Mech. Sci., 13, 979–990, 2022
https://doi.org/10.5194/ms-13-979-2022


Mechanical Sciences
Open Access

# A real-time and accurate detection approach for bucket teeth falling off based on improved YOLOX

**Jinnan Lu and Yang Liu**

School of Mechanical Engineering, Liaoning Technical University, Fuxin 123000, China

**Correspondence:** Yang Liu (121185350@qq.com)

Received: 16 July 2022 – Revised: 12 October 2022 – Accepted: 9 November 2022 – Published: 25 November 2022

**Abstract.** An electric shovel is a bucket-equipped mining excavator widely used in open-pit mining today. The prolonged direct impact between the bucket teeth and the ore during the mining process will cause the teeth to loosen prematurely or even break, resulting in unplanned downtime and productivity losses. To solve this problem, we propose a real-time and accurate detection algorithm of bucket teeth falling off based on improved YOLOX. Firstly, to solve the problem of poor detection effect caused by uneven illumination, the dilated convolution attention mechanism is added to enhance the feature expression ability of the target in complex backgrounds so as to improve the detection accuracy of the target. Secondly, considering the high computing cost and large delay of the embedded device, the deep separable convolution is used to replace the traditional convolution in the feature pyramid network, and the model compression strategy is used to prune the redundant channels in the network, reduce the model volume, and improve the detection speed. The performance test is carried out on the self-constructed dataset of WK-10 electric shovel. The experimental results show that, compared with the YOLOX model, the mean average precision of the algorithm in this paper reaches 95.26 %, only 0.33 % lower, while the detection speed is 50.8 fps, 11.9 fps higher, and the model volume is 28.42 MB, which is reduced to 29.46 % of the original. Compared with many other existing methods, the target detection algorithm proposed in this paper has the advantages of higher precision, smaller model volume, and faster speed. It can meet the requirements of real-time and accurate detection of the bucket teeth falling off.

## 1 Introduction

With the continuous progress of science and technology, the mining of mineral resources is developing in a more complex, automated, and intelligent direction (Wang et al., 2018; Chen et al., 2017). Electric shovels are the main excavation equipment in mine production. The prolonged direct impact of the bucket teeth on ore during excavation can cause the bucket teeth to be subject to strong impact, friction, and bending force, which will cause the bucket teeth to loosen or even break unexpectedly from the teeth base, resulting in serious economic losses and waste of human and material resources. Firstly, the failure of the bucket teeth will increase the frictional resistance of the electric shovel during the excavation operation, making it difficult for the bucket to excavate, thereby affecting the efficiency of the excavation operation and accelerating the loss of the electric shovel (Akter and Basak, 2022). Secondly, since most of bucket teeth are made of high manganese steel or alloy steel, the hardness is much higher than that of ore and other materials. Once the broken bucket teeth are unloaded into the crusher together with ore, it will cause crusher failure and affect the whole mine crushing production line (Singla et al., 2014). The looseness or fracture of the bucket teeth is not detected in a timely and accurate manner, which will lead to serious safety hazards. Therefore, it is necessary to propose a high-precision, high real-time bucket teeth target detection algorithm.

In recent years, domestic and foreign scholars and research institutions have been trying to use traditional image processing algorithms to realise intelligent detection of the bucket teeth on existing video monitoring systems. For example, He et al. (2011) applied support vector machine (SVM) on the image's gray gradient histogram and applied structural feature constraints to the detection results based on the positional relationships between bucket teeth to improve the ac-

curacy of object detection. Alma'aitah and Hassanein (2014) used Sprouts sensor network platform to detect bucket teeth, and embedded wireless power transmission units and auxiliary antennas in bucket teeth to locate broken bucket teeth. Lim et al. (2016) determined the position of bucket teeth through the image samples of the bucket teeth movement process, used the optical flow method and frame difference method to track and correct the bucket teeth position, and finally used the image grayscale feature to judge the lack of bucket teeth. However, traditional algorithms are often based on the surface layer information of images, and need continuous parameter updates and template matching to obtain optimal features. Each feature template defines a specific task, which cannot meet the real-time and accurate detection requirement on the bucket teeth falling off.

With the continuous development of deep learning, breakthroughs have been made in the field of computer vision. The convolutional neural network has become an important means of image processing, and it has been introduced into the field of target detection, making the real-time and accurate detection of the bucket teeth falling off possible (Sun et al., 2022; Shen et al., 2022; Ji et al., 2022). Ji et al. (2021) proposed an intelligent monitoring system for missing bucket teeth, and used Faster R-CNN for bucket teeth target detection, which improved the detection accuracy of bucket teeth. Based on the DeepLabV3+ detection algorithm, Liu et al. (2021) completed the semantic segmentation of the image and obtained the profile information of the bucket teeth by improving the loss function and adding the attention mechanism. However, due to the deep network model and the large amount of parameters, it is difficult to deploy to embedded mobile devices. Therefore, under the premise of ensuring detection accuracy, it is very important to compress and lighten the convolutional neural network model (Zheng et al., 2022).

At present, the development of compression methods and lightweight models of convolutional neural networks are rapid. The model compression methods mainly include knowledge distillation, low-rank decomposition, quantisation, and channel pruning. Li et al. (2019) introduced knowledge distillation into convolutional neural networks to achieve model compression and distilled redundant network models to simple models; Denton et al. (2014) used approximate matrix products to replace the weight matrix of the convolutional neural networks, and decomposed the convolutional neural networks with low-rank, so as to reduce the amount of parameters. Liu et al. (2017) used the channel pruning strategy to prune redundant parameters of the convolutional layer, which effectively reduced the amount of computation and parameters while ensuring the integrity of the convolutional neural network. For lightweight models, single-shot multi-box detector (SSD) network structure based on deep feature fusion was proposed (Bai et al., 2022; Huang et al., 2022), which enhances the complementarity of high-level features and improves the detection performance of the SSD network for targets of different scales. Redmon et al. (2016) proposed the YOLO (you only look once) network, which takes the object detection process as a regression task and predicts multiple bounding boxes directly from the input image. In response to the problem that YOLOv1 only supports the same resolution of the input image with the training image, YOLOv2 (Redmon and Farhadi, 2017) improves the network structure and position prediction mechanism, which can adapt the model to multi-scale image input. YOLOv3 (J. Huang et al., 2021) uses a logical classifier to achieve multi-label classification but due to the insufficient feature extraction capability of the network itself, the accuracy of small target recognition is low. The subsequent improved YOLOv4 (Tan et al., 2021) on this basis uses the feature pyramid network structure to integrate the deep features and shallow features to solve the problem of small target recognition. In 2021, Ge et al. (2021) proposed a high-performance detection model YOLOX, which does not use anchor and dynamically matches samples for targets of different sizes. In the previous version of YOLO, the decoupling head was integrated; that is, classification and regression were implemented in a $1 \times 1$ convolution (L. Huang et al., 2021). YOLOX, it was believed, adversely affected the identification of the network, so in YOLOX, YoloHead was divided into two parts and implemented separately and finally integrated during the prediction phase. Compared with YOLOv3 and YOLOv4, the detection accuracy and speed of YOLOX model have been improved, the end-to-end deployment is more flexible, and the model size and memory occupation are more suitable for embedded mobile devices. However, the overall detection performance of YOLOX algorithm needs to be further improved.

In order to achieve real-time and accurate detection of bucket teeth falling off, an improved YOLOX algorithm is proposed in this paper. The main contributions and innovations of this paper are as follows: (1) firstly, the dilated convolution attention mechanism is added to enhance the saliency of the target in the complex background; secondly, the deep separable convolution is used to replace traditional convolution in the feature pyramid network; finally, the model compression strategy is used to prune redundant channels in the network, reduce model volume, and improve detection speed. (2) WK-10 and WK-35 electric shovel datasets were collected under real working conditions for training deep learning models. (3) Comparing multiple sets of experiments and evaluating the detection results, the mean average precision of our algorithm can reach 95.26 %, the detection speed is 50.8 fps, and the model volume is 28.42 MB. The experimental results on the migration dataset show that the improved model has strong generalisation ability.

## 2 Bucket teeth intelligent monitoring system

The bucket teeth intelligent monitoring system is mainly composed of image acquisition module, Ethernet commu-

nication module, detection module, and display module. As shown in Fig. 1, firstly, a top-view camera is installed under the crown wheel of the shovel boom and a front camera is installed on the rotary platform, and thus the image information of bucket teeth is collected to the greatest extent by using the complementary method of two monitoring spots. Then, it is transmitted to the embedded Xavier processor through Ethernet to detect bucket teeth based on the deep learning algorithm. Finally, the detection results are transmitted to the industrial touch all-in-one machine through UDP for display.

In view of the bad working environment of the electric shovel and the poor performance of the embedded equipment, a target detection algorithm that meets the requirements of accuracy and speed is selected, and the problem of bucket teeth falling off is converted into the problem of detecting the type, number, and position of the input image target.

## 3 YOLOX algorithm

YOLOX is an end-to-end target detection algorithm which adds many optimisation strategies on the basis of YOLOv3 to greatly improve the detection performance. Its network structure is shown in Fig. 2. YOLOX is divided into four main parts which are input, feature extraction backbone network, neck, and prediction module. The input uses two methods to enhance features: mosaic data enhancement and mixUP. The former one aims to improve the model detection capacity on small objects in the image. The input data are sliced by Focus before entering the feature extraction backbone network. The Focus structure performs self-copying and slicing operations on the image to obtain a down-sampling feature map with twice the information (Wang et al., 2022). The CBS (composed of convolution, batch normalisation, and SiLU activation function) module is used for feature extraction, and the cross-stage partial (CSP) structure is used to optimise the gradient information in the network, reduce the amount of inference calculation, and speed up the calculation of the network. The spatial pyramid pooling (SPP) layer performs feature extraction on the maximum pooling of different pooling kernel sizes, improving the receptive field of the network, and enhancing the nonlinear expression ability of the network. Neck improves the feature pyramid network + path aggregation network (FPN + PAN) structure to enhance the feature fusion of the network and the transmission of inference information in the network (Bello et al., 2021). Prediction replaces the original classification and regression layer couple head with decouple head, adopting anchor-free frame mechanism and simOTA label assignment method to solve the optimal transmission problem and reduce the amount of model parameters, which significantly improves the regression speed and accuracy of the network.

## 4 Improved YOLOX algorithm

The improved YOLOX network structure is shown in Fig. 3. First, the images are fed into the backbone network, and the feature maps are obtained after several convolution operations and pooling operations. Then, the obtained feature map is fed into the feature pyramid structure, and the feature extraction is enhanced by using the FPN structure and PAN structure, and the up-sampling and down-sampling operations are completed at the same time to combine the shallow feature map semantic information with the deep feature map semantic information. Finally, the target detection work is completed in the YoloHead structure.

### 4.1 Dilated CBAM (convolutional block attention module)

When the electric shovel is working, the background often has uneven illumination, which makes the target features inconspicuous. By adding an attention mechanism, the interference caused by the environment can be alleviated and the detection accuracy can be improved. At present, most of the existing attention mechanisms are channel or spatial attention mechanisms, which cannot make the network pay attention to the most contributing channel and the most contributing space at the same time. To take both into account, this paper adds a channel and spatial CBAM (convolutional block attention module) (Zhang et al., 2020). This module takes up less computation and can significantly improve the detection performance. However, in order to increase the network receptive field and improve the connection between target and background information, we change the convolutional layer to a dilated convolutional layer. However, Yu et al. (2017) pointed out that dilated convolution can cause gridding artifacts when the frequency of feature maps is higher than the sampling frequency of dilated convolution. To remove the gridding artifacts, two convolutional layers with a smaller dilation rate are added after the convolutional layer with dilation rate of 4; the structure is shown in Fig. 4. Among them, a dilated CBAM module is added after each CSP module, and up-sampling and down-sampling to calculate the weight information of the feature map in the channel and spatial position. According to the weight distribution, the ability of the network to distinguish target from background is improved, and a good detection result is achieved in practical applications.

The dilated CBAM module first sends the input features $F$ into the channel attention module, which obtains the information of each channel through average pooling and max pooling, superimposes the obtained parameters through a MLP (multi-layer perceptron) to generate different channel attention map, and activates by the sigmoid function to the final channel attention map. The calculation formula is shown as Eq. (1):

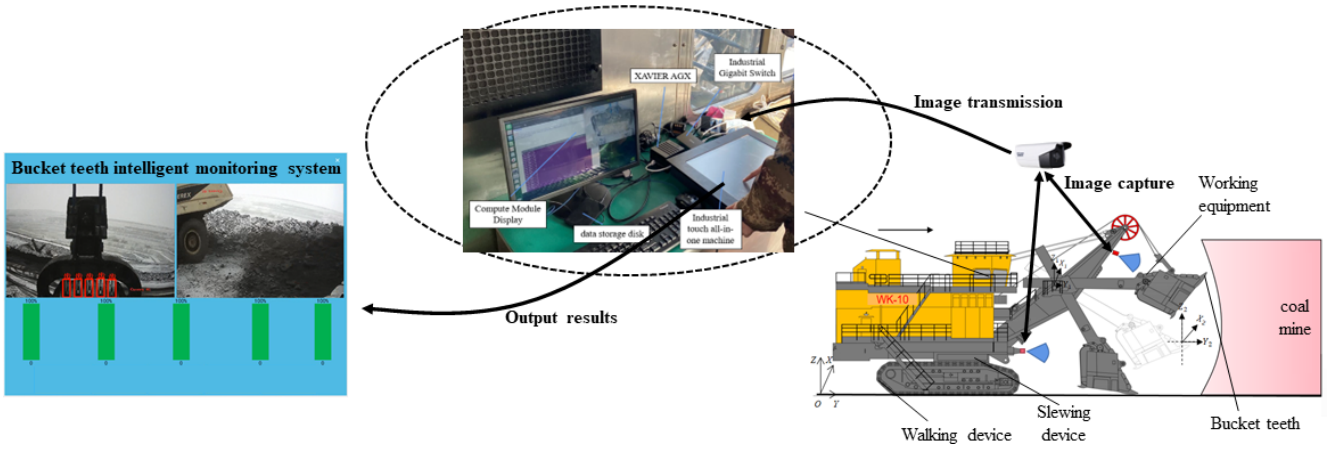


Figure 1. Intelligent monitoring system of bucket teeth.

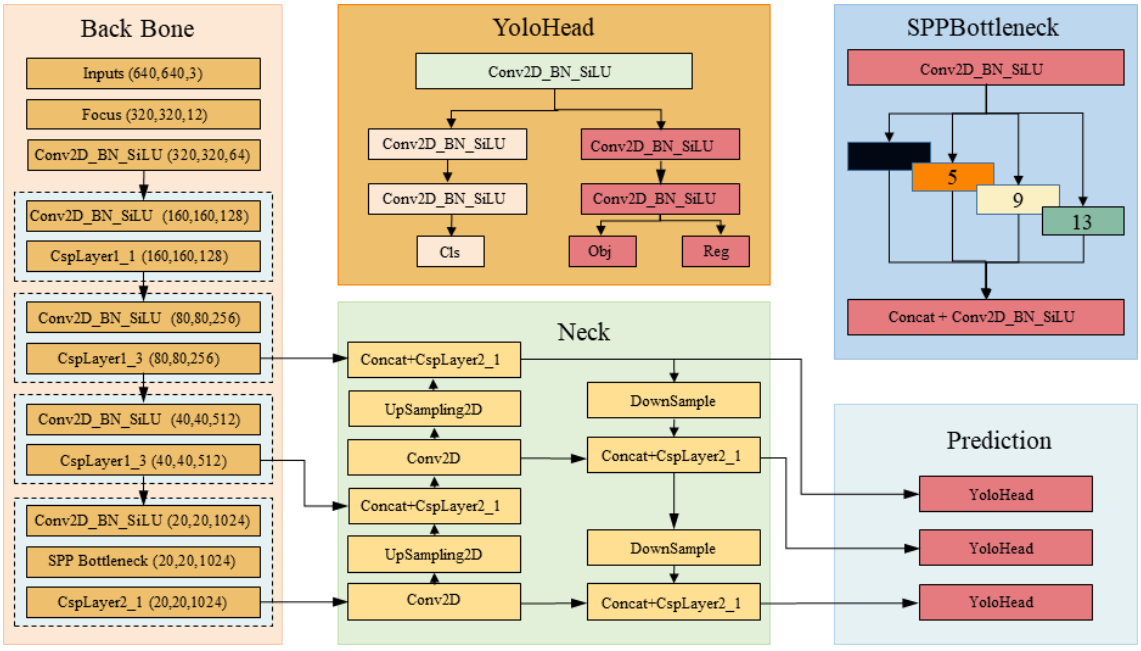


Figure 2. YOLOX network structure.

$$M_c(F) = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MLP}(\mathrm{MaxPool}(F))) = \sigma\left(W_1\left(W_0\left(F^c_{avg}\right)\right) + W_1\left(W_0\left(F^c_{max}\right)\right)\right), \quad (1)$$

where $\sigma(\cdot)$ is the sigmoid nonlinear activation function; $W_0$ and $W_1$ are the hidden layer weights and output layer weights, respectively; and $F^c_{avg}$ and $F^c_{max}$ are the average pooling and max pooling operations of the channel attention module, respectively.

After the feature map $F_x$ is sent to the spatial attention module, the spatial information is gathered along the channel dimension through average pooling and max pooling to generate a spatial feature map. Then, through the dilated convolution calculations with dilation rates of 4, 2, and 1, and activation of the sigmoid function, the spatial attention feature is obtained. The calculation formula is shown as Eq. (2):

$$M_s(F_x) = \sigma\left(f_4^{3\times3}\left(f_2^{3\times3}\left(f_1^{3\times3}(\mathrm{AvgPool}(F_x); \mathrm{MaxPool}(F_x))\right)\right)\right) = \sigma\left(f_4^{3\times3}\left(f_2^{3\times3}\left(f_1^{3\times3}\left(F^s_{avg}; F^s_{max}\right)\right)\right)\right), \quad (2)$$

where $f_4^{3\times3}$, $f_2^{3\times3}$, and $f_1^{3\times3}$ are the dilated convolution calculations with the convolution kernel size of $3 \times 3$ and dilation rates of 4, 2, and 1, respectively; and $F^s_{avg}$ and $F^s_{max}$ are

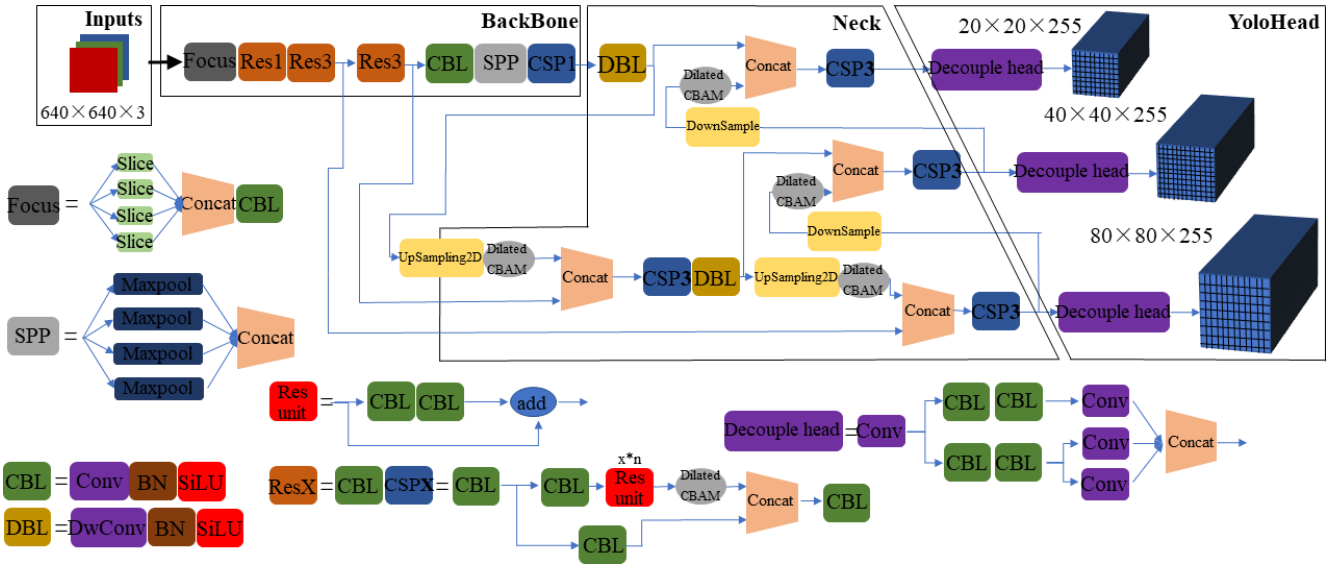


Figure 3. Improved YOLOX network structure.

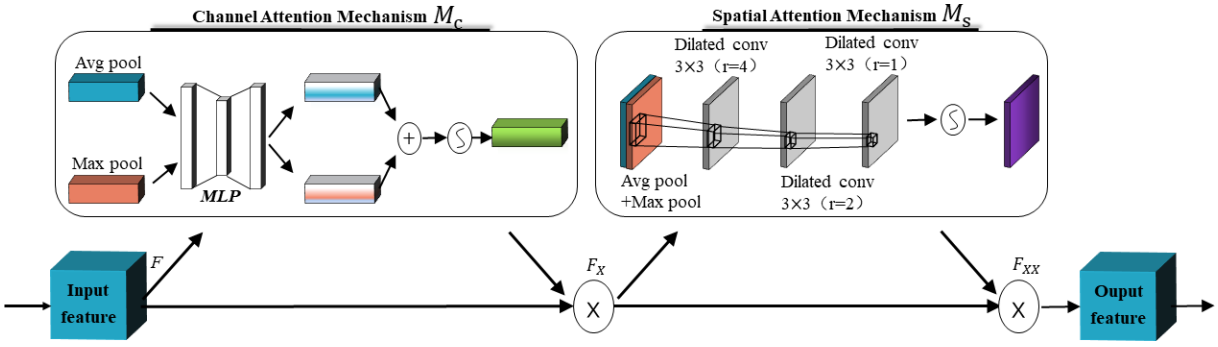


Figure 4. Dilated CBAM module.

the average pooling and max pooling operations of the spatial attention module, respectively. In addition, a dilated CBAM module is added after CSP unit in the CSPX structure; its structure is shown in Fig. 5.

## 4.2 Deep separable convolution

Deep separable convolution (Buiu et al., 2020; Yun et al., 2022) is the core of the lightweight network MobileNet, which can achieve the same feature extraction effect as traditional convolution. Replacing the traditional convolution in PANet (path aggregation network) with the deep separable convolution can improve the problem that the running speed cannot be improved due to the high amount of computation in the traditional convolution, and at the same time reduce the model volume and network parameters. Its structure is shown in Fig. 6.

The deep separable convolution decomposes the convolution into one channel-by-channel convolution and one point-by-point convolution, which reduces the computation and

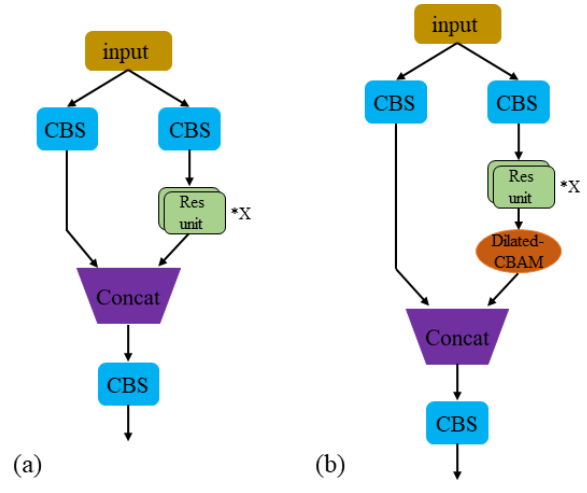


Figure 5. Improvement of CSPX. **(a)** CSPX of the original model. **(b)** CSPX of the algorithm in this paper.

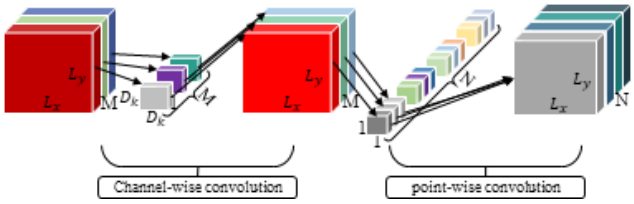


**Figure 6.** Deep separable convolution structure.

model volume of the network and significantly improves the detection speed of the network. It is assumed that the input feature map size is $L_x \times L_y \times M$, the output feature map size is $L_x \times L_y \times N$, and the convolution kernel size is $D_k$. Among them, $L_x$ and $L_y$ are the width and height of the feature map, $M$ and $N$ are the number of input and output channels, and the calculation of the traditional convolution is shown as Eq. (3):

$$L_x \times L_y \times M \times N \times D_k \times D_k. \tag{3}$$

The calculations of channel-by-channel convolution and point-by-point convolution are

$$L_x \times L_y \times M \times D_k \times D_k \tag{4}$$
$$L_x \times L_y \times M \times N. \tag{5}$$

Combining Eqs. (4) and (5), the deep separable convolution is calculated as

$$L_x \times L_y \times M \times D_k \times D_k + L_x \times L_y \times M \times N. \tag{6}$$

The reduction ratio of the deep separable convolution compared to traditional convolution is

$$\frac{L_x \times L_y \times M \times D_k \times D_k + L_x \times L_y \times M \times N}{L_x \times L_y \times M \times N \times D_k \times D_k} = \frac{1}{N} + \frac{1}{D_k^2}. \tag{7}$$

### 4.3 Network layer model compression

Due to the large volume of the current convolutional neural network models, they suffer from the problems of high computing power cost and large delay in many practical projects. Using model compression, the redundant parts of the network and unimportant channels can be automatically identified and pruned in the training process, so as to obtain a compact model with high accuracy (Yu et al., 2021).

#### 4.3.1 Sparse training

Model channel sparse training can distinguish important channels and unimportant channels as a preparation for pruning unimportant channels. An indicator needs to be set for determining the importance of the channel, and BN (batch normalisation) layer has the functions of moving parameters and channel scaling, so the parameters in the BN layer can be used as the evaluation indicator. Chen et al. (2020) pointed out that the parameter $\gamma$ in the BN layer directly affects the output $y$, and the smaller the value of $\gamma$, the less important the channel information is. Therefore, this paper takes the parameter $\gamma$ as the quantification indicator of channel importance and calls it the channel scaling factor. The loss function of the channel pruning algorithm for introducing the scaling factor $\gamma$ is shown as Eq. (8):

$$L = \sum_{(x,y)} l(f(x, W), y) + s \sum_{\gamma \in \Gamma} g(\gamma), \tag{8}$$

where the first term is the training loss of YOLOX, $x$ and $y$ are the input and output of training, respectively, $W$ is the training parameter of the network, and the second term is the L1 regular constraint term of the BN layer parameter $\gamma$, which is the penalty caused by the sparsity of the scaling factor. It takes $g(\gamma) = |\gamma|$ in this paper, and $s$ balances these two terms as a penalty factor set in this experiment.

#### 4.3.2 Channel pruning and fine-tuning

After model channel sparse training, a model with many scaling factors close to zero is obtained, as shown in Fig. 7. Assuming that the threshold value is set to 0.1, the convolution layers below the threshold value are shown in yellow and the convolution layers higher than the threshold value are shown in blue. By pruning the input and output connections of the yellow convolution layers and the corresponding weights, the model volume can be reduced and the model integrity can be guaranteed. However, when the set threshold value is too large, the accuracy of the pruned model will be greatly reduced, so it is necessary to use fine-tuning to restore the accuracy of the pruned network. In many cases of model compression, the fine-tuned narrow network can even obtain higher accuracy than the original unpruned network.

## 5 Experiment and result analysis

The software and hardware configurations used in the experiment are shown in Table 1.

### 5.1 Dataset preparation

In this paper, industrial cameras are used to collect samples, which have more light input than ordinary surveillance cameras. In the open-air low-light environment, those cameras can also present clear pictures. At the same time, angle-adjustable camera brackets and mixed fill lights are equipped, which is convenient to adjust the recording angle and fill light at night, ensuring that the best image can be obtained even in bad working conditions. A total of 1200 key frame images are collected in this experiment including 500 at night and 700 during the day. The image resolution is $1280 \times 960$.

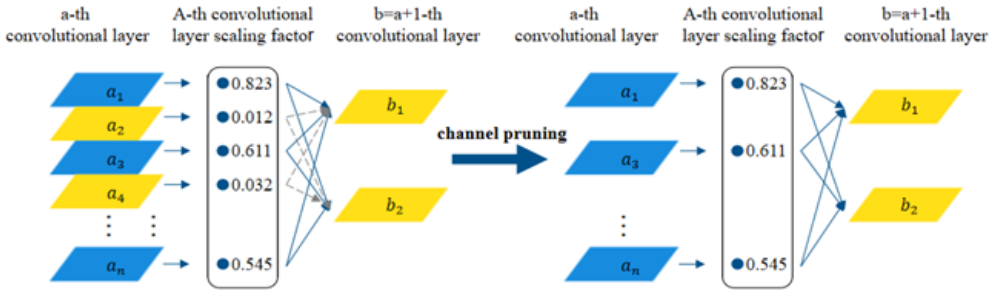


**Figure 7.** Channel pruning strategy.

**Table 1.** Software and hardware configurations.

| Configurations | Version parameters |
|---|---|
| Operating system | Ubuntu18.04LTS |
| GPU | NVIDIA GeForce RTX 3080 Ti |
| CPU | Intel Core i5-10400F@2.90GHz × 6 CPUs |
| Deep learning framework | Pytorch |

The sample categories cover bucket teeth, bucket, pedestrian, dump truck, patrol vehicle, and other targets in the working scene. Traditional data enhancement methods are used for image rotation, mosaic data enhancement, and Gaussian filtering, and thus the dataset is expanded to 3600 images. Among them, 70 % of 3600 images are used as training sample set, 20 % are used as validation sample set, and 10 % are used as the test sample set. Some dataset examples are shown in Fig. 8.

## 5.2 Model evaluation metrics

In order to verify the performance of the model, this experiment uses three indicators – mAP (mean average precision), model volume, and FPS (frames per second) – as the evaluation standard. The calculation formula of mAP is shown as Eq. (9):

$$\mathrm{mAP} = \frac{\sum_{n=1}^{n} \mathrm{AP}(n)}{n} \times 100\,\%, \tag{9}$$

where $n$ (taken as 9) is the number of detected categories and mAP is the average value of detection accuracy of all categories. The larger the value, the higher the detection accuracy is. Among them, AP is numerically equal to the area enclosed by the $P$–$R$ curve and the coordinate axis.

The calculation formulas of $P$ and $R$ are as follows:

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \times 100\,\% \tag{10}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \times 100\,\%, \tag{11}$$

where $P$ is the precision, $R$ is the recall rate, and TP, FP, and FN is the true positive, false positive, and false negative, respectively.

## 5.3 Model improvement effectiveness

### 5.3.1 Experimental analysis of YOLOX-CD (YOLOX-dilated cbam + deep separable convolution)

The initial learning rate is set to 0.001, the momentum coefficient is 0.949, the number of iterations is 4000, and the learning rate is changed to 0.0001 and 0.00001 at 3200 and 3600 iterations, respectively. The loss function value change curves of model training are shown in Figs. 9 and 10. It can be seen from the figure that the loss value of YOLOX gradually decreased to 1.2 after 3300 iterations, and finally stabilised to about 1. However, the loss value of YOLOX-CD is lower and the convergence speed is faster, the loss value continued to stabilise to about 0.7 at 2800 iterations, and the model shows a convergence trend.

After the model training is completed, the test sample set is input into each model for testing, and the test result of YOLOX (baseline), YOLOX-C with the dilated convolution attention, YOLOX-D with the deep separable convolution introduced in PANet, and YOLOX-CD model with both are compared, so as to verify the effectiveness of the model improvement. The experimental results are shown in Table 2.

It can be clearly seen from Table 2 that after introducing both the dilated convolution attention mechanism and the deep separable convolution, the performances are significantly improved compared with the original model. Among them, the mAP is increased by 0.83 %, the detection speed is increased by 2.4 fps, and the model volume is reduced by 6.34 MB.

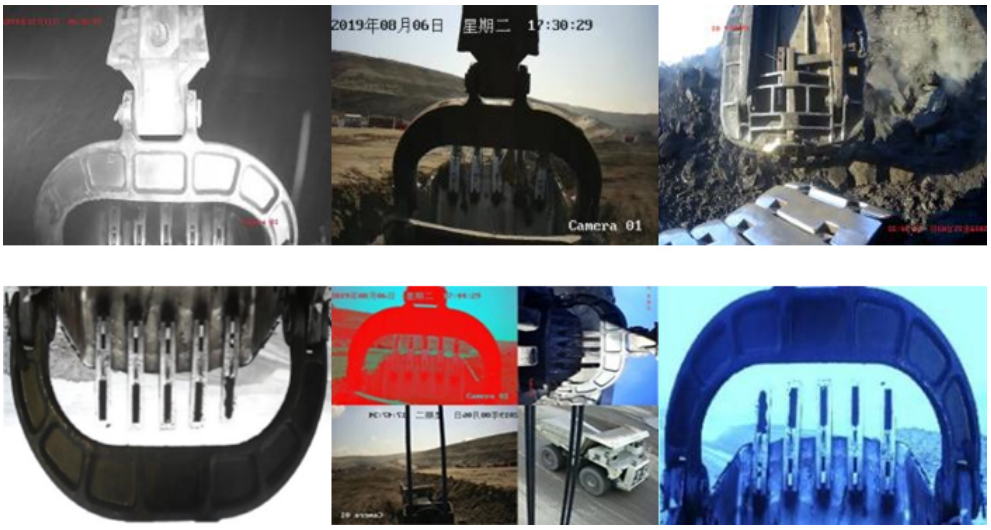


Figure 8. Dataset examples.

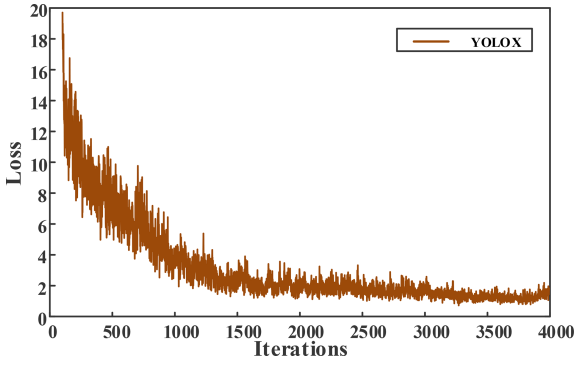


Figure 9. Loss function value change curve of YOLOX.

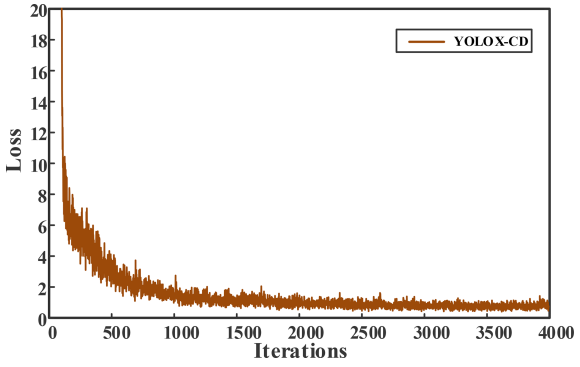


Figure 10. Loss function value change curve of YOLOX-CD.

### 5.3.2 Experimental analysis of model compression

Under the same conditions as the initial training environment, the penalty factor $s$ is set as 0.0001, 0.001, 0.005, and 0.01, respectively, and the optimal penalty factor is selected by comparing the model performance during the sparse training process. Figure 11 shows the mAP change curve under different penalty factors. It can be seen from the figure that the detection performance of the YOLOX-CD reaches the best when the penalty factor is $s = 0.001$. Therefore, $s = 0.001$ is selected as the penalty factor for the final channel sparse training in this paper.

After sparse training, YOLOX-CD is pruned, and the model is named YOLOX-CDP. During the pruning process, the mAP of YOLOX-CDP will decrease, which needs to be

Table 2. Comparison of experimental results.

| Model name | mAP (%) | Model volume (MB) | FPS |
|---|---|---|---|
| YOLOX | 95.59 | 96.46 | 38.9 |
| YOLOX-C | 97.61 | 99.27 | 35.2 |
| YOLOX-D | 94.44 | 87.69 | 44.5 |
| YOLOX-CD | 96.42 | 90.12 | 41.3 |

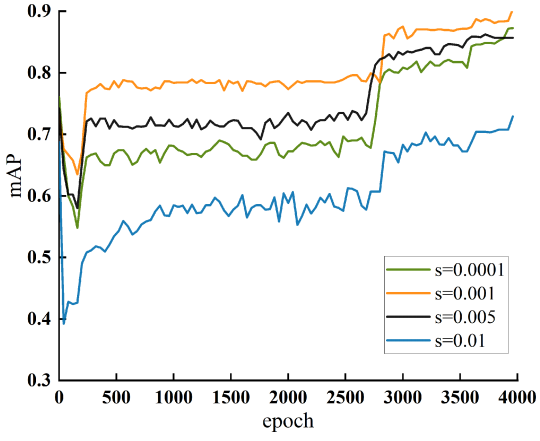


Figure 11. mAP curve under different penalty factors.

**Table 3.** Model performance comparison with different pruning rates.

| Model name | mAP (%) | Model volume (MB) | FPS |
|---|---|---|---|
| YOLOX | 95.59 | 96.46 | 38.9 |
| YOLOX-CD-0.5 | 96.35 | 47.73 | 45.1 |
| YOLOX-CD-0.7 | 95.26 | 28.42 | 50.8 |
| YOLOX-CD-0.9 | 56.72 | 8.34 | 57.2 |

**Table 4.** Performance comparison of different algorithms.

| Model name | mAP (%) | Model volume (MB) | FPS |
|---|---|---|---|
| Faster R-CNN | 93.34 | 522.91 | 13.0 |
| YOLOv3 | 89.73 | 236.20 | 31.2 |
| YOLOv4 | 92.06 | 245.00 | 35.7 |
| YOLOv5 | 94.82 | 89.54 | 43.0 |
| OURS | 95.26 | 28.42 | 50.8 |

fine-tuned. Among them, the pruning rate is set to 0.5, 0.7, and 0.9, respectively, and 20 training epochs are carried out in fine-tuning. The model performance with different pruning rates are shown in Table 3.

In Table 3, as the pruning rate increases, both the model volume and the prediction time decrease continuously. Compared with the YOLOX model, although the mAP of YOLOX-CD-0.7 decreased by 0.33 %, the detection speed increased by 11.9 fps, and the model volume decreased to 29.46 % of the original YOLOX. The experimental results show that the model compression method can reduce the model volume and improve the detection speed under the premise of ensuring high precision. To more intuitively reflect the effect of YOLOX-CD-0.7 model bucket teeth falling off real-time and accurate detection, samples are selected by classification in the test set to carry out multiple groups of comparative experiments. The experimental results are shown in Fig. 12.

Figure 12a shows the detection results of the top and bottom video images in the normal working scene of the electric shovel, including top-view excavation, bottom-view excavation, top-view loading, and bottom-view loading. It can be seen from the detection results that the YOLOX-CD-0.7 model can effectively detect targets such as bucket teeth, bucket, dump truck, etc. and the accuracy of the framed range is high. Figure 12b shows the detection results of dangerous scenarios that may be encountered during the operation of the electric shovel, including occlusion detection, broken teeth detection, vehicle intrusion, and personnel intrusion. As can be seen from the figure, the algorithm proposed in this paper can not only judge whether the bucket teeth fall off but also identify the invading vehicles and personnel, which can effectively prevent the occurrence of dangerous accidents. Figure 12c shows the image detection results in complex and illuminated scenes. As can be seen from the figure, even under complex and uneven lighting conditions, the algorithm in this paper can still detect small, weak, and occluded targets well.

This paper takes the dataset of the WK-35 electric shovel as an example to carry out the migration experiment. Samples under different scenarios are randomly selected from the dataset for testing; the test results are shown in Fig. 13. It can be seen from the experimental results that for the untrained new sample data, the original model cannot accurately identify the target in complex scenes, while the improved model can complete the real-time and accurate detection of targets in various complex scenes, which proves that improved YOLOX has good generalisation ability.

### 5.4 Comparison experiment of different algorithms

In order to further verify the advantages of the algorithm in this paper, it is compared with four classic target detection algorithms of Faster R-CNN, YOLOv3, YOLOv4, and YOLOv5 in the same dataset and training environment. The comparison results are shown in Fig. 14.

It can be seen from the figure that the $P-R$ curve of our algorithm is located at the top of all the curves, and the performance of detecting bucket teeth is better than that of the other four algorithms. The above experimental results are listed in Table 4. It can be seen that the algorithm in this paper has improved mAP by 1.92 %, 5.53 %, 3.2 %, and 0.44 %, respectively, compared with Faster R-CNN, YOLOv3, YOLOv4, and YOLOv5, with smaller model volume and faster detection speed. Comprehensive performance of all aspects show that improved YOLOX can meet the requirements of real-time and accurate detection of bucket teeth falling off during the working process of the electric shovel.

## 6 Conclusions

We propose a real-time and accurate detection algorithm of bucket teeth falling off based on improved YOLOX. Through the field experiment on the working scene of electric shovel in a large open pit mine, the following conclusions are drawn.

Adding the dilated CBAM attention mechanism solves the problem of weak target saliency in complex environments and improves the target detection accuracy. The deep separable convolution is used to replace the traditional convolution in PANet, which reduces the overall computation of the network. The channel pruning strategy is used to compress the model, which significantly improves the network detection speed and reduces the model volume on the basis of ensuring the integrity of the model. The experimental results show that the improved YOLOX detection accuracy is 95.26 %, only 0.33 % lower, while the detection speed is 50.8 fps, 11.9 fps higher, and the model volume is 28.42 MB, which is reduced to 29.46 % of the original.

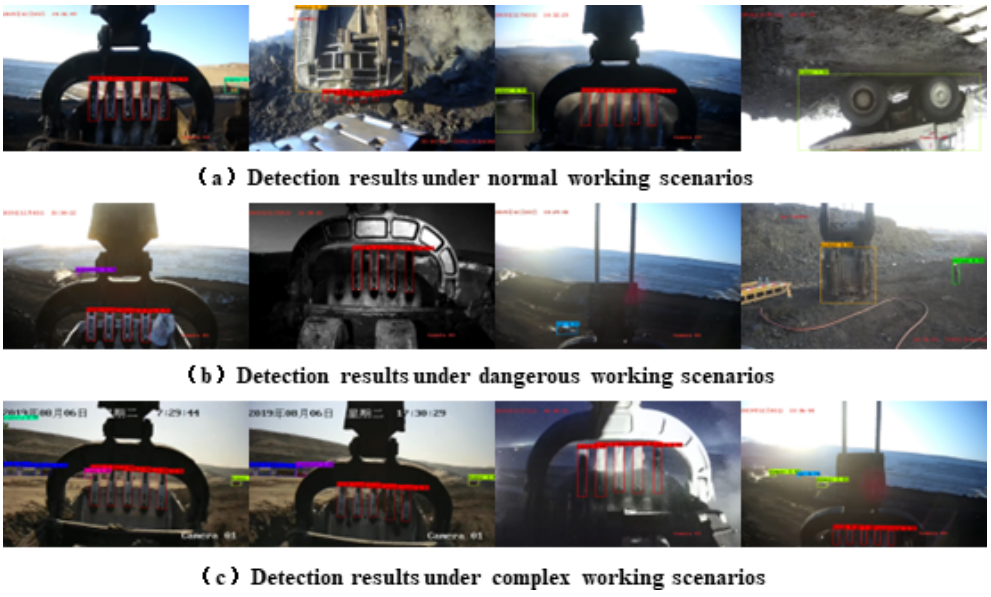


**Figure 12.** The detection results of YOLOX-CD-0.7 on the test sample set.

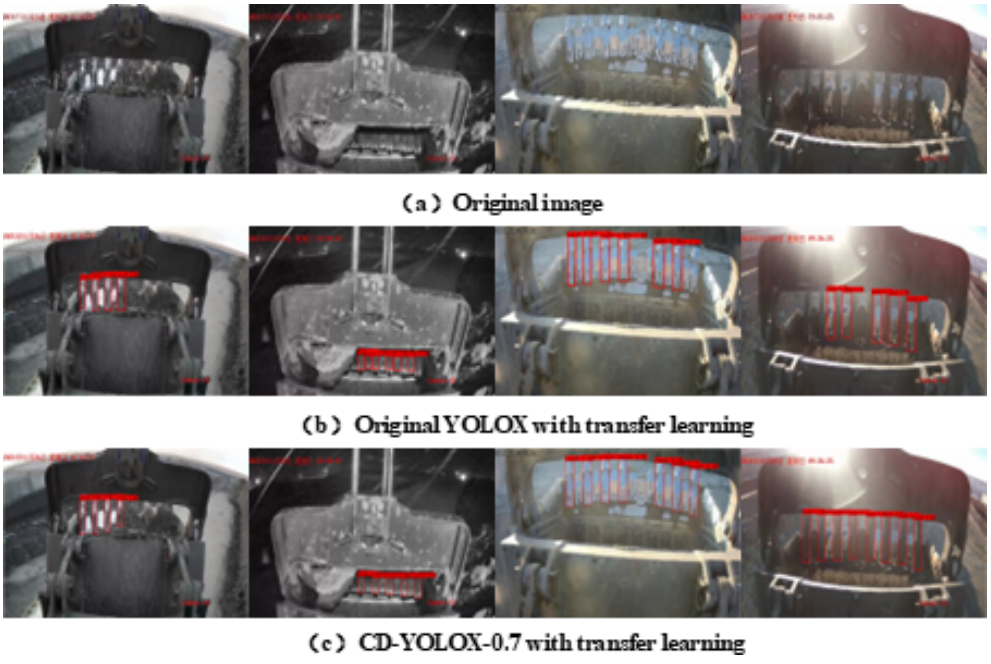


**Figure 13.** Comparison of migration test results.

Compared with Faster R-CNN, YOLOv3, YOLOv4, and YOLOv5 on the self-built WK-10 electric shovel dataset, the improved YOLOX algorithm has great advantages in detection accuracy, speed, and model volume, and is suitable for deployment in embedded mobile devices for detection and calculation.

The migration experiment results show that the improved model has strong generalisation ability and can complete the real-time and accurate detection of targets of different types of electric shovels in complex scenarios, which provides a theoretical basis and technical support for the development of intelligent mines and mining intelligence.

The working environment of the open pit mine is harsh, which will produce obvious fog phenomenon, affecting the detection performance of the algorithm in this paper. In future work, image defogging module is introduced to improve the detection accuracy of our algorithm in a fog environment.

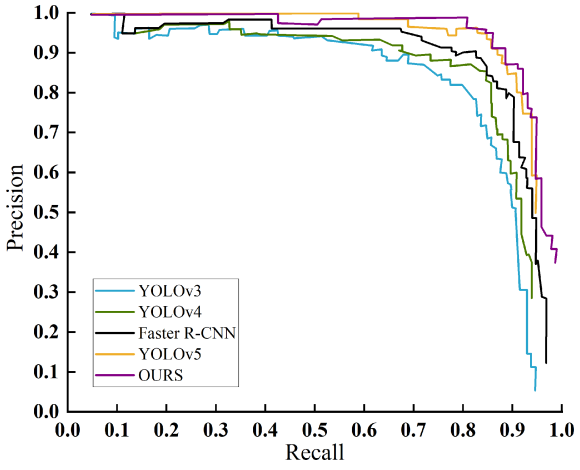


**Figure 14.** *P*–*R* curves of different algorithms.

**Data availability.** All the code/data used in this paper can be obtained upon request to the corresponding author.

**Author contributions.** JL provided the ideas, reviewed the overall process for the work, and revised the paper. YL provided the program, made the figures, and wrote the majority of the paper.

**Competing interests.** The contact author has declared that neither of the authors has any competing interests.

**Disclaimer.** Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Acknowledgements.** The authors would like to thank the National Natural Science Foundation of China for supporting this research. We also greatly appreciate the efforts of the reviewers and our colleagues.

**Financial support.** This research has been supported by the National Natural Science Foundation of China (grant nos. 51774162 and 51874158).

**Review statement.** This paper was edited by Zi Bin and reviewed by three anonymous referees.

## References

Akter, A. and Basak, H.: Design and analysis of biomimetics based excavator bucket and tooth, P. I. Mech. Eng. E-J. Pro, 236, 1167–1175, https://doi.org/10.1177/09544089211057645, 2022.

Alma'aitah, A. and Hassanein, H.: Utilizing Sprouts WSN platform for equipment detection and localization in harsh environments, in: 39th Annual IEEE Conference on Local Computer Networks Workshops, 8–11 September 2014, Edmonton, Canada, 777–783, https://doi.org/10.1109/lcnw.2014.6927734, 2014.

Bai, D., Sun, Y., Tao, B., Tong, X., Xu, M., Jiang, G., Chen, B., Cao, Y., Sun, N., and Li, Z.: Improved single shot multibox detector target detection method based on deep feature fusion, Concurr. Comp.-Pract. E., 34, e6614, https://doi.org/10.1002/CPE.6614, 2022.

Bello, R., Mohamed, A., and Talib, A.: Enhanced Mask R-CNN for herd segmentation, Int. J. Agr. Biol. Eng., 14, 238–244, https://doi.org/10.25165/j.ijabe.20211404.6398, 2021.

Buiu, C., Danaila, V., and Raduta, C.: MobileNetV2 ensemble for cervical precancerous lesions classification, Processes, 8, 595, https://doi.org/10.3390/pr8050595, 2020.

Chen, S., Zhan, R., Wang, W., and Zhang, J.: Learning Slimming SAR Ship Object Detector Through Network Pruning and Knowledge Distillation, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 1267–1282, https://doi.org/10.1109/jstars.2020.3041783, 2020.

Chen, W., Li, X., He, H., and Wang, L.: A Review of Fine-Scale Land Use and Land Cover Classification in Open-Pit Mining Areas by Remote Sensing Techniques, Remote Sens., 10, 15, https://doi.org/10.3390/rs10010015, 2017.

Denton, E., Zaremba, W., Bruna, J., LeCun, Y., and Fergus, R.: Exploiting Linear Structure Within Convolutional Networks for Efficient Evaluation, in: Advances in Neural Information Processing Systems, 8–13 December 2014, Montreal, Canada, https://arxiv.org/abs/1404.0736, 2014.

Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J.: YOLOX: Exceeding yolo series in 2021, arXiv [preprint], https://doi.org/10.48550/arxiv.2107.08430, 6 August 2021.

He, L., Wang, H., and Zhang, H.: Object detection by parts using appearance, structural and shape features, in: 2011 IEEE International Conference on Mechatronics and Automation, Beijing, China, 7–10 August 2011, 489–494, https://doi.org/10.1109/icma.2011.5985611, 2011.

Huang, J., Zhang, H., Wang, L., Zhang, Z., and Zhao, C.: Improved YOLOv3 Model for miniature camera detection, Opt. Laser Technol., 142, 107133, https://doi.org/10.1016/j.optlastec.2021.107133, 2021.

Huang, L., Fu, Q., He, M., Jiang, D., and Hao, Z.: Detection algorithm of safety helmet wearing based on deep learning, Concurr. Comp.-Pract. E., 33, e6234, https://doi.org/10.1002/cpe.6234, 2021.

Huang, L., Chen, C., Yun, J., Sun, Y., Tian, J., Hao, Z., Yu, H., and Ma, H.: Multi-Scale Feature Fusion Convolutional Neural Network for Indoor Small Target Detection, Frontiers in Neurorobitics, 16, 881021, https://doi.org/10.3389/fnbot.2022.881021, 2022.

Ji, S., Li, W., Zhang, B., Zhou, L., and Duan, C.: Bucket Teeth Detection Based on Faster Region Convolutional Neural Network, IEEE Access, 9, 17649–17661, https://doi.org/10.1109/access.2021.3054436, 2021.

Ji, Y., Lu, Q., and Yao, Q.: Short communication: A case study of stress monitoring with non-destructive stress measurement and deep learning algorithms, Mech. Sci., 13, 291–296, https://doi.org/10.5194/ms-13-291-2022, 2022.

Li, Z., Xu, X., Xie, L., and Su, H.: Learning Slimming SSD through Pruning and Knowledge Distillation, in: 2019 Chinese Automation Congress, 22–24 November 2019, Hangzhou, China, 2701–2705, https://doi.org/10.1109/cac48633.2019.8996995, 2019.

Lim, S., Soares, J., and Zhou, N.: Tooth guard: A vision system for detecting missing tooth in rope mine shovel, in: 2016 IEEE Winter Conference on Applications of Computer Vision, 7–10 March 2016, New York, United States, 1–7, https://doi.org/10.1109/wacv.2016.7477594, 2016.

Liu, X., Qi, X., and Jiang, Y.: Electric Shovel Teeth Missing Detection Method Based on Deep Learning, Comput. Intel. Neurosc., 2021, 6503029, https://doi.org/10.1155/2021/6503029, 2021.

Liu, Z., Li, J., Shen, Z., Huang, G., Yan, S., and Zhang, C.: Learning efficient convolutional networks through network slimming, in: 2017 IEEE International Conference on Computer Vision, 22–29 October 2017, Venice, Italy, 2736–2744, https://doi.org/10.1109/iccv.2017.298, 2017.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, 27–30 June 2016, Washington, United States, 779–788, https://doi.org/10.1109/cvpr.2016.91, 2016.

Redmon, J. and Farhadi, A.: YOLO9000: Better, Faster, Stronger, in: 30th IEEE Conference on Computer Vision and Pattern Recognition, 21–26 July 2017, Honolulu, United States, 6517–6525, https://doi.org/10.1109/cvpr.2017.690, 2017.

Shen, X., Huang, Q., and Xiong, G.: Modelling and predictive investigation on the vibration response of a propeller shaft based on a convolutional neural network, Mech. Sci., 13, 485–494, https://doi.org/10.5194/ms-13-485-2022, 2022.

Singla, S., Kang, A., Grewal, J., and Cheema, G.: Wear behavior of weld overlays on excavator bucket teeth, Proc. Mat. Sci., 5, 256–266, https://doi.org/10.1016/j.mspro.2014.07.265, 2014.

Sun, Y., Zhao, Z., Jiang, D., Tong, X., Tao, B., Jiang, G., Kong, J., Yun, J., Liu, X., Zhao, G., and Fang, Z.: Low-illumination image enhancement algorithm based on improved multi-scale Retinex and ABC algorithm optimization, Frontiers in Bioengineering and Biotechnology, 10, 865820, https://doi.org/10.3389/fbioe.2022.865820, 2022.

Tan, L., Lv, X., Lian, X., and Wang, G.: YOLOv4_Drone: UAV image target detection based on an improved YOLOv4 algorithm, Comput. Electr. Eng., 93, 107261, https://doi.org/10.1016/j.compeleceng.2021.107261, 2021.

Wang, G., Zheng, H., and Zhang, X.: A robust checkerboard corner detection method for camera calibration based on improved YOLOX, Front. Phys., 9, 819019, https://doi.org/10.3389/fphy.2021.819019, 2022.

Wang, X., Sun, W., Li, E., and Song, X.: Energy-minimum optimization of the intelligent excavating process for large cable shovel through trajectory planning, Struct. Multidiscip. O., 58, 2219–2237, https://doi.org/10.1007/s00158-018-2011-6, 2018.

Yu, F., Koltun, V., and Funkhouser, T.: Dilated residual networks, in: 30th IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, United States, 21–26 July 2017, 636–644, https://doi.org/10.1109/cvpr.2017.75, 2017.

Yu, Z., Shen, Y., and Shen, C.: A real-time detection approach for bridge cracks based on YOLOv4-FPM, Automat. Constr., 122, 103514, https://doi.org/10.1016/j.autcon.2020.103514, 2021.

Yun, J., Jiang, D., Liu, Y., Sun, Y., Tao, B., Kong, J., Tian, J., Tong, X., Xu, M., and Fang, Z.: Real-time target detection method based on lightweight convolutional neural network, Frontiers in Bioengineering and Biotechnology, 10, 861286, https://doi.org/10.3389/fbioe.2022.861286, 2022.

Zhang, Y., Huang, J., and Cai, F.: On Bridge Surface Crack Detection Based on an Improved YOLO v3 Algorithm, IFAC-PapersOnLine, 53, 8205–8210, https://doi.org/10.1016/j.ifacol.2020.12.1994, 2020.