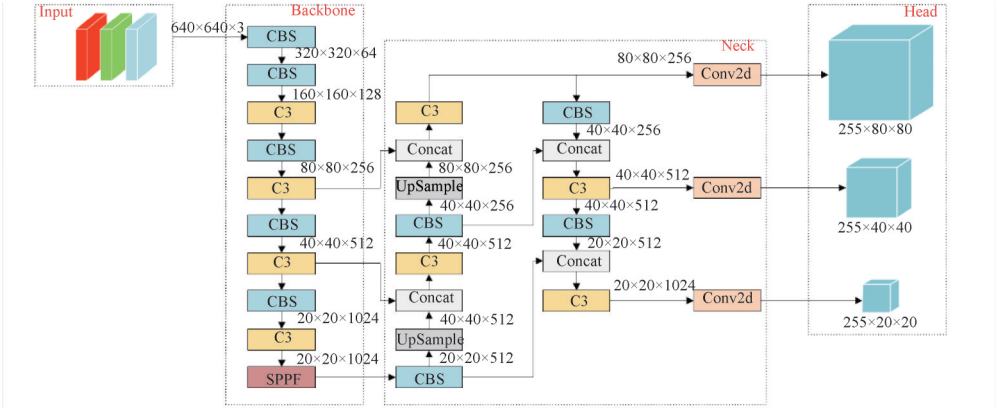


Fig.1 Structural diagram of YOLOv5s

2.2 Improved YOLOv5 Network Structure

The construction of convolutional neural networks typically employs a large number of small convolutional kernels to enhance image detection performance[15]. However, small convolutional kernels possess a limited receptive field and are constrained in their ability to extract object shape information. In contrast, this paper utilizes the RepLKNet module, which features a large convolutional kernel in place of the C3 module, to enhance target deformation feature information and expand the receptive field. RepLKNet's capacity to capture more spatial and shape-related features is particularly advantageous for ore waste detection, where irregular object shapes and varying textures complicate the detection process. Additionally, the reduced network depth enabled by the RepLKNet module helps mitigate optimization challenges to some extent. In a comparative experiment, Ding et al. [16] employed a simple backbone replacement method to modify the Cascade Mask R-CNN backbone and conducted performance testing on the

COCO dataset. The experimental results indicated that RepLKNet achieved an 8.07% improvement in AP value, with a model parameter reduction of 3M. Furthermore, to mitigate the computational burden associated with large convolutional kernels, we not only design an enhancement algorithm for the neck network to reduce computations but also adopt SIOU to replace the original CIOU loss function of YOLOv5s, thereby improving detection accuracy. SIOU improves upon CIOU by considering both the distance and aspect ratio between predicted and ground truth boxes, which leads to more precise localization, especially for small or highly deformed objects like ore waste. The subsequent sections will explore the improvements made to YOLOv5s in detail across three aspects: the backbone network, the neck network, and the loss function. The structure of the improved network is illustrated in Fig. 2.

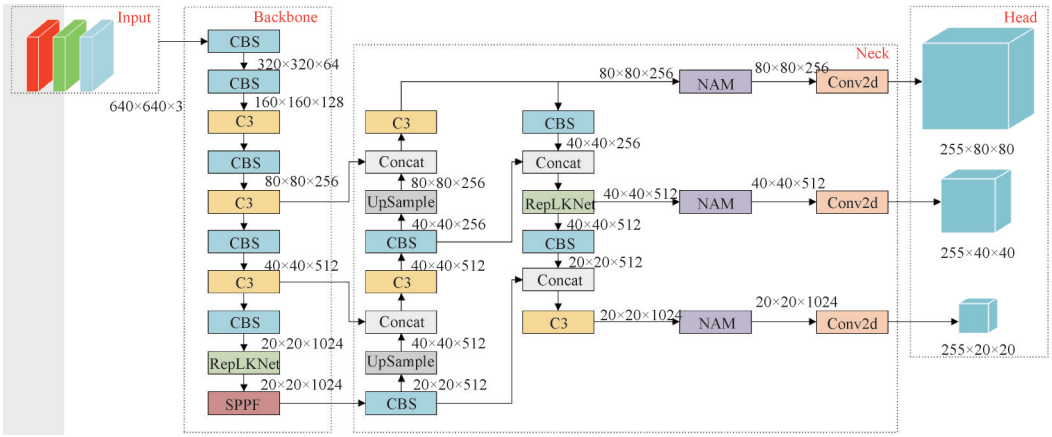


Fig.2 Improved YOLOv5 structure

2.2.1 Backbone Network

To further enhance the detection accuracy of YOLOv5s and obtain more effective graphical and positional features during image feature extraction, this study replaces the C3 module at the end of the original backbone network structure with the RepLKNet module proposed by Ding et al.[16]. The detailed structure of the RepLKNet module, which consists of the input feature processing module (Stem), the working module (Stage), and the transition module (Transition), is illustrated in Fig.3. The Stem module comprises sequentially arranged convolutional layers of 3×3, 3×3DW, 1×1, and another 3×3DW. Here, DW refers to depthwise convolution. Feature information is initially captured by connecting a 3×3convolutional layer with a stride of 2 to a 3×3DW convolutional layer, immediately followed by subsampling via a 3×3DW convolutional layer connected to a 1x1 convolutional layer. The Stage and Transition modules are added sequentially following the Stem module. The Stage module comprises multiple RepLK Block modules and employs a skip connection method, thereby enhancing the receptive field and aggregating spatial information[17]. Each RepLK Block module is subsequently followed by a ConvFFN module, which consists of batch normalization (BN), two 1×1 convolutional layers, and the skip of the GELU activation function, aimed at increasing model depth and enhancing characterization capabilities[18]. The Transition module increases the channel dimension using a 1×1 convolutional layer and performs subsampling with a $3\times$

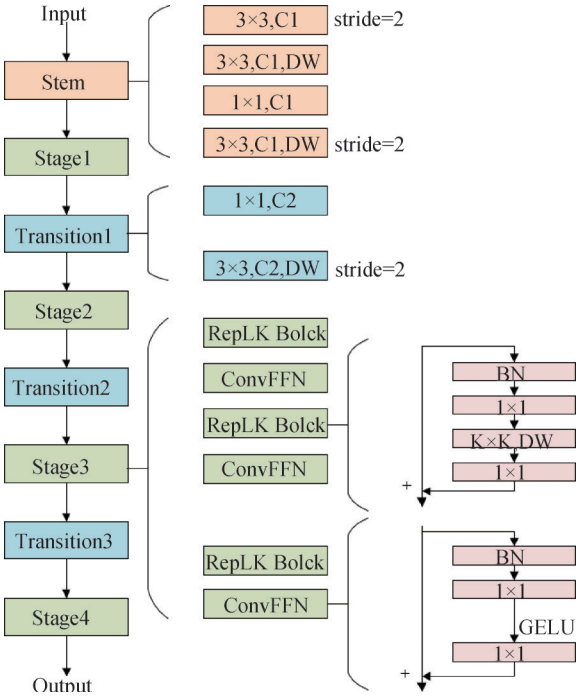


Fig.3 RepLKNet module

3DW convolutional layer with a stride of 2.

Experiments indicate that by strictly following the skip connection methodology and employing a reparameterized small kernel to address optimization challenges, large convolutional kernels perform comparably well, even on small feature maps, and exhibit superior performance on downstream tasks[16].

2.2.2 Neck Network

In downstream tasks, lower accuracy in small target detection is often attributed to the network's insufficient capability to extract relevant feature information, while irrelevant information in feature feedback consistently hinders network detection efficiency. To enhance the model's detection accuracy, the network integrates a lightweight attention module known as the Normalization-based Attention Module (NAM) [19] and replaces the first C3 module in the Feature Pyramid Network subsampling with the previously described RepLKNet module. This approach aims to improve feature parsing ability, enhance feature attention following the neck network, and boost detection efficiency and accuracy without increasing model parameters.

NAM is an attention mechanism grounded in normalization techniques that compute attention weights using scaling factors derived from Batch Normalization (BN). It avoids the use of fully connected layers[20] found in attention mechanisms such as Squeeze-and-Excitation Networks (SE) and Convolutional Block Attention Module (CBAM), thereby reducing computational complexity and the number of parameters. This method enhances computational efficiency while effectively suppressing non-significant features. In NAM, feature batch normalization is achieved through the following equation:

$$\mathrm{BN}(\mathbf{B}) = \boldsymbol{\gamma}\frac{\mathbf{B}-\mu_{\mathrm{B}}}{\sqrt{\sigma_{\mathrm{B}}^2+\varepsilon}} + \boldsymbol{\beta} \tag{2.1}$$

In the formula (2.1), B represents the original features. μ_{B} and σ_{B} represent the mean and variance of the batch. ε represents a very small constant (usually set to 3) used to prevent the denominator from being zero due to zero variance. β and γ represent the same linear transformation parameters as the feature dimensions, which are obtained through training. The expression for the attention module in NAM can be formulated as

$$\mathbf{M}_{\mathrm{C}} = \mathrm{sigmoid}\left(\mathrm{M}_{\gamma}\left(\mathrm{BN}\left(\mathbf{F}_1\right)\right)\right) \tag{2.2}$$

$$\mathrm{W}_{\gamma} = \frac{\gamma_i}{\sum_{j=0}\gamma_j} \tag{2.3}$$

In the equation, $\mathbf{M}_{\mathrm{C}}$ represents the output feature, γ is the scale factor for each channel with weight W_{γ}, and $\mathbf{F}_1$ is the input feature. The scale factor is also applied to the spatial dimension, called pixel normalization. The expression for the corresponding spatial attention submodule can be formulated as

$$\mathbf{M}_{\mathrm{S}} = \mathrm{sigmoid}\left(\mathrm{W}_{\lambda}\left(\mathrm{BN}_{\mathrm{S}}\left(\mathbf{F}_2\right)\right)\right) \tag{2.4}$$

$$\mathrm{W}_{\lambda} = \frac{\lambda_i}{\sum_{j=0}\lambda_j} \tag{2.5}$$

Where the output is denoted as $\mathbf{M}_S$, λ is the scale factor weights as W_{λ}, and $\mathbf{F}_2$ is the input features.

This paper employs a serial integration model, where the channels are sequentially arranged first, followed by spatial arrangement. The attention maps generated by this method are more fine-grained compared to those produced through parallel processing. The overall structure of NAM is illustrated in Fig. 4.

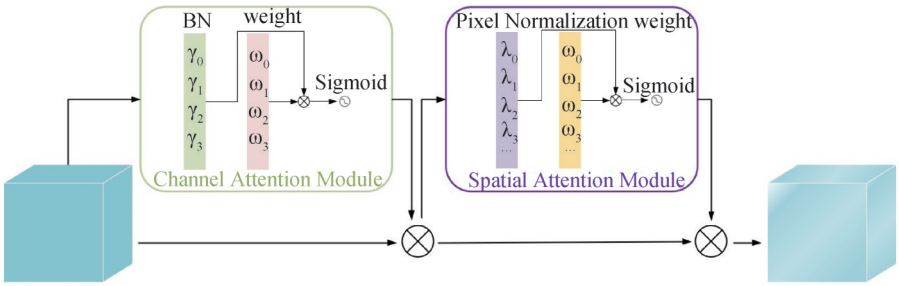


Fig.4 NAM module

2.2.3 Loss Function

The default loss function in YOLOv5 is CIOU[21], which does not consider the angular deviation between the ground truth box and the predicted box in its design. This oversight may cause the predicted box to drift during the training process. To address these issues, this paper adopts SIOU to replace CIOU to enhance object detection accuracy. SIOU introduces an angle loss function Λ, a distance loss functionΔ, and a shape loss function Ω based on IOU[22]. These functions facilitate faster convergence of the predicted box to the correct position during training, thereby enhancing training efficiency[23].

The expression for angular loss can be formulated as

$$\Lambda = 1 - 2\sin^2\left(\arcsin\left(\frac{C_h}{\sigma}\right) - \frac{\pi}{4}\right) \tag{2.6}$$

In the equation, C_h is the height difference between the center points of the ground truth box and the prediction box, and σ is the distance between the center points of the truth box and the prediction box.

The expression for the distance loss can be expressed as

$$\Delta = 2 - \mathrm{e}^{\gamma\rho_x} - \mathrm{e}^{\gamma\rho_y} \tag{2.7}$$

$$\rho_{\mathrm{x}}=\left(\frac{b_{\mathrm{c_x}}^{\mathrm{gt}}-b_{\mathrm{c_x}}}{c_{\mathrm{w}}}\right)^2 \tag{2.8}$$

$$\rho_{\mathrm{y}}=\left(\frac{b_{\mathrm{c_y}}^{\mathrm{gt}}-b_{\mathrm{c_y}}}{c_{\mathrm{h}}}\right)^2 \tag{2.9}$$

Among them, $b_{\mathrm{c_x}}^{\mathrm{gt}}-b_{\mathrm{c_x}}$ represents the distance between the centroids of the ground truth box and the prediction box in the x-axis direction. $b_{\mathrm{c_y}}^{\mathrm{gt}}-b_{\mathrm{c_y}}$ represents the distance between the centroids of the ground truth box and the prediction box in the y-axis direction. c_{w} and c_h represent the width and height of the smallest packet closure area, $\gamma=2-\Lambda$.

The expression for shape loss can be expressed as

$$\Omega=\left(1-\mathrm{e}^{\varphi_w}\right)^{\theta}+\left(1-\mathrm{e}^{\varphi_h}\right)^{\theta} \tag{2.10}$$

$$\varphi_w=\frac{\left|w-w^{\mathrm{gt}}\right|}{\max\left(w,w^{\mathrm{gt}}\right)} \tag{2.11}$$

$$\varphi_h=\frac{\left|h-h^{\mathrm{gt}}\right|}{\max\left(h,h^{\mathrm{gt}}\right)} \tag{2.12}$$

In the equation, (w,h) and $\left(w^{\mathrm{gt}},h^{\mathrm{gt}}\right)$ are the widths and heights of the prediction box and Ground truth box, and θ controls the attention of the shape loss, with the result taken as 4.

Combining the angle, distance, and shape loss functions, the SIOU loss function can be derived as

$$Loss_{\mathrm{SIOU}}=1-\mathrm{IOU}+\frac{\Delta+\Omega}{2} \tag{2.13}$$

3 Experimental Analysis

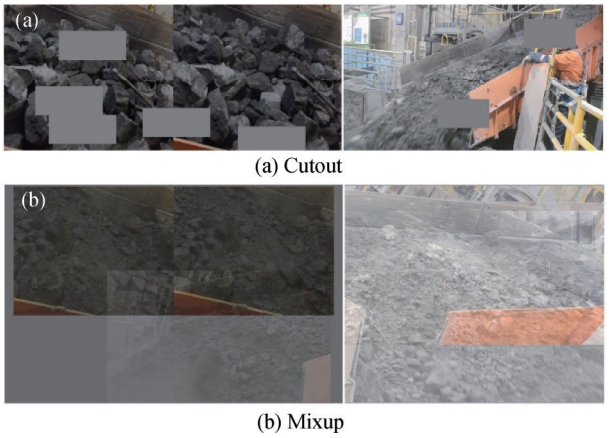
3.1 Dataset

The dataset presented in this paper was compiled by the research team in collaboration with Jinchuan Group Ltd by capturing real-time photographs of the conveyor unit at the No. 2 mine's crushing station. To enhance the information regarding angle, illumination, and distance in the images, the image acquisition process utilized both conventional cameras and a stereo vision camera in fixed positions and handheld mobile mode. The conventional camera has a pixel size of 1.75×1.75μm and a maximum resolution of 3840×2160. The stereo vision camera features a pixel size of 2×2 μm, a focal length of 4 mm, and a maximum resolution of 2560×720. The dataset comprises 949 images from the conventional camera and 379 photographs from the stereo vision camera, encompassing two types of iron waste-wires and bolts-captured from various viewpoints, as illustrated in Fig. 5.

Given the relatively small size of the dataset and the need to accommodate the proposed large convolutional kernel model while enhancing the reliability of the validation methods, data augmentation was applied to the original dataset using the Mixup[24] and Cutout[25] techniques.

Fig.5 Example of a self-built dataset

Specifically, 1, 315 images were augmented using the Cutout method, while 2,000 images were enhanced using the Mixup method. An example of the enhanced images is depicted in Fig. 6.



(a) Cutout

(b) Mixup

Fig.6 Data Enhancement

3.2 Experimental Settings

This study employed two types of computing devices: a desktop PC for model training and a laptop for field deployment. The detailed performance parameters of these devices are presented in Table 2.

Table 2 Device parameters

Name	Processor	Memory	TFLOPS
Desktop computers	GeForce RTX3060Ti	8G	16.2
Notebooks	GeForce RTX1050Ti	4G	2.1

Fig. 7 illustrates the network training and testing processes for ore waste detection. First, the augmented dataset is randomly split using cross-validation, with 80% allocated for training and the remaining 20% for validation. Secondly, a genetic algorithm is employed to select the optimal initial learning rate for the model (lr0= 0.01) over 300 iterations. Parameter iteration is carried out by randomly selecting the previous hyperparameter as the baseline parameter, based on the model's weights. The genetic operator involves crossover mutations with a probability of 0.8 and a variance of 0.2. The mutation results are then evaluated using a fitness function, and the

generation with the highest fitness, representing the optimal solution, is selected and saved. Additionally, based on pre-training convergence data and computational device specifications, initial parameters are set (batch-size=16, epochs=300) to begin model training on desktop PCs. The best network model is saved upon completion of training. The trained network model is then deployed on a laptop, with a passive binocular camera used as the data acquisition device. This setup is positioned at the manually operated platform guardrail of a crushing station, where a stable light source ensures real-time monitoring of ore stream transmission, and video is captured to evaluate the model's performance. Finally, save the model's test results. It is important to emphasize that the YOLOv5 algorithm in this study serves as a baseline for comparing the performance of the proposed improved algorithm. The test results are shown in Fig.8 and Table 3. The figure indicates that the improved network structure demonstrates higher confidence in detecting both larger bolts and smaller wires, confirming its effectiveness in such detections.

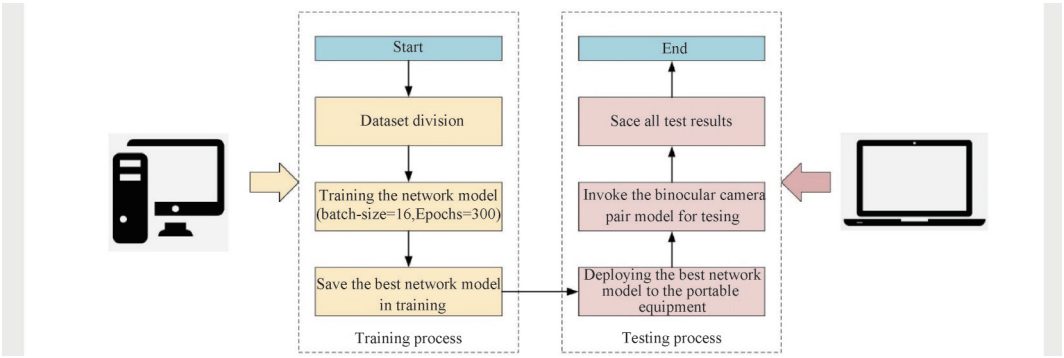
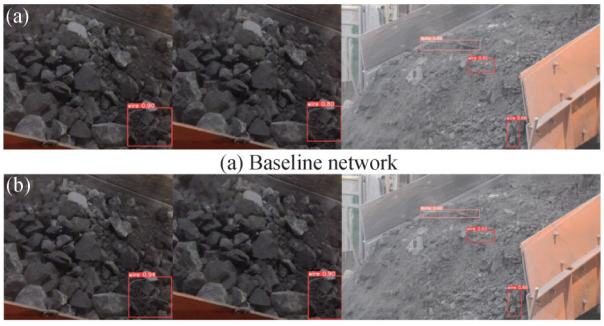


Fig.7 Training and Experimental Procedures for the Improved Algorithm. The training phase was conducted on a PC (lr0 = 0.01, batch size = 16, epochs = 300), with the dataset randomly split into training and validation sets at a ratio of 80:20. The testing phase was performed on a laptop, with the experiments conducted at a downhole crushing station.



(a) Baseline network

(b) Improved network. Confidence Gains of Up to 0.1

Fig.8 Comparison of Detection Using Binocular and Monocular Images

4 Results

This section evaluates the performance of the improved YOLOv5s algorithm proposed in this paper for ore waste detection, discussing it in detail from two perspectives: various attention mechanisms and different lightweight object detection algorithms. The accuracy metrics utilized include precision, recall, and mean average precision (mAP), defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.2}$$

Table 3 performance comparison

Model	Precision	Recall	mAP (0.5)	mAP (0.5:0.95)	Parameters(M)	FLOPs(G)
YOLOv5s	90.4%	78.0%	80.2%	49.3%	7.03	16.0
Ours	96.8%	96.1%	98.5%	77.2%	6.66	15.4

In Eqs. (4.1) and (4.2), True Positive (TP): indicates that the target is a positive example and the predicted outcome is the same as a positive example. False Positive (FP): indicates that the target is a counterexample, but the predicted outcome is a positive example. False Negative (FN): indicates that the target is a positive example and the predicted outcome is a negative example.

The area under the Precision-Recall (PR) curve, plotted with Recall on the horizontal axis and Precision on the vertical axis, is defined as Average Precision (AP). The following Eq.(4.3) can be used to determine AP.

$$\text{AP} = \frac{1}{m}\sum_{i}^{m} \text{p}_i \tag{4.3}$$

Among them, m is the sample of positive examples among all samples, and p is the maximum precision of each recalled sample. mAP is calculated by averaging the AP of all classes in the dataset. mAP(0.5) is the IOU set to 0.5 and mAP (0.5: 0.95) is the average mAP for

different IOU thresholds. The IOU varies from 0.5 to 0.95 in steps of 0.05. Additionally, this study assesses model complexity based on metrics such as Single Image Detection Time, Frames Per Second (FPS), Parameters, and GFLOPs to evaluate algorithm performance on hardware.

4.1 Performance Comparison of Different Attention Mechanisms

To validate the superiority of the added NAM in the improved structure, it is compared with three typical attention modules: SimAM[26], SKA[27], and CBAM[20]. To facilitate the presentation of detection results, a heat map was generated using the Grad-CAM method[28], as shown in Fig. 9.

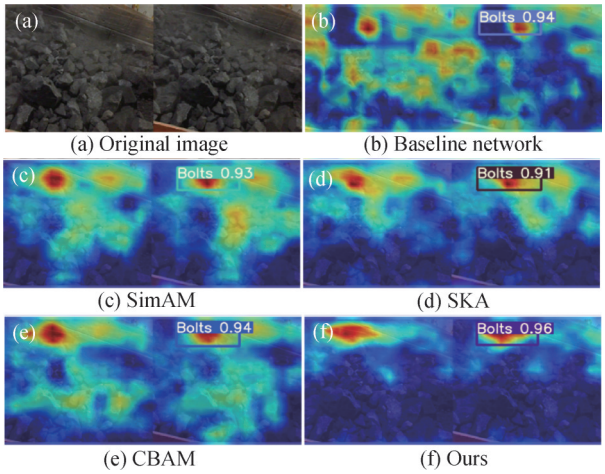


Fig.9 Thermal distribution test

The analysis yields the following conclusions:

1) Incorporating the Attention Mechanism module significantly enhances the algorithm's capacity to focus on object features during detection, thereby improving its ability to extract valid features and filter out irrelevant ones. The ranking of feature screening capabilities is as follows: NAM > SKA > SimAM > CBAM.

2) The accuracy of the model in detecting ore waste slightly fluctuated after adding different attention mechanisms. The detection accuracy improved with the addition of NAM. The detection accuracy of the CBAM module corresponding to the network structure remains almost unchanged, while the detection accuracy of SimAM and SKA corresponding to the network structure is slightly decreased.

The results for detecting ore waste using the aforementioned network structure are summarized in Table 2. The results indicate that the network structure incorporating NAM demonstrates the highest detection accuracy. The Precision, recall, mAP (0.5), and mAP (0.5: 0.95) for object detection are 96.8%, 96.1%, 98.5%, and 77.2%, respectively. The above metrics have been improved by 6.4%, 18.1%, 18.3%, and 27.9% compared to the original network structure. Furthermore, the improved network structure reduces the model's parameter count and complexity while maintaining detection accuracy. The number of parameters and complexity are 6.66 million and 15.4 GFLOPs, respectively, which are the minimum values of Parameters and GFLOPs in Table 4 and decreased by 5.26% and 3.75% compared to the original network structure.

Table 4 Results of the comparison of different attention mechanisms

Model	Precision	Recall	mAP (0.5)	mAP (0.5:0.95)	Parameters(M)	FLOPs(G)
Original	90.4%	78.0%	80.2%	49.3%	7.03	16.0
SimAM[26]	91.6%	77.4%	80.3%	47.2%	7.03	16.0
SKA[27]	92.4%	75.4%	80.6%	47.7%	29.14	33.6
CBAM[20]	89.9%	78.4%	80.7%	47.8%	7.06	16.0
Ours	96.8%	96.1%	98.5%	77.2%	6.66	15.4

4.2 Performance Comparison of Different Lightweight Detection Algorithms

To further validate the method's performance, this section deploys it on desktop PCs and laptops, comparing it with existing mainstream lightweight detection algorithms to assess its detection speed and accuracy. Fig. 10 illustrates the response times of the various algorithms on both devices. The following conclusions can be drawn:

1) From the perspective of single image detection time, on a desktop PC, the improved YOLOv5s is 1 ms faster than Elan, Rghostnet, and Pwconv2, and 8 ms faster than Afpn. On the laptop, it is 3.9 ms and 4.9 ms faster than the Shuffle and Afpn algorithms, respectively. Overall, the improved YOLOv5s algorithm demonstrates shorter image detection times. The experimental results are presented in Fig. 10(a).

2) In terms of the frames per second (FPS) transmission rate, the improved YOLOv5s achieved 55.6 FPS on the desktop PC, which is comparable to most mainstream algorithms. On the laptop, it surpassed the Shuffle, Fasternet, and Afpn algorithms by 6.3 FPS, 16.4 FPS, and 8.9 FPS, respectively. Overall, the improved YOLOv5s algorithm exhibits a high transmission frame rate per second. The experimental results are shown in Fig. 10(b).

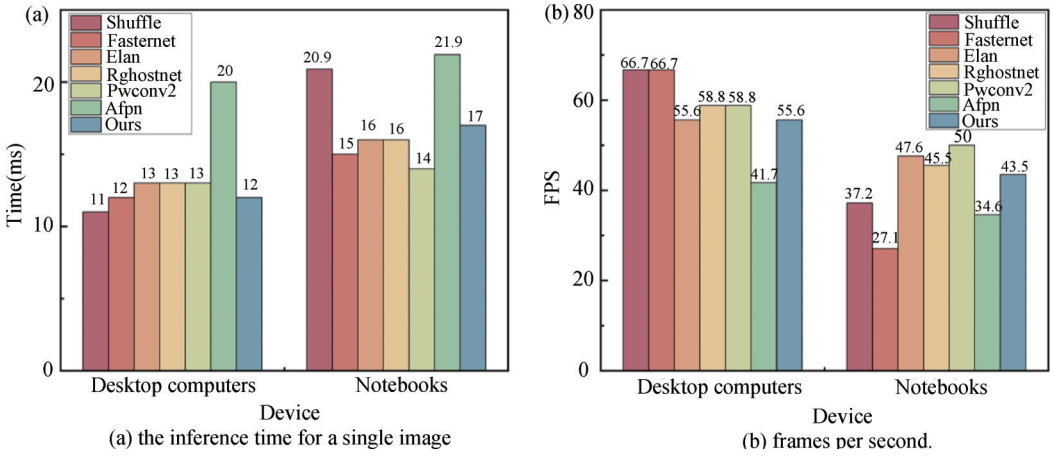


Fig.10 Comparison of response time under different algorithms

The results for detecting iron waste using the aforementioned network structure are summarized in Table 5. The improved YOLOv5 algorithm demonstrates superior detection accuracy and speed. Compared to the Afpn algorithm, which has a relatively balanced detection accuracy and model complexity, its model parameter count and model complexity are increased by only 21.16% and 16.23%. The Precision is improved by 6.6%, the recall is improved by 17.9%, the mAP (0.5) is improved by 17.5%, and the mAP (0.5: 0.95) is even improved by 32.1%. These improvements can be attributed to the incorporation of NAM.

Table 5 Comparison of results of different lightweight models

Model	Precision	Recall	mAP (0.5)	mAP (0.5:0.95)	Parameters (M)	GFLOPs
ShuffleNet[29]	87.6%	61.3%	65.7%	31.8%	3.19	5.9
FasterNet[30]	90.4%	76.5%	79.5%	46.0%	6.37	14.7
ELAN[31]	89.0%	74.0%	77.0%	45.0%	5.52	11.6
RepGhost[32]	89.4%	73.0%	77.0%	44.3%	6.05	13.2
PWConv2[30]	90.0%	76.5%	79.1%	46.9%	6.11	14.3
Afpn[33]	90.2%	78.2%	81.0%	45.1%	5.25	12.9
Ours	96.8%	96.1%	98.5%	77.2%	6.66	15.4

4.3 Ablation Study

To verify the effectiveness of the improvements proposed in this study, ablation experiments were conducted using iron waste identification. The analysis and comparison were based on a unified dataset with modular modifications. The experimental results are presented in Table 6. As shown in Table 6, the application of two data augmentation methods significantly enriches the data features, leading to improved detection efficiency. Secondly, the NA module and the SIOU loss function contribute to better target focus and model convergence, thereby enhancing target recall to a greater extent. Finally, the use of a large convolution kernel

Table 6 Ablation study

Number	DA	SIOU	RepLKNet	NAM	mAP(0.5)	Precision	Recall
1	×	×	×	×	80.2%	90.4%	78.0%
2	√	×	×	×	96.9%	94.8%	94.4%
3	√	√	×	×	98.0%	94.7%	95.9%
4	√	×	√	×	96.5%	96.5%	92.4%
5	√	×	×	√	98.3%	97.1%	96.1%
6	√	√	√	×	96.7%	95.6%	93.7%
7	√	×	√	√	98.4%	97.6%	94.9%
Ours	√	√	√	√	98.5%	96.8%	96.1%

improves feature extraction capabilities, resulting in a moderate increase in detection precision. Additionally, while the method proposed in this study does not achieve the highest detection precision and recall rate in the experiments, it demonstrates the highest average detection precision, with a more balanced overall performance.

In conclusion, the improved YOLOv5s algorithm proposed in this paper enhances the speed of ore waste detection while significantly improving detection accuracy, aligning with the operational needs of actual metallurgical and mining production activities.

5 Conclusion

To enhance the efficiency of ore waste processing and advance technological development in mining production-related industries, this paper proposes an efficient algorithm based on YOLOv5s for addressing the issue of iron waste screening during underground crushing. Firstly, we replace the two C3 modules in the backbone and neck networks with RepLKNet modules to enhance deformation information and expand the effective receptive field during feature extraction, thereby improving detection accuracy. Secondly, the less computationally intensive NAM was incorporated into the neck network to mitigate the negative impact of RepLKNet modules, reduce the target range during feature parsing, and accelerate feature fusion, thus lowering computational complexity. Finally, the loss function is modified to SIOU to address the computational resource demands of anchor angle matching and to shorten the time-consuming training phase. Finally, the loss function is modified to SIOU to address the computational resource demands of anchor angle matching and to shorten the time-consuming training phase.

In future work, further research should focus on the fusion of infrared and visible images to increase the feature information of the detected images and to recognize the ore waste more efficiently. Furthermore, improving the stereo-matching algorithm is also a key area for further research. Such advancements may facilitate the modernization of ore waste screening processes in crushing stations.

Author Contribution:

Kaiyu Yan: Conceptualization, methodology, data organization, writing-original manuscript preparation, visualization. All authors reviewed the manuscript.

Foundation Information:

This work is supported by the Department of science and technology of Shaanxi Province (NO. 2023-ZDLGY-24).

Data availability

The data that support the findings of this study are available on request from the corresponding author, [Wang],upon reasonable request.

Conflicts of Interest:

The authors declare no competing interests.

Dates:

Received 26 November 2024; Accepted 16 April 2025; Published online 30 June 2025

References

[1] Saran G, Ganguly A, Tripathi V, et al (**2022**) Multi-Modal Imaging-Based Foreign Particle Detection System on Coal Conveyor Belt. *Transactions of the Indian Institute of Metals* 75(9): 2231-2240. https: //doi. org/10.1007/s12666-021-02492-3

[2] Yang Z, Ge Z (**2022**) On Paradigm of Industrial Big Data Analytics: From Evolution to Revolution. *IEEE Transactions on Industrial Informatics* 18(12): 8373-8388. https://doi. org/ 10.1109/TII. 2022.3190394

[3] Zou Z, Chen K, Shi Z, et al (**2023**) Object Detection in 20 Years: A Survey. In: *Proceedings of the IEEE* 111(3): 257–276. https://doi.org/ 10. 1109/JPROC. 2023.3238524

[4] Guo Z, Wang C, Yang G, et al (**2022**) MSFT-YOLO: Improved YOLOv5 Based on Transformer for Detecting Defects of Steel Surface. *Sensors* 22(9): 3467. https://doi.org/ 10.3390/s22093467

[5] Girshick R (**2015**) Fast R-CNN. In*:* Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1440-1448. https://doi. org/ 10. 1109/ICCV.2015.169

[6] Redmon J, Divvala S, Girshick R, et al (**2016**) You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR), PP. 779-788. https:// doi.org/ 10.1109/CVPR.2016.91

[7] Ge Z, Liu S, Wang F, et al (**2017**) YOLOX: Exceeding YOLO Series in 2021. arXiv: 2107. 08430.

[8] Wang C-Y, Bochkovskiy A, Liao H-Y M (**2023**) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In*:* Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 7464-7475. https://doi. org/ 10.1109 10.1109

[9] Redmon J, Farhadi A (**2017**) YOLO9000: Better, Faster, Stronger. In*:* Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 6517-6525. https://doi.org/10.1109/CVPR.2017.690

[10] Terven J, Cordova-Esparza D (**2023**) A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction* 5(4): 1680-1716. https://doi. org/

10. 3390/make5040083

[11] Liu B, Luo H (**2022**) An Improved YOLOv5 for Multi-Rotor UAV Detection. *Electronics* 11(15): 2330. https://doi. org/ 10.3390/electronics 11152330

[12] Liu G, Hu Y, Chen Z, et al (**2023**) Lightweight object detection algorithm for robots with improved YOLOv5. *Engineering Applications of Artificial Intelligence* 123: 106217. https://doi.org/ 10.1016/j.engappai.2023.106217

[13] Xiao D, Kang Z, Yu H, et al (**2022**) Research on belt foreign body detection method based on deep learning. *Transactions of the Institute of Measurement and Control* 44(15): 2919-2927. https://doi.org/10.1177/01423312221094393

[14] Xiao D, Liu P, Wang J, et al (**2023**) Mining belt foreign body detection method based on YOLOv4_GECA model. *Scientific Reports* 13(1): 8881. https://doi. org/10.1038/s41598-023-35962-3

[15] Tan M, Le Q (**2019**) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv* : 1905.11946. http: //arxiv.org/abs/1905. 11946

[16] Ding X, Zhang X, Han J, et al (**2022**) Scaling Up Your Kernels to 31x31: Revisiting Large Kernel Design in CNNs. In*:* Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 11953-11965. https://doi.org/10.1109/CVPR52688.2022.01166

[17] Ding X, Zhang X, Ma N, et al (**2021**) RepVGG: Making VGG-style ConvNets Great Again. In*:* Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition **2101**.03697: 13728-13737. https:// arxiv. org/abs/2101.03697

[18] Hendrycks D, Gimpel K (**2016**) Gaussian Error Linear Units (GELUs). arXiv: 1606.08415.

[19] Liu Y, Shao Z, Teng Y, et al (**2021**) NAM: Normalization-based Attention Module. *Neural Information Processing Systems* , pp. 1-5.

[20] Woo S, Park J, Lee J Y, et al (**2018**) CBAM: Convolutional Block Attention Module. *Lecture Notes in Computer Science* 11211:3-19. https:// doi.org/10.1007/978-3-030-01234-2_1

[21] Zheng Z, Wang P, Liu W, et al (**2019**) Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. In: Proceedings of the AAAI Conference on Artificial Intelligence 34(07): 12993-13000.

[22] Yu J, Jiang Y, Wang Z, et al (**2016**) UnitBox: An Advanced Object Detection Network. In*:* Proceedings of the 24th ACM International Conference on Multimedia, pp. 516-520.

[23] Gevorgyan Z (**2022**) SIoU Loss: More Powerful Learning for Bounding Box Regression. arXiv: 2205.12740.

[24] Zhang H, Cisse M, Dauphin Y N, et al (**2018**) mixup: Beyond Empirical Risk Minimization. In*:* Proceedings of the International Conference on Learning Representations, pp. 1-12.

[25] DeVries T, Taylor G W (**2022**) Improved Regularization of Convolutional Neural Networks with Cutout. In*:* Proceedings of the International Conference on Artificial Intelligence and Security(ICAIS) **2022**: 1587.

[26] Yang L, Zhang R Y, Li L, et al (**2021**) SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In*:* Proceedings of the International Conference on Machine Learning(ICML), pp. 1-11.

[27] Li X, Wang W, Hu X, et al (**2019**) Selective Kernel Networks. In*:* Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 510-519. https://doi.org/ 10.1109/CVPR.**2019**.00060

[28] Selvaraju R R, Cogswell M, Das A, et al (**2017**) Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In*:* Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 618 – 626.

[29] Zhang X, Zhou X, Lin M, et al (**2018**) ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In*:* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6848-6856.

[30] Chen J, Kao S, He H, et al (**2023**) Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 12021-12031.

[31] Wang C Y, Yeh I H, Liao H Y M (**2024**) YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. arXiv: 2402.13616.

[32] Chen C, Guo Z, Zeng H, et al (**2022**) RepGhost: A Hardware-Efficient Ghost Module via Re-parameterization. arXiv: 2211.06088.

[33] Yang G, Lei J, Zhu Z, et al (**2023**) AFPN: Asymptotic Feature Pyramid Network for Object Detection. In*:* Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2184-2189. https://doi. org /10.1109/ SMC53992.2023.10394415