applied sciences

MDPI

Article

Real-Time Obstacle Detection Method in the Driving Process of Driverless Rail Locomotives Based on DeblurGANv2 and Improved YOLOv4

Wenshan Wang, Shuang Wang *, Yanqiu Zhao *, Jiale Tong, Tun Yang and Deyong Li

School of Mechanical Engineering, Anhui University of Science and Technology, Huainan 232001, China
* Correspondence: swang@aust.edu.cn (S.W.); 2020100054@aust.edu.cn (Y.Z.);
Tel.: +86-0554-6668949 (S.W. & Y.Z.)

Abstract: In order to improve the detection accuracy of an algorithm in the complex environment of a coal mine, including low-illumination, motion-blur, occlusions, small-targets, and background-interference conditions; reduce the number of model parameters; improve the detection speed of the algorithm; and make it meet the real-time detection requirements of edge equipment, a real-time obstacle detection method in the driving of driverless rail locomotives based on DeblurGANv2 and improved YOLOv4 is proposed in this study. A blurred image was deblurred using DeblurGANv2. The improved design was based on YOLOv4, and the lightweight feature extraction network MobileNetv2 was used to replace the original CSPDarknet53 network to improve the detection speed of the algorithm. There was a high amount of background interference in the target detection of the coal mine scene. In order to strengthen the attention paid to the target, the SANet attention module was embedded in the Neck network to improve the detection accuracy of the algorithm under low-illumination, target-occlusion, small-target, and other conditions. To further improve the detection accuracy of the algorithm, the K-means++ algorithm was adopted to cluster prior frames, and the focal loss function was introduced to increase the weight loss of small-target samples. The experimental results show that the deblurring of the motion-blurred image can effectively improve the detection accuracy of obstacles and reduce missed detections. Compared with the original YOLOv4 algorithm, the improved YOLOv4 algorithm increases the detection speed by 65.85% to 68 FPS and the detection accuracy by 0.68% to 98.02%.

Keywords: YOLOv4 algorithm; Mobilenetv2; SA attention mechanism; focal loss function; K-means++; DeblurGANv2

check for updates

Citation: Wang, W.; Wang, S.; Zhao, Y.; Tong, J.; Yang, T.; Li, D. Real-Time Obstacle Detection Method in the Driving Process of Driverless Rail Locomotives Based on DeblurGANv2 and Improved YOLOv4. *Appl. Sci.* **2023**, *13*, 3861. https://doi.org/10.3390/app13063861

Academic Editor: João M. F. Rodrigues

Received: 1 February 2023
Revised: 9 March 2023
Accepted: 14 March 2023
Published: 17 March 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

At present, China's coal mining industry has gradually progressed from mechanization and automation to intelligence, and the development goal of intelligent mines and unmanned pits has become the consensus of the industry. With unmanned mining activities, coal mine auxiliary transportation has become a bottleneck restricting unmanned coal mines, and it is frequently linked to sporadic, underground accidents. The auxiliary transportation system urgently needs to develop in the direction of standardization, intelligence, and being unmanned [1–3]. The coal mine rail locomotive is one of the main methods of auxiliary transportation, and the realization of an unmanned rail locomotives is of great significance for the development of coal mine intelligence. Considering that the underground workways in coal mines are not a completely closed working environments, obstacles, such as people, rail locomotives, and falling stones, sometimes appear on the transport line. The real-time and accurate detection of these obstacles is one of the key technologies of driverless rail locomotives.

Compared with other on-board sensors, cameras can obtain the most abundant information about the surrounding environment. Moreover, with their low cost and highly

developed hardware, cameras have unique advantages in the field of autonomous driving. However, a camera is susceptible to environmental factors. When the illumination changes or the background becomes complex, the accuracy and robustness of the target detection algorithm decreases. At present, the research on the intelligent-vehicle visual-object detection algorithm has made considerable progress; however, there are still a series of problems and difficulties in its application in underground coal mines. In general, the factors affecting the overall performance of the algorithm mainly include complex-scene and network-model factors in the coal mine: (1) complex-scene factors in a coal mine mainly include low illumination levels, motion blur, occlusion, small targets, background interference, etc.; (2) network-model factors include an existing detection model with a complex architecture and many parameters, which cannot meet the requirements of the real-time detection of edge devices. On the premise of ensuring high detection accuracy, it becomes particularly important to improve the detection speed of the model.

The visual equipment of the automatic rail locomotive is installed at the front of the locomotive, and the uneven track, operation of the rail locomotive, and movement of obstacles blur the pictures collected by the visual equipment, which seriously hinder the intelligent detection based on the image processing technology. It is necessary to deblur the collected pictures. With the rapid development of deep learning and big data technology, image deblurring using the Convolution Neural Network (CNN) has been widely studied. Goodfellow et al. [4] proposed a Generative Adversarial Network (GAN) in 2014, which relies on the idea of two network games to conduct targeted learning, and, on this basis, can be expanded to generate realistic and clear images to achieve the blind deblurring effect. Radford et al. [5] proposed the deep convolution generative adversarial network in 2016, which was improved on the basis of GAN to address the defect of GAN learning instability. The generative model introduces the convolution network structure, which effectively improves the learning level of the network. Ledig et al. [6] proposed the SRGAN neural network in 2017. Based on the perceptual loss optimization algorithm, super-resolution image restoration was realized. Mirza et al. [7] proposed the Conditional Generative Adversarial Network (CGAN). Based on GAN, additional conditional information was added to the corresponding generator and discriminator to realize the conditional generation model. In 2018, Kupyn et al. [8] first applied the GAN network structure in the field of deblurring, restored blurred images through generators, differentiated clear images from blurred images by discriminators, and trained them against each other to remove the blur caused by the camera's motion. In 2019, the feature pyramid network was introduced on this basis, and the performance was improved further [9].

The research on target detection can be divided into traditional methods and methods based on deep learning. Traditional methods use appearance and color features for target detection purposes [10,11]; however, feature extraction depends on manual work and has numerous limitations, which cannot meet the real-time and accuracy requirements of obstacle detection methods when driving automatic rail locomotives. In 2014, R-CNN [12] used a convolution network for target detection for the first time. To date, deep learning-based methods have been widely used for target detection. The target detection algorithm based on the deep learning method is mainly divided into two categories: the one- and two-stage detection algorithms. One-stage detection algorithms mainly include the YOLO series [13–15], EfficientDet [16], and SSD [17]. Two-stage detection algorithms mainly include Fast R-CNN [18] and Faster R-CNN [19]. He et al. [20] proposed a track-obstacle detection algorithm based on improved R-CNN. By introducing a new up-sampling parallel structure and context extraction module (CEM) into the architecture of R-CNN, the accuracy of obstacle detection was improved. He et al. [21] proposed a rail-transit-obstacle detection algorithm based on improved Mask R-CNN, which uses an SSwin-Le Transformer as the feature extraction network and ME-PAPN as a feature fusion network. A variety of multi-scale enhancement methods were integrated to improve the accuracy of small-target detection. He et al. [22] proposed an obstacle detection algorithm for dangerous track areas, which uses the Mask R-CNN model based on ResNet101 as the backbone feature extraction network. The experimental results

show that the network has a high detection accuracy for small targets. He et al. [23] proposed a track-obstacle detection algorithm based on improved YOLOv4. D-CSPDarknet was designed as a feature extraction network, and the feature fusion network combines path aggregation and feature pyramid networks to establish a spatial pyramid aggregation network in each fusion layer to improve the detection accuracy of medium- and long-distance obstacles. He et al. [24] proposed a flexible and efficient multi-scale single-stage target detector, FE-YOLO, for track-image-obstacle detection, and designed a repeatable bi-directional cross-scale path aggregation module as the core of the feature fusion network to improve the accuracy of track-obstacle detection. Wang et al. [25] proposed a method of track-obstacle detection based on improved YOLOv3. A four-scale detection structure was formed by adding a scale to the three scales of the original YOLOv3, and the detection accuracy of small-target stones was increased. The reference [20–25] mainly focuses on improving the detection accuracy of the algorithm but ignores the detection speed of the algorithm. For the obstacle detection of an electric locomotive driving scene, it is more necessary to achieve rapid-detection networks. Chen et al. [26] proposed a lightweight network-based foreign-body-intrusion detection method for the railway regions of interest. Sparse and channel-pruning methods were used to compress the YOLOv3 model, and a lightweight railway foreign-body-intrusion detection model was constructed. This model is effective for target detection in an ideal environment, but is poor for complex environments, such as those with low illumination. Han et al. [27] proposed a vehicle-target detection algorithm based on improved the YOLOv4-tiny algorithm, which improves the detection speed of the algorithm while losing detection accuracy, and does not achieve a good balance between detection accuracy and speed. Dong et al. [28] proposed a target detection algorithm based on improved YOLOv5, which replaces all convolution modules in YOLOV5 with the Ghost module with low computational complexity and simplifies the backbone network to improve network performance. However, because the improved algorithm is very lightweight, it is not effective for environment detection purposes, such as mutual occlusion and background interference. Hao et al. [29] proposed a type of foreign-body detection device in a coal mine conveyor belt based on CBAM-YOLOv5. In order to solve the problem that the foreign-body target was difficult to accurately detect due to the low-illumination conditions, the convolution block attention model was introduced into the YOLOv5 detection network to improve the salience of the foreign-body target in the image, enhance the feature-expression ability of the foreign-body target in the detection network, and then improve the detection accuracy of the foreign-body target. However, the speed of the algorithm needs to be improved further.

Based on this result, a real-time obstacle detection method when driving automatic rail locomotives based on DeblurGANv2 and improved YOLOv4 is proposed. DeblurGANv2 is used to solve the problem of motion blur. The YOLOv4 algorithm is improved to ameliorate the detection performance in an environment with low illumination, occlusions, small targets, and background interference, and the detection speed of the algorithm is improved to meet the requirements of the real-time detection of edge equipment. This paper (1) calculates the blur of the image by the Laplace operator to accurately judge whether the image needs to be deblurred or not, and the blurred image is deblurred by DeblurGANv2. (2) The improved design is based on YOLOv4, and the lightweight feature extraction network MobileNetv2 [30] is used to replace the original CSPDarknet53 network to improve the detection speed of the algorithm. There is a high amount of background interference information in the target detection of the coal mine scene. In order to improve the attention paid to the target, the SANet [31] attention module is embedded in the Neck network to improve the detection accuracy of the algorithm under low-illumination, target-occlusion, small-target, and other conditions. To further improve the detection accuracy of the algorithm, the K-means++ algorithm is adopted to cluster prior frames, and the focal [32] loss function is introduced to increase the weight loss of small-target samples. (3) The obstacle datasets of the rail locomotive operation area are constructed to provide a test environment for all kinds of obstacle detection algorithms, for which dataset 1 is

composed of clear pictures, dataset 2 is composed of blurred pictures, and dataset 3 is composed of pictures deblurred by DeblurGANv2. The experimental results show that the deblurring of the motion-blurred picture can effectively improve the detection accuracy of obstacles and reduce missed detections. Compared with other commonly used models, the improved YOLOv4 algorithm presents a better balance and can ensure the detection accuracy and speed of the model simultaneously.

2. Methods

In order to improve the accuracy and speed of obstacle detections during the driving of rail locomotives, the research conducted considering two aspects: motion-blur image processing and the improved detection algorithm. In order to facilitate the readers' understanding, we refer to reference [33] and describe the work by presenting a pseudo-code (Algorithm 1).

Algorithm 1: pseudo code of the proposed algorithm [33]

Input: A picture with obstacle targets.
Execute the algorithm in following order to get the desired result.
begin
    Blur judgment
        do
            Use formula (2) to judge whether the image is blurred or not.
            Using DeblurGANv2 to deblur the blurred image.
        end
    Improved YOLOv4 algorithm
        do
            Replace the backbone network and reduce the number of channels in the Neck and Head parts.
            Introduce the SANet attention mechanism.
            the K-means++ algorithm is adopted to cluster prior frames.
            the Focal loss function was introduced to increase the loss weight of small target samples.
        end
    Training network
        do
            Build dataset.
            Training parameter configuration.
            Weights: use pre-trained VOC dataset.
            Train the network and generate model weights.
        end
end
Output: obstacles were detected through the Bounding box around.

2.1. Image Deblurring via DeblurGANv2

DeblurGANv2 is a deblurring method based on deep learning. DeblurGANv2 uses a general feature pyramid network as the backbone feature extraction network, allowing users to select Inception-ResNet-v2 [34], MobileNetv2 [30], and MobileNetDSC in the backbone, which enables users to optimize the accuracy, speed, or both. The running times of MobileNetDSC and MobileNet are less than 0.06 s, which makes real-time deblurring possible. The DeblurGANv2 model with a MobileNetv2 backbone feature extraction network was used to deblur images. MobileNetV2 was used as our backbone feature extraction network because it has a rapid runtime and maintains almost the same accuracy as the Inception-ResNet-v2 backbone.

Not all the pictures collected by the camera are blurred. If all the collected pictures are deblurred, it not only wastes resources, but also causes serious information loss and reduces the accuracy of the obstacle detection process. Therefore, this study used the blur-judgment mechanism proposed by Zhou et al. [35] to perform a blur judgment of the

collected pictures. The motion-blur image-processing flow is shown in Figure 1, in which the image blur value can be calculated by the Laplace operator.

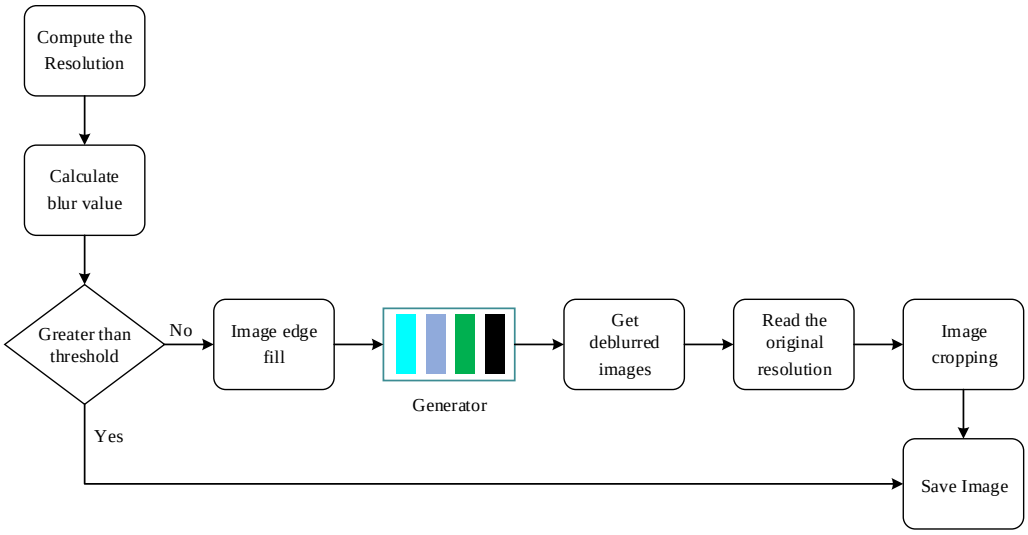


Figure 1. Motion-blurred-image processing flow.

Through the Laplace operator, we could determine whether there was more or less high-frequency information in the picture. If the picture has more high-frequency information, the picture is considered clear; on the contrary, it is considered that the picture is relatively blurred. The Laplace operator is defined in Equation (1):

$$\Delta f = \nabla^2 f = \nabla \cdot \nabla f \tag{1}$$

where f is a second-order differentiable real function. The template e operator is presented in Figure 2.

0	1	0
1	−4	1
0	1	0

Figure 2. Template of the Laplace operator.

$D(f)$ is defined in Equation (2):

$$D(f) = \sum_{y} \sum_{x} |G(x,y)| G(x,y) > T \tag{2}$$

where $G(x, y)$ is the convolution of the Laplace operator at the pixel point (x, y). T is the given threshold.

2.2. MobileNetv2 as the YOLOv4 Backbone Network

The original YOLOv4 network uses CSPDarknet-53 as the feature extraction network, which is a kind of full-convolution network with a complex structure and has a strong feature extraction ability; however, with the increase in the number of residual units and

network channels, it leads to a sharp increase in the number of network parameters, which affects the speed of model detection. For the multi-target detection of the electric locomotive driving scene, it is more necessary for the network to perform rapid detection; therefore, there is no need for a feature extraction network that is too complex. Therefore, taking advantage of the deep separable convolution and inverse residual structure of lightweight network Mobilenetv2, it was used as the feature extraction network of YOLOv4 to reduce the model's capacity and number of parameters, improve the detection speed of the model, and solve the problem of slow network training caused by the limited hardware.

Depthwise separable convolution is a form of convolution that divides a standard convolution into two parts, namely, depth (*D*Conv) and point (*P*Conv) convolutions. Figure 3 shows the general process of depthwise separable convolution. Firstly, a filter is input into each channel to generate a feature map, namely, *D*Conv. *D*Conv can be expressed as:

$$DConv(W_d, x)_{(i,j)} = \sum_{m=0}^{M} \sum_{n=0}^{N} W_d \cdot x_{(i+m,j+n)} \tag{3}$$

where W_d is the weight matrix of *D*Conv; x is the input feature mapping of the convolution layer; (i, j) is the coordinate point of output feature mapping; M and N represent the height and width of the input layer, respectively; and m and n represent the two-dimensional space of the convolution kernel.

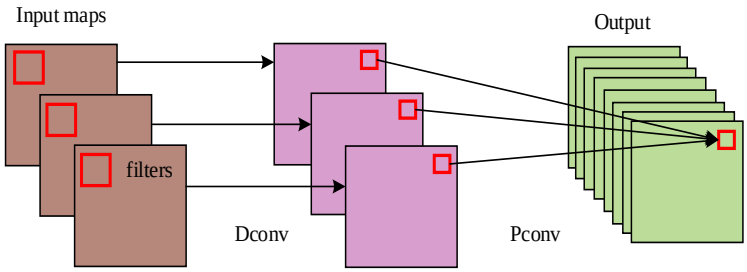


Figure 3. Depths-wise separable convolution.

Then, *P*Conv is used to combine the output of *D*Conv, which can effectively extract the spatial features. *P*Conv can be expressed as follows:

$$PConv(W_p, x)_{(i,j)} = \sum_{k=0}^{K} W_p \cdot x_{(i,j)} \tag{4}$$

where W_p is the weight matrix of *P*Conv; K represents the input depth of this layer; k represents the convolution kernel.

In summary, the overall process of depthwise separable convolution can be expressed as:

$$DSConv(W_p, W_d, x)_{(i,j)} = PConv(W_p, x)_{(i,j)} \times (W_p, DConv(W_d, x)_{(i,j)}) \tag{5}$$

The computational complexities of ordinary and depthwise separable convolutions are compared below. Let $D_k \times D_k \times U$ be the size of the input feature map and $D_F \times D_F \times U$ be the size of the convolution kernel, whose quantity is X. The computational level of the ordinary convolution is:

$$D_k \times D_k \times D_F \times D_F \times U \times X \tag{6}$$

The computational level of the depth-separable convolution is:

$$D_k \times D_k \times D_F \times D_F \times U + U \times X \times D_k \times D_k \tag{7}$$

The computational cost ratio of depth separable to ordinary convolutions is:

$$\frac{D_k \times D_k \times D_F \times D_F \times U + U \times X \times D_k \times D_k}{D_k \times D_k \times D_F \times D_F \times U \times X} = \frac{1}{X} + \frac{1}{D_F^2} \quad (8)$$

It can be observed from the calculation results above that the calculation for depthwise separable convolution is much lower than that of the ordinary convolution. Mobilenetv2 with the depthwise separable convolution was used as the backbone feature extraction network, which can effectively reduce the number of model parameters and computations.

Although depthwise separable convolution can reduce the number of model parameters well, it extends the number of network layers, resulting in the disappearance of gradients, and the residual structure can solve this situation very well. Figure 4 shows the residual structure and Figure 5 presents the inverse residual structure. It is not difficult to observe from Figures 4 and 5 that the residual structure is a process of "dimensionality fall–convolution–dimensionality rise", while the inverse residual structure is a process of "dimensionality rise–depth convolution–dimensionality fall". This is because the depthwise separable convolution finds it difficult to extract enough feature information as the standard convolution method; if it is extracted after a decrease in the dimensionality, it reduces the features extracted by the network and affects the network's performance. Therefore, the increase in dimensionality and the depthwise convolution not only ensure that sufficient features are extracted from the backbone network, but also reduce the number of parameters and the computations.

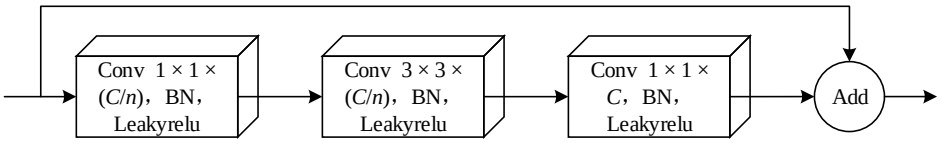
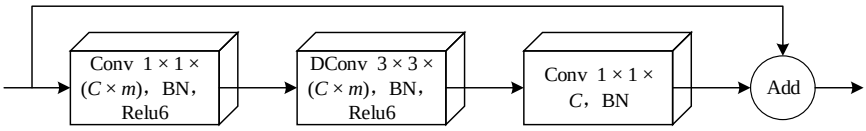
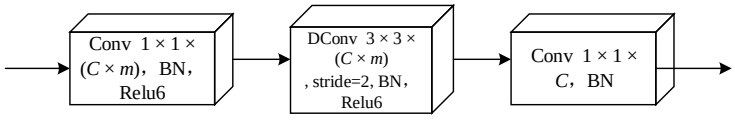


Figure 4. Residual structure.



(**a**) Structure 1.



(**b**) Structure 2.

Figure 5. Inverse residual structure.

The MobileNetv2 structure used in this study is presented in Table 1, with a total of 72 network layers. Layers 39 (52 × 52 × 84), 58 (26 × 26 × 576), and 72 (13 × 13 × 320) were selected as the three output layers for the backbone feature extraction network.

Table 1. Adopted MobileNetv2 structure.

Input Size	Type	Output Channel Number	Stride
416 × 416 × 3	Conv 3 × 3	32	2
208 × 208 × 32	Block1	16	1
208 × 208 × 16	Block2	24	2
104 × 104 × 24	Block1	24	1
104 × 104 × 24	Block2	32	2
52 × 52 × 32	Block1	32	1
52 × 52 × 32	Block1	32	1
52 × 52 × 32	Block1	64	1
52 × 52 × 64	Block1	64	1
52 × 52 × 64	Block1	64	1
52 × 52 × 64	Block1	64	1
52 × 52 × 64	Block2	96	2
26 × 26 × 96	Block1	96	1
26 × 26 × 96	Block1	96	1
26 × 26 × 96	Block1	96	1
26 × 26 × 96	Block1	96	1
26 × 26 × 96	Block2	160	2
13 × 13 × 160	Block1	160	1
13 × 13 × 160	Block1	160	1
13 × 13 × 160	Block1	320	1

2.3. SA Attention Mechanism

The SA attention mechanism can allocate the limited computing resources to the section with the highest amount of information in the image, which can gather the attention of the network model from the recognized object and reduce the influence of the image background better. Therefore, the SA module was introduced to enhance the attention of the network, highlight the key features, and improve the detection accuracy of the algorithm under the conditions of low illumination, target occlusions, and small targets.

In neural network learning, there are mainly two kinds of attention mechanisms: channel and spatial. The Convolutional Block Attention Module (CBAM) [36] combines channel and spatial attention mechanisms to achieve a better performance; however it produces a high amount of computation and is difficult to converge. The SA attention module uses a shuffle unit to simultaneously combine channel and spatial attention mechanisms. As shown in Figure 6, feature graph X with an input size of $H \times W \times C$ is divided into g groups along the channel size, where g utilizes 64 to obtain the matrix $[X_1, \ldots, X_g]$, and then divides each feature graph X_k (X_k size is $H \times W \times C/g$) into X_{k1}, X_{k2} (X_{k1} and X_{k2} sizes are both $H \times W \times C/2g$) along the channel; X_{k1} and X_{k2} are processed by channel and spatial attention mechanism, respectively. Channel attention begins with Global Average Pooling (*GAP*), utilizing the following expression:

$$s = GAP(X_{k1}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{k1}(i, j) \tag{9}$$

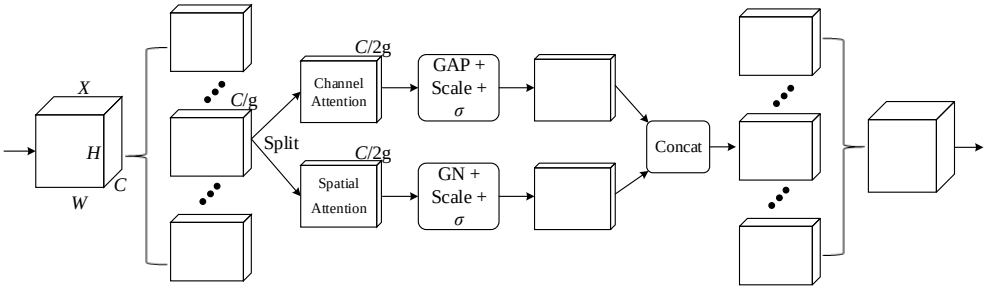


Figure 6. SA module-structure diagram.

Then, continue with the scale operation, use parameter scaling W_1 and move channel vector b_1, and finally continue with the activation function σ processing:

$$X'_{k1} = \sigma(W_1 s + b_1) \cdot X_{k1} \tag{10}$$

The spatial attention operation first performs the Group Norm (GN) and then performs the same operation as the channel attention mechanism. The specific expression is as follows:

$$X'_{k2} = \sigma(W_2 \cdot GN(X_{k2}) + b_2) \cdot X_{k2} \tag{11}$$

Finally, channel and spatial attention mechanism are fused, so that there are two-directional attention mechanisms on each divided feature graph, and the divided feature graph is shuffled to realize the flow of cross-group information along the channel dimension.

2.4. Optimal Design of Prior Frame

Selecting the appropriate prior frame plays a key role in improving the effect of network training. Due to the different sizes and shapes of obstacles, the initial prior frame of the original YOLOv4 cannot meet the detection needs, and the K-means algorithm randomly selects the initial value; therefore, the clustering effect and stability are not good. In order to reduce the error caused by the prior-frame size problem, the K-means++ algorithm [37] was selected to cluster the labels of the obstacle dataset and generate 9 groups of anchors with different aspect ratios. The clustering results are shown in Table 2. The clustering effect of this algorithm is more stable, and the generated prior frame is closer to the actual size distribution of the dataset.

Table 2. Nine sets of prior boxes obtained by K-means++.

Feature Map	Receptive Field	Anchor
13 × 13	Big	(74,57), (114,204), (126,109)
26 × 26	Medium	(45,31), (50,85), (62,148)
52 × 52	Small	(20,29), (23,15), (30,54)

2.5. Optimization of Loss Function

The YOLOv4 algorithm needs to deal with a high number of prior frames during training. When most of the prior frames do not contain targets to be detected, this causes an imbalance between the number of positive and negative training samples. Through an analysis of the constructed obstacle dataset, it was observed that most of the small-target obstacles, such as stones, occupied a small proportion of the image, which can easily lead to an imbalance in the number of positive and negative samples in the training process of the algorithm. Based on this result, focal loss function was introduced to replace the cross-entropy loss function, which reduces the weight of the simple sample background, causes the model to focus on target-object detection, and avoids a tendency to obtain a high

number of background samples when the algorithm performs a prediction. The focal loss function is shown in Equations (12)–(14):

$$FL(p_i) = -\alpha_i(1-p_i)^{\zeta}\log(p_i) \tag{12}$$

$$\alpha_i = \begin{cases} \alpha, & \text{if } y = 1 \\ 1-\alpha, & \text{otherwise} \end{cases} \tag{13}$$

$$p_i = \begin{cases} p, & \text{if } y = 1 \\ 1-p, & \text{otherwise} \end{cases} \tag{14}$$

2.6. Improved Network Structure

The detection of obstacles during the driving of rail locomotives not only requires a high detection accuracy, but also a faster detection speed. Compared with the original YOLOv4 algorithm, the improved YOLOv4 algorithm achieves a higher detection speed while ensuring higher detection accuracy, so as to satisfy real-time detection. MobileNetv2, the SA attention module, and the YOLOv4 detection network were combined to construct an improved YOLOv4 target detection network. The network structure is shown in Figure 7.

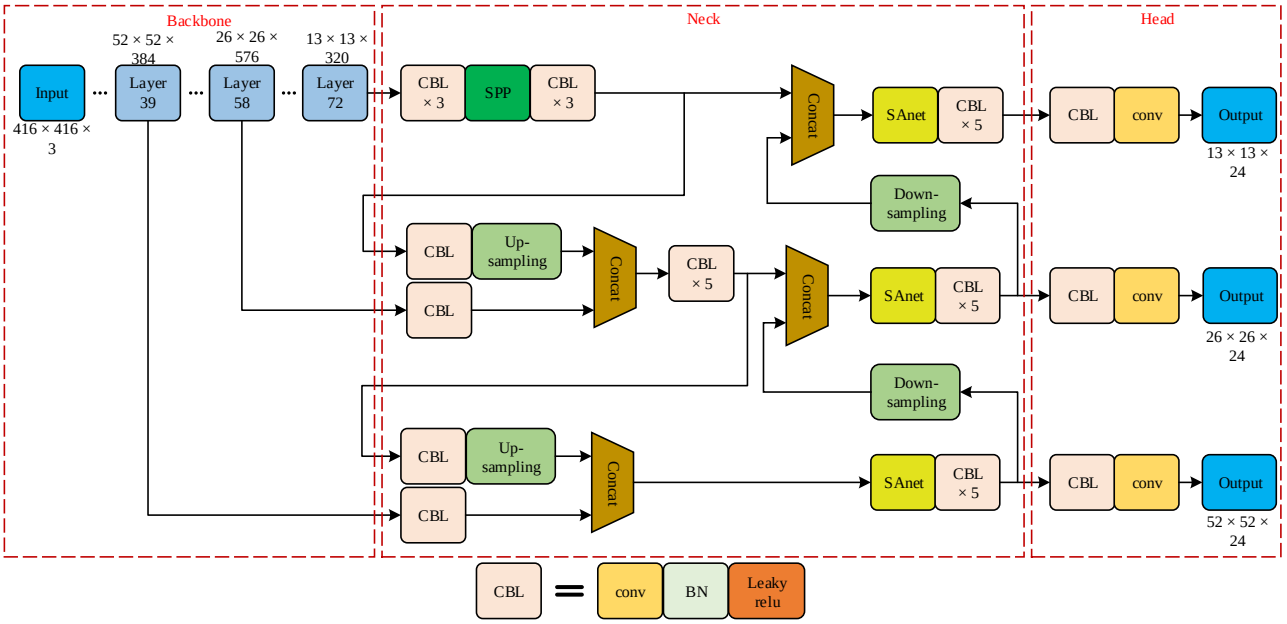


Figure 7. Improved network structure.

3. Results and Discussion

3.1. Create a Dataset

The experimental dataset was obtained from a clear video recorded in Yuandian No. 1 Mine in Huaibei. A total of 1600 pairs of blurred–clear pictures were generated by averaging the adjacent 7~13 frames of each video. Clear images were compiled into dataset 1, and some clear images are shown in Figure 8a–c. Blurred images were compiled into dataset 2, and some blurred images are shown in Figure 8d–f. DeblurGANv2 was used to deblur the blur images in dataset 2, and the processed images were compiled into dataset 3. Some images are shown in Figure 8g–i.

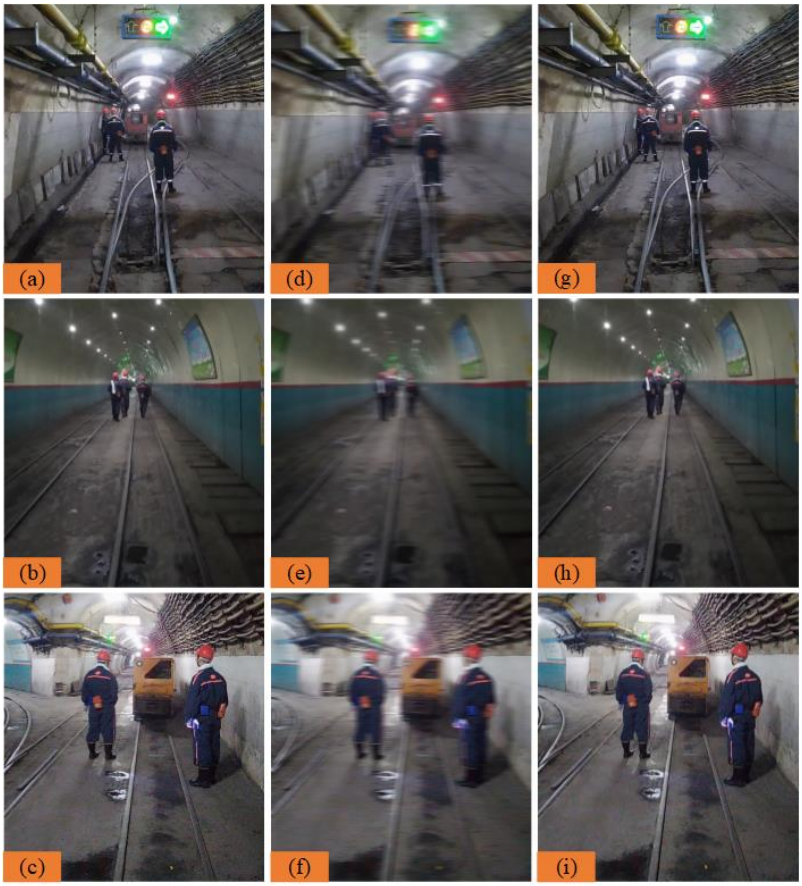


Figure 8. Partial dataset pictures. (**a**–**c**) Clear; (**d**–**f**) blurred; and (**g**–**i**) deblurred pictures.

3.2. Test Parameter Configuration

Training and testing were run on a computer with an Intel (R) Core (TM) CPU model, i7 CPU @ 2.90 GHz, and a NVIDIA GeForce RTX2060 GPU model. The parallel computer framework of CUDA10.0 version and the deep learning acceleration library of Cudnn7.3 version were installed. On the TensorFlow deep learning framework, Python3.6 was used as the programming language to implement this research.

3.3. Model Training and Evaluation Index

Model training details: 32 images were loaded in each iteration and Forward Propagation (FP) process was completed in 16 batches. After 32 images completed the forward propagation process, the parameters were updated by Back Propagation (BP). The maximum number of iterations was set as 10,000, and the initial learning rate, weight attenuation coefficient, and momentum factor were set as 0.001, 0.0005, and 0.9, respectively. The multi-scale training strategy was used to randomly select the input size of the image from {320,352,384,416,448, 480,512,544,576,608} every 10 iterations [38]. Flip, zoom, translation, rotation, noise, and other methods were randomly combined to generate more target samples that could be trained by the network, so as to enhance the robustness and generalization of the network [39,40].

Evaluation index: the performance of the model was evaluated by Precision (P), Recall (R), Average Precision (AP), mean Average Precision (mAP), and detection speed (FPS). The calculation formula is shown in (15)–(18):

$$\mathrm{P} = \frac{TP}{TP + FP} \tag{15}$$

$$\mathrm{R} = \frac{TP}{TP + FN} \tag{16}$$

$$\mathrm{AP} = \int_0^1 P(R)dR \tag{17}$$

$$\mathrm{mAP} = \frac{1}{c} c_i \mathrm{AP} \tag{18}$$

where *TP* indicates the number of positive samples identified as positive samples, *FP* indicates the number of positive samples mistakenly identified as positive samples, *FN* indicates the number of positive samples misjudged as negative samples, and c represents the number of categories in the sample.

3.4. Analysis of Experimental Results

3.4.1. Experimental Results

In order to verify the contribution to the accuracy of the obstacle detection method by using DeblurGANv2 to deblur blurred images, the improved YOLOv4 algorithm was used to conduct experiments on three datasets. The experimental results are presented in Table 3, and some of the obstacle detection results are shown in Figure 9. The mAP based on dataset 1 was 98.02%, mAP based on dataset 2 was 88.28%, and mAP based on dataset 3 was 97.19%. Using DeblurGANv2 to deblur blurred images can effectively improve obstacle detection accuracy. It is easy to observe from Figure 6 that using the improved YOLOv4 algorithm to detect blurred images makes it prone to the problems of missing detection and inaccurate location, and deblurring can solve this problem.

Table 3. Experimental results based on three datasets.

Dataset	Detection Model	Average Precision (AP)/%			mAP/%
		E-L	People	Stone	
1	Improved YOLOv4 proposed	99.16	98.65	96.24	98.02
2	Improved YOLOv4 proposed	96.27	88.12	80.46	88.28
3	Improved YOLOv4 proposed	98.83	97.56	95.18	97.19

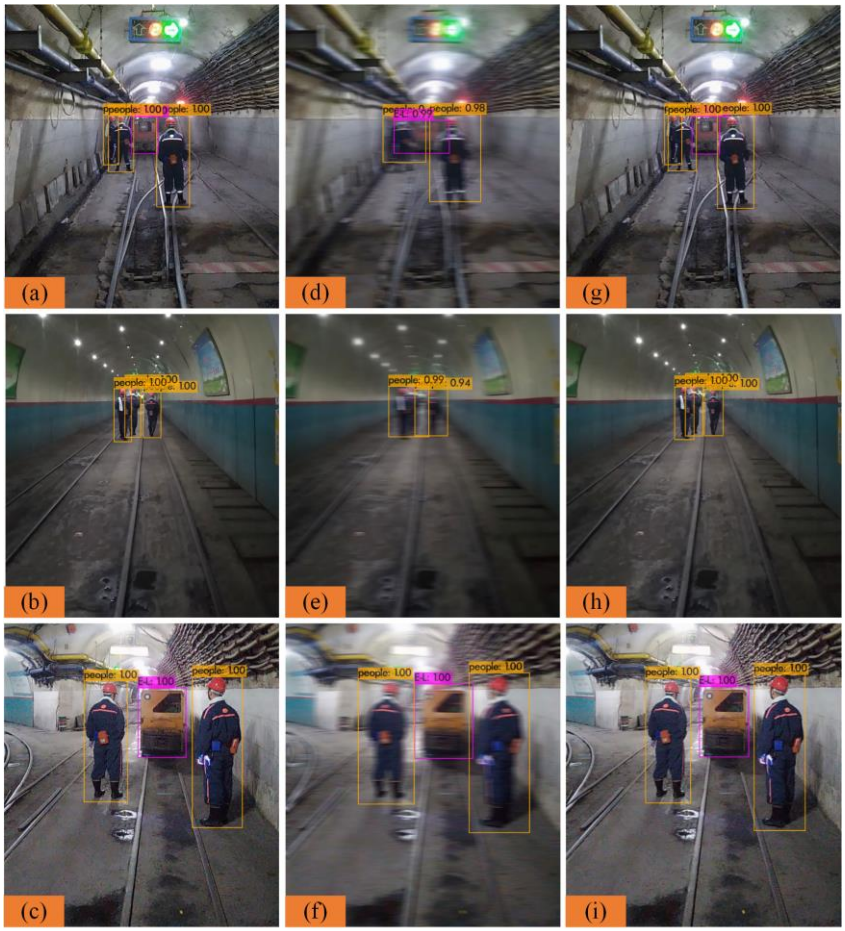


Figure 9. Detection results for some images in the three datasets. (**a**–**c**) Clear pictures in dataset 1; (**d**–**f**) blurred pictures in dataset 2; and (**g**–**i**) deblurred images in dataset 3.

Figure 10 shows the detection results of the improved YOLOv4 in occlusion and small-target scenes, and Figure 11 shows the detection results of the improved YOLOv4 algorithm in low-light scenes. As can be observed from Figures 10 and 11, the algorithm has better detection results in three complex scenarios when adding the SA attention mechanism.

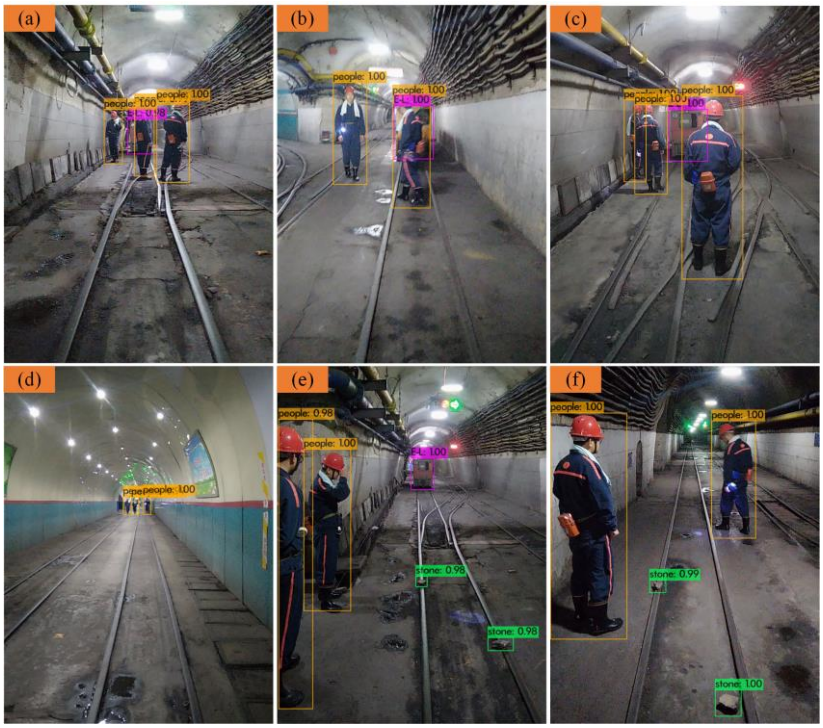


Figure 10. Detection results for two scenes with occlusions and small targets. (**a**–**c**) Occlusion targets; (**d**–**f**) small targets.



Figure 11. Improved YOLOv4 detection results for low-light scenes. (**a**–**f**) Example of test result.

3.4.2. Ablation Experiment

In order to perform a more comprehensive analysis of the contribution of each improved strategy to the performance of obstacle detection, ablation experiments were conducted on dataset 1, and the experimental results are presented in Table 4. Network A is the original YOLOv4+K-means clustering network; network B is the model after Mobilenetv2 was selected as the YOLOv4 backbone network; network C adds the SA module on the basis of network B; network D uses K-means++ clustering on the basis of network C; and network E introduces the focal loss function on the basis of network D.

Table 4. Ablation experiment results.

Network	YOLOv4 and Its Improvements	Average Precision (AP)/%			mAP/%	FPS
		E-L	People	Stone		
A	YOLOv4 + K-means clustering	98.51	97.67	95.83	97.34	41
B	A + Mobilenetv2	98.33	99.20	88.26	95.26	72
C	B + SA module	99.32	98.41	92.15	96.63	68
D	C + K-means++ clustering	99.22	98.86	94.38	97.49	68
E	D + Focal loss function	99.16	98.65	96.24	98.02	68

As can be observed from Table 4, using Mobilenetv2 as the backbone of YOLOv4 can effectively improve the detection speed of the network, which increases by 75.61% to 72 FPS; however, the detection accuracy also decreases. The addition of the SA attention mechanism strengthens the feature extraction ability of the network, which can effectively suppress background interferences, improve the detection accuracy of obstacles in a complex background environment, and improve mAP by 1.37%. The K-means++ algorithm was used to cluster the prior frames, detection accuracy was improved further without affecting the detection speed, and mAP increased by 0.86%. The focal loss function was introduced to replace the cross-entropy loss function to reduce the amount of simple-sample background, and the model focused on the detection of target objects. Without affecting the detection speed, detection accuracy was further improved and mAP increased by 0.53%.

3.4.3. Comparative Experiments on Different Attention Mechanisms

In order to present the advantages of the SA attention mechanism over other attention mechanisms, this experiment added the SA, SE, CBAM, and CA modules at the same location of the Neck network of the basic Mobilenetv2-YOLOv4 network. The experimental results are presented in Figure 12. After adding the attention module, the number of parameters in the whole network slightly increased, which caused a decrease in the reasoning speed of the model; however, it still reached a value higher than 65FPS. Several attention mechanisms can improve the accuracy of obstacle detection methods. Among them, after adding the SA attention module, model accuracy was the improved the most, and mAP was improved by 1.37%.

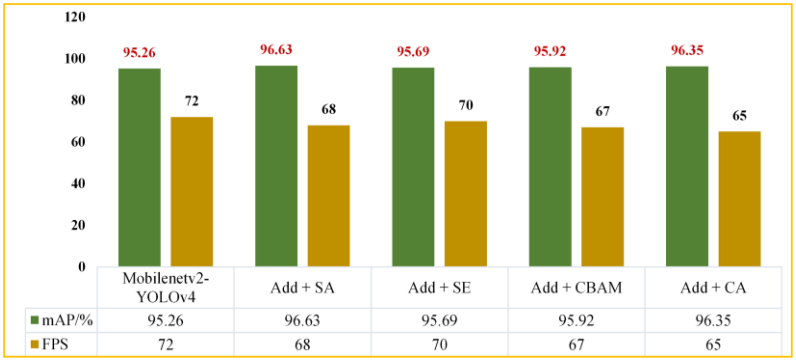


Figure 12. Comparative experimental results for different attention mechanisms.

3.4.4. Comparative Experiment for Different Algorithms

In order to compare the improved YOLOv4 network with the mainstream object detection network used at present, faster R-CNN with relatively high accuracy in the two-stage target detection network and YOLOv3, YOLOV3-tiny, YOLOv4, and YOLOV4-tiny in the single-stage target detection network were selected to conduct comparative tests on dataset 1. All algorithms were trained with pre-trained weights. The experimental results are presented in Table 5. Among the several algorithms, the improved YOLOv4 algorithm achieved the best detection accuracy. Compared with the original YOLOv4 network, the improved YOLOv4 detection speed increased by 65.85% to 68FPS. Although YOLOv3-tiny and YOLOv4-tiny were better than the improved YOLOv4 in terms of detection speed, the detection accuracy was lower. The improved YOLOv4 algorithm was better balanced and could guarantee the high detection accuracy and speed of the model simultaneously.

Table 5. Comparative experimental results for different target detection algorithms.

Model	Average Precision (AP)/%			mAP/%	FPS
	E-L	People	Stone		
Faster R-CNN [19]	99.48	92.53	91.24	94.42	10
YOLOv3 [13]	99.87	98.84	90.78	96.50	39
YOLOv3-tiny [41]	98.43	98.46	81.83	92.91	104
YOLOv4 [15]	98.51	97.67	95.83	97.34	41
YOLOv4-tiny [42]	98.49	97.61	85.48	93.86	109
Improved YOLOv4	99.16	98.65	96.24	98.02	68

4. Conclusions

In order to improve the detection accuracy of an algorithm in the complex environment of a coal mine, including low illumination, motion blur, occlusions, small targets, and background interference; reduce the number of model parameters; improve the detection speed of the algorithm; and cause it to meet the needs of the real-time detection of edge equipment, our research was conducted from two angles: motion-blur image processing and the improved detection algorithm. The blur value of the image was calculated by the Laplace operator to accurately judge whether the image needed to be deblurred or not, and the blurred image was deblurred using DeblurGANv2. With YOLOv4 as the infrastructure, we made the following improvements: (1) replaced the original CSPDarknet53 network with the lightweight feature extraction network MobileNetv2; (2) embedded the SA attention module in the Neck network; (3) used the K-means++ algorithm to cluster the prior frames, and introduced the focal loss function to replace the cross-entropy loss function. The experimental results show that the deblurring of the motion-blurred image can effectively improve the detection accuracy of obstacles and reduce the number of missed detections. Compared with other commonly used models, the improved YOLOv4 algorithm has better results and can ensure the high-accuracy detection and speed of the model simultaneously.

In this paper, experiments were conducted on our datasets, and the algorithm presented some limitations. The subsequent step is to increase the diversity and complexity of the data and further improve the network structure to improve the target detection accuracy under complex working conditions.

The development of multi-sensor information fusion technology will be the focus of future research. This paper mainly studied the perception algorithm based on visual information, and the collection of the visual information was affected by shooting angles for the videos, a low-illumination environment, obstacle occlusions, rapid movements, and so on. Multi-sensor information fusion can make up for the shortcomings of a single sensor and realize the complementary advantages of multi-source information. At present, the domestic research concerning this technology is still in the initial stage, and the data fusion

method and implementation of the fusion system are key issues for the research conducted in the future.

Author Contributions: S.W. conceived the idea. W.W. and Y.Z. performed the data analyses and wrote the manuscript. J.T. and T.Y. edited the manuscript. All authors discussed the results and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Anhui Province University Outstanding Youth Research Project under Grant No. 2022AH020056, the National Natural Science Foundation of China under Grant No. 52274152, the Collaborative Innovation Project of Universities in Anhui Province under Grant No. GXXT-2020-60, and the Graduate Innovation Fund of Anhui University of Science and Technology under Grant No. 2021CX1008.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and/or analyzed during the present study are available from the corresponding author on reasonable request.

Acknowledgments: Thank you for the experimental environment support provided by Huaibei Mining Group for this project.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Han, J.; Xing, W.; Yang, L.; Zhen, W.; Yun'an, C.; Lei, C. Driverless technology of underground locomotive in coal mine. *J. China Coal Soc.* **2020**, *45*, 2104–2115.
2. Yangyang, C.; Zhenlong, H.; Zhiwei, L. Development trend and key technology of coal mine transportation robot in China. *Coal Sci. Technol.* **2020**, *48*, 233–242.
3. Shirong, G.; Eryi, H.; Wenliang, P. Classification system and key technology of coal mine robot. *J. China Coal Soc.* **2020**, *45*, 455–463.
4. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Proceedings of International Conference on Neural Information Processing Systems*; MIT Press: Kuching, Malaysia, 2014; pp. 2672–2680.
5. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
6. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
7. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv preprint* **2014**, arXiv:1411.1784. [CrossRef]
8. Kupyn, O.; Budzan, V.; Mykhailych, M.; Mishkin, D.; Matas, J. DeblurGAN: Blind motion deblurring using conditional adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 8183–8192. [CrossRef]
9. Kupyn, O.; Martyniuk, T.; Wu, J.; Wang, Z. Deblurgan-v2: Eblurring (orders-of-magnitude) faster and better. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8878–8887.
10. Souani, C.; Faiedh, H.; Besbes, K. Efficient algorithm for automatic road sign recognition and its hardware implementation. *J. Real Time Image Process.* **2014**, *9*, 79–93. [CrossRef]
11. Maldonado, B.S.; Lafuente, A.S.; Gil, J.P.; Gomez-Moreno, H.; Lopez-Ferreras, F. Road-sign detection and recognition based on support vector machines. *IEEE Trans. Intell. Transp. Syst.* **2007**, *8*, 264–278. [CrossRef]
12. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 23–28 June 2014; pp. 580–587.
13. Tian, Y.; Yang, G.; Wang, Z.; Wang, H.; Li, E.; Liang, Z. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput. Electron. Agric.* **2019**, *157*, 417–426. [CrossRef]
14. Hendry, H.; Chen, R. Automatic License Plate Recognition via sliding-window darknet-YOLO deep learning. *Image Vis. Comput.* **2019**, *87*, 47–56. [CrossRef]
15. Bochkovskiy, A.; Wang, C.Y.; Lia, H.-Y.M.O. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2014**, arXiv:2004.10934.
16. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 16–18 June 2020; pp. 10781–10790.
17. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single shot multibox detector. In *Proceedings of European Conference on Computer Vision*; Springer: Berlin, Germany, 2016; pp. 21–37.

18. Girshick, R. Fast R-CNN. In *IEEE International Conference on Computer Vision*; IEEE Press: Washington, DC, USA, 2015; pp. 1440–1448.
19. Ren, S.Q.; He, K.M.; Girshick, R.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
20. He, D.; Ren, R.; Li, K.; Zou, Z.; Ma, R.; Qin, Y.; Yang, W. Urban rail transit obstacle detection based on Improved R-CNN. *Measurement* **2022**, *196*, 111277. [CrossRef]
21. He, D.; Qiu, Y.; Miao, J.; Zou, Z.; Li, K.; Ren, C.; Shen, G. Improved Mask R-CNN for obstacle detection of rail transit. *Measurement* **2022**, *190*, 110728. [CrossRef]
22. He, D.; Li, K.; Chen, Y.; Miao, J.; Li, X.; Shan, S.; Ren, R. Obstacle detection in dangerous railway track areas by a convolutional neural network. *Meas. Sci. Technol.* **2021**, *32*, 105401. [CrossRef]
23. He, D.; Zou, Z.; Chen, Y.; Liu, B.; Yao, X.; Shan, S. Obstacle detection of rail transit based on deep learning. *Measurement* **2021**, *176*, 109241. [CrossRef]
24. He, D.; Zou, Z.; Chen, Y.; Liu, B.; Miao, J. Rail Transit Obstacle Detection Based on Improved CNN. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 2515114. [CrossRef]
25. Wang, W.; Wang, S.; Guo, Y.; Zhao, Y. Obstacle detection method of unmanned electric locomotive in coal mine based on YOLOv3–4L. *J. Electron. Imaging* **2022**, *31*, 023032. [CrossRef]
26. Chen, Y.; Lu, C.; Wang, Z. Detection of foreign object intrusion in railway region of interest based on lightweight network. *J. Jilin Univ.* **2021**, *52*, 2405–2418.
27. Han, L.; Zheng, P.; Li, H.; Jiangfan, C.; Zexi, H.; Zutao, Z. A novel early warning strategy for right-turning blind zone based on vulnerable road users detection. *Neural Comput. Applic* **2022**, *34*, 6187–6206. [CrossRef]
28. Dong, C.; Pang, C.; Li, Z.; Zeng, X.; Hu, X. PG-YOLO: A Novel Lightweight Object Detection Method for Edge Devices in Industrial Internet of Things. *IEEE Access* **2022**, *10*, 123736–123745. [CrossRef]
29. Hao, S.; Zhang, X.; Ma, X. Foreign object detection in coal mine conveyor belt based on CBAM-YOLOv5. *J. China Coal Soc.* **2022**, *47*, 4147–4156.
30. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Juan, PR, USA, 17–19 June 2018; pp. 4510–4520. [CrossRef]
31. Zhang, Q.-L.; Yang, Y.-B. SA-Net: Shuffle Attention for Deep Convolutional Neural Networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 6–11 June 2021; pp. 2235–2239. [CrossRef]
32. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [CrossRef] [PubMed]
33. Farid, A.; Hussain, F.; Khan, K.; Shahzad, M.; Khan, U.; Mahmood, Z. A Fast and Accurate Real-Time Vehicle Detection Method Using Deep Learning for Unconstrained Environments. *Appl. Sci.* **2023**, *13*, 3059. [CrossRef]
34. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
35. Zhou, L.; Min, W.; Lin, D.; Han, Q.; Liu, R. Detecting Motion Blurred Vehicle Logo in IoV Using Filter-DeblurGAN and VL-YOLO. *IEEE Trans. Veh. Technol.* **2020**, *69*, 3604–3614. [CrossRef]
36. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the 15th European Conference; Springer: Munich, Germany, 2018; pp. 3–19.
37. Esteves, R.M.; Hacker, T.; Rong, C. Competitive K-Means, a New Accurate and Distributed K-Means Algorithm for Large Datasets. In Proceedings of the IEEE 5th International Conference on Cloud Computing Technology and Science, Bristol, UK, 2–5 December 2013; pp. 17–24.
38. Wang, W.; Wang, S.; Guo, Y.; Zhao, Y.; Tong, J.; Yang, T. Detection method of obstacles in the dangerous area of electric locomotive driving based on MSE-YOLOv4-Tiny. *Meas. Sci. Technol.* **2022**, *33*, 115403. [CrossRef]
39. Bargoti, S.; Underwood, J. Deep fruit detection in orchards. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 3626–3633.
40. Ding, W.; Taylor, G. Automatic moth detection from trap images for pest management. *Comput. Electron. Agric.* **2016**, *123*, 17–28. [CrossRef]
41. Ge, P.; Guo, L.; He, D.; Huang, L. Light-weighted vehicle detection network based on improved YOLOv3-tiny. *Int. J. Distrib. Sens. Netw.* **2022**, *18*, 15501329221080665. [CrossRef]
42. Li, X.; Pan, J.; Xie, F.; Zeng, J.; Li, Q.; Huang, X.; Liu, D.; Wang, X.; Huang, X. Fast and accurate green pepper detection in complex backgrounds via an improved Yolov4-tiny model. *Comput. Electron.* **2021**, *191*, 106–115. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.