applied sciences

MDPI

Article

Research on Automatic Counting of Drill Pipes for Underground Gas Drainage in Coal Mines Based on YOLOv7-GFCA Model

Tiyao Chen *, Lihong Dong and Xiangyang She

College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an 710600, China
* Correspondence: yao7368@126.com

Abstract: Gas explosions threaten the safety of underground coal mining. Mining companies use drilling rigs to extract the gas to reduce its concentration. Drainage depth is a key indicator of gas drainage; accidents will be caused by going too deep. Since each drill pipe has the same length, the actual extraction depth is equivalent to the number of drill pipes multiplied by the length of a single drill pipe. Unnecessary labor is consumed and low precision is achieved by manual counting. Therefore, the drill pipe counting method of YOLOv7-GFCA target detection is proposed, and the counting is realized by detecting the movement trajectory of the drilling machine in the video. First, Lightweight GhostNetV2 is used as the feature extraction network of the model to improve the detection speed. Second, the (Fasternet-Coordinate-Attention) FCA network is fused into a feature fusion network, which improves the expression ability of the rig in complex backgrounds such as coal dust and strong light. Finally, Normalized Gaussian Wasserstein Distance (NWD) loss function is used to improve rig positioning accuracy. The experimental results show that the improved algorithm reaches 99.5%, the model parameters are reduced by 2.325×10^6, the weight file size is reduced by 17.8 M, and the detection speed reaches 80 frames per second. The movement trajectory of the drilling rig target can be accurately obtained by YOLOv7-GFCA, and the number of drill pipes can be obtained through coordinate signal filtering. The accuracy of drill pipe counting reaches 99.8%, thus verifying the feasibility and practicability of the method.

Keywords: underground coal mine; gas extraction; YOLOv7-GFCA; GhostNetV2; FCA; NWD loss function; target detection

Citation: Chen, T.; Dong, L.; She, X. Research on Automatic Counting of Drill Pipes for Underground Gas Drainage in Coal Mines Based on YOLOv7-GFCA Model. *Appl. Sci.* **2023**, *13*, 10240. https://doi.org/10.3390/app131810240

Academic Editors: Yosoon Choi and Nikolaos Koukouzas

Received: 25 July 2023
Revised: 31 August 2023
Accepted: 11 September 2023
Published: 12 September 2023

1. Introduction

China's growing economy has resulted in an increased demand for coal resources [1]. However, the drilling site environment poses unique challenges due to its complexity, including the presence of significant coal dust and the risk of gas explosions. To mitigate these risks, coal mining enterprises commonly employ drilling rigs for gas extraction [2]. Not only is gas recovered as an energy source through this approach, but also the occurrence of gas explosions is minimized. The critical parameter in gas drainage operations is served by the drilling depth, as safety hazards can be induced by drilling too deep. In practical production, the drilling depth is determined by the total length of the drill pipe. Since each drill pipe has a consistent length, counting the number of drainage drill pipes enables the calculation of the drilling depth, ensuring safe gas extraction.

As coal mines embrace intelligent construction, researchers are increasingly utilizing image processing algorithms to enable intelligent production using existing video surveillance systems. Dong et al. [3] introduced an enhanced Camshift algorithm for the real-time capture of drill pipe targets to facilitate accurate drill pipe counting. Meanwhile, Dong et al. [4] proposed a corner detection method combined with the pyramid optical flow technique to estimate the optical flow field of moving drilling rigs, employing frame

skipping. However, limitations are encountered by the traditional manual feature extraction approach in the complex mine environment, which is characterized by uneven lighting and interference from coal dust. As a result, inaccurate statistical outcomes are yielded by this traditional method.

In recent years, significant advancements and successful applications in various image-related tasks were witnessed in deep learning. In line with this trend, researchers have explored its potential for drill pipe counting. Gao et al. [5] presented an improved version of the ResNet network for binary classification. They utilized the integral method to filter video classification confidence and counted the number of falling edges in the confidence curve, thereby achieving quantitative drill pipe counting. Du et al. [6] involved the use of spatio-temporal graph convolutional neural networks to identify the unloading action of drilling rigs, enabling the determination of the number of drill pipes. However, the focus of these methods is primarily on distinguishing the actions of rods being loaded and unloaded by workers, and their applicability is limited across different working scenarios. Additionally, the drill pipe counting speed in these approaches is slow, posing challenges in meeting the demands of real working environments.

In the gas extraction process, the drilling rig identifies the reciprocating drilling rig target along a specific straight line and subsequently counts the number of reciprocations. Drilling depth is determined by the number of drill pipes. To address this, a target detection-based method is proposed for the efficient and accurate counting of light drilling rigs in complex environments. Advanced target detection technology is employed by this method to swiftly and precisely detect the moving targets of drilling rigs, ultimately enhancing the overall safety and efficiency of gas drainage operations.

Object detection algorithms can be categorized into one-stage and two-stage approaches. Representative two-stage algorithms include Faster Regions with CNN features(R-CNN) [7] and Mask-RCNN [8]. While higher detection accuracy is offered by these algorithms, their speed is comparatively slower. In the context of coal mine drilling sites, these algorithms are unsuitable for real-time deployment of industrial equipment. On the other hand, first-stage algorithms like Single Shot MultiBox Detector(SSD) [9] and You Only Look Once(YOLO) series [10] exhibit fast detection speeds. In fact, in certain, specific domains, comparable detection accuracy can be achieved by these algorithms to that of two-stage algorithms.

At present, a significant amount of research is conducted by many scholars on coal mine underground targets using target detection technology, leading to the achievement of certain research results [11–13]. However, targets such as miners, safety equipment, and coal blocks are mainly detected in the literature [11–13], and the drilling rig targets in the gas drainage environment are not studied.

In view of this, later scholars conducted research on drilling rig targets. Yu et al. [14] combined YOLOv5 with the DeepSort algorithm, and the detection speed of the drilling rig in the drilling field environment reached 64 FPS, but the accuracy is only 90.5%, and there is still room for improvement. Tan et al. [15] used computer vision technology to achieve accurate counting of drill rods and calculated the corresponding drilling depth, and to track the drilling rig head to form a trajectory to quantify the number of drilling holes, but the detection speed is slow and does not meet the real-time requirements.

Considering the aforementioned challenges, a novel rig detection algorithm YOLOv7-GFCA, is introduced for drill pipe counting. By capturing real-time location information of the drilling rig, the determination of the number of drill pipes is enabled, leading to accurate measurement of drilling depth and enhanced gas drainage efficiency. The key contributions of this research are summarized as follows:

(i) GhostNetV2 is used as the backbone network to reduce the amount of parameters brought about by redundant feature calculation and improve the detection speed;
(ii) The Lightweight FasterNet and the improved coordinate attention(CA) are integrated as a feature fusion network (FCA) to reduce the interference caused by coal dust and light;

(iii) The Normalized Gaussian Wasserstein Distance(NWD) loss function is introduced to improve the positioning accuracy of the drilling rig target.

2. Related Work and Research Methods

2.1. YOLOV7 Target Detection Model

The YOLOv7 [16] Model is representative of a one-stage target detection algorithm. The network consists of four parts, namely image input, backbone network, feature fusion network, and output. The backbone network is used to extract features, and the feature fusion network is used to fuse features of different scales. The output of the network is three feature maps of different scales.

The backbone network is composed of Efficient Layer Aggregation Network (ELAN) and Max Pooling(MP) structures. ELAN is composed of multiple Conv-BatchNorm-Silu (CBS) modules, and its input and output feature sizes remain unchanged, but the structure is relatively redundant and there are too many parameters. The MP layer is simultaneously down-sampled through maximum pooling, and the number of channels is combined through Concat (an operation for concatenating tensors). The E-ELAN layer and the MPConv layer alternately halve the length and width to extract features. The Spatial Pyramid Pooling and Cross Stage Partial (SPPCSPC) module stitches the backbone network and the detection head network through the maximum pooling of different scales.

Finally, the number of channels is adjusted through the RepConv detection head network, and the three parts of prediction confidence, category and regression box are predicted. Figure 1 is a network structure diagram of YOLOV7.

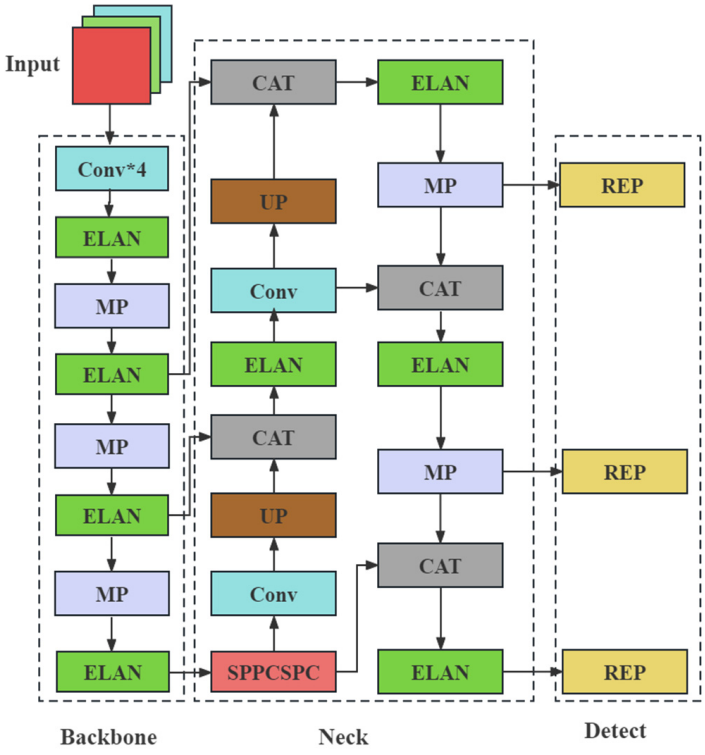


Figure 1. YOLOv7 target detection algorithm module diagram. (Conv is the convolution operation, ELAN is an efficient grid structure, MP is maximum pooling operation, CAT is splicing operation, UP is upsampling operation, and REP is detection head network, Conv*4 is the input tensor passing through four convolutional layers continuously).

2.2. Lightweight Network

The term "Lightweight network" refers to a specialized technology wherein high precision is maintained in neural networks while reducing computational complexity. In

these networks, efforts are made to minimize complexity and computational costs. Prominent lightweight networks, such as ShuffleNet [17], GhostNetV2 [18], and FasterNet [19], extract image features efficiently, enabling the accomplishment of various visual tasks with decreased computational resources. Addressing the total number of channels in GhostNetV2 involves rearranging them within the block using a small number of channels, leading to a reduction in model storage space and computation. The efficiency of model training is enhanced in FasterNet through the utilization of spatially separable convolution and multi-scale training strategies. In scenarios characterized by constrained resources, effective solutions are presented by these lightweight networks, resulting in comparable performance to more intricate counterparts being achieved.

2.3. Attention Mechanism

The attention mechanism is designed to adaptively allocate attention weights based on task objectives. This enables the prioritization of relevant regions by neural networks while diminishing the significance of irrelevant parts. The attention mechanism covers five categories: time domain, space domain, layer domain, hybrid domain, and channel domain. Various lightweight attention mechanisms, such as SE (Squeeze Excitation) [20], ECA (Efficient Channel Attention) [21], and CA [22] attention have been developed. These mechanisms allow attention values to be adjusted for focusing the model on the target area, thereby mitigating background interference and ultimately enhancing recognition accuracy, especially when dealing with complex backgrounds. By intelligently guiding attention, essential information can be effectively highlighted by the model while the emphasis on irrelevant details is reduced.

2.4. Loss Function

A loss function involving a classification loss and a localization loss is encompassed by object detection tasks. The loss function is used to compute the difference between predicted objects and ground truth, effectively assessing the accuracy of object localization. A higher level of alignment between the predicted and actual positions is indicated by a smaller loss value, signifying more accurate position predictions. However, inaccurate positioning when tracking the drilling rig target can be caused by factors such as vibration and drill sticking. Therefore, a need exists to optimize the calculation of positioning coordinates in order to enhance the accuracy of the positioning process. Efforts focused on addressing these challenges and refining the positioning algorithms can lead to an improved precision in tracking the drilling rig target.

3. Drill Pipe Counting Method Based on YOLOV7-GFCA Network

To further enhance the detection speed of the drilling rig and fulfill the real-time counting requirements of the complex drilling site environment downhole, a new detection network, YOLOV7-GFCA, is proposed in this paper, based on GhostNetV2, and incorporates FCA within the YOLOV7 model. After the model is trained, we input the original working video of the drilling rig into the YOLOV7-GFCA network to obtain the movement trajectory of the drilling rig, filter the coordinate information of the detection frame vibration caused by the movement of the drilling rig through a filtering algorithm, count the number of peaks in the waveform, and calculate the number of drill pipes extracted by the drilling rig to get the final drilling depth; Figure 2 is a flow chart of the overall drill pipe counting framework.

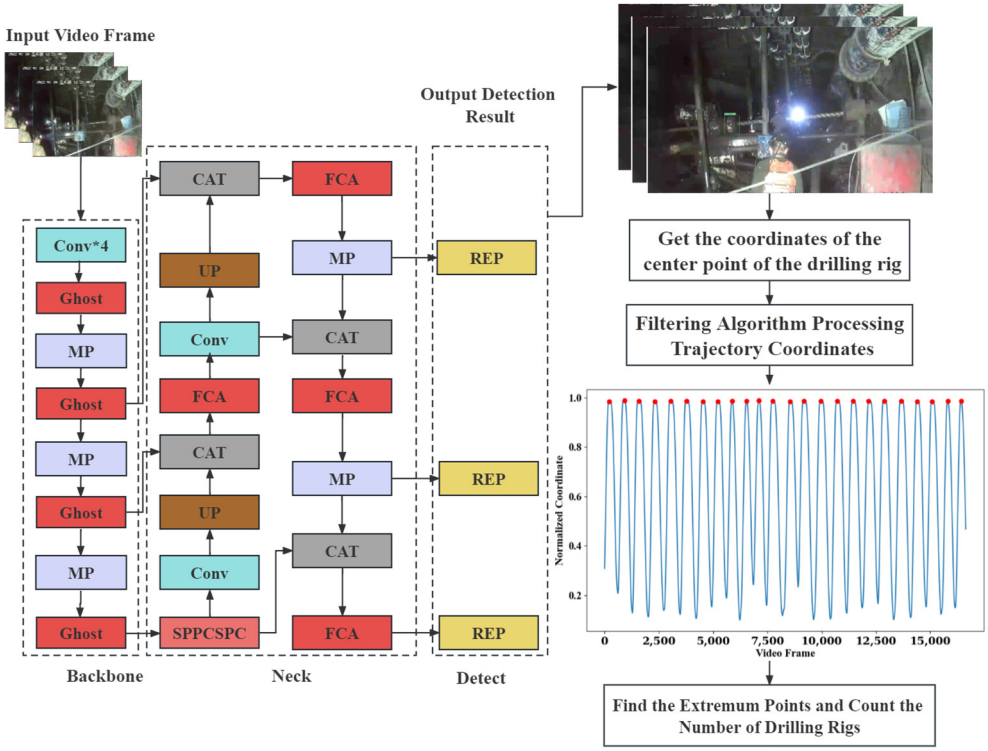


Figure 2. Flow chart of drill pipe counting based on YOLOV7-GFCA target detection network (the red part in the backbone network Ghost and neck network FCA is our improvement, and Conv*4 is the input tensor passing through four convolutional layers continuously).

3.1. Backbone Network Improvements

In the coal mine drilling site environment, a crucial role is played by the YOLOv7 backbone network in the extraction of essential features, including the appearance, color, and shape of the drilling rig. However, the utilization of the ELAN module by the original backbone network introduces feature representation redundancy, leading to increased computational requirements and model complexity. To address these issues, a proposal is made in this study to replace the ELAN structure with GhostNetV2, serving as a more efficient backbone network for feature extraction. With its ability to reduce redundancy and computational overhead, higher efficiency is offered by GhostNetV2, rendering it a suitable choice for enhancing the performance of the detection system in the coal mine drilling environment.

Lightweight GhostNetV2 is comprised of two Ghost modules and one DFC module. The generation of feature maps with an increased number of channels is undertaken by the first Ghost module. Subsequently, the channel dimension of the feature maps is effectively compressed by the second Ghost module, reducing redundancy. Furthermore, the decoupled fully connected (DFC) attention mechanism is employed to capture long-range pixel dependencies both horizontally and vertically. Through the collaborative operation of these combined components, the network's capacity for extracting meaningful features is enhanced. A visual representation of the proposed GhostNetV2 architecture is provided in Figure 3 for reference.

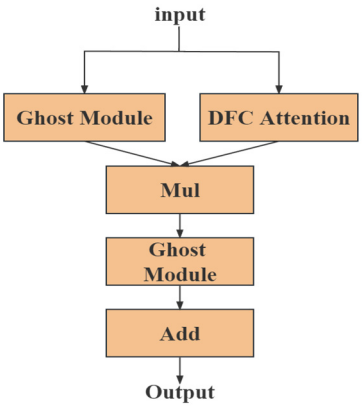


Figure 3. The GhostNetV2 module is composed of a Ghost module and two DFC modules; input and output are the input and output of the feature map, respectively. Mul is feature map multiplication. Add is feature map addition.

Designed for the generation of more feature maps using fewer parameters, the Ghost module serves as a key component in GhostNetV2. In each convolutional layer of the neural network, numerous multiplication and accumulation operations are involved in the convolution operation. This results in a substantial computational resource demand during the model's training and inference processes. To generate feature maps, the Ghost module employs simple linear transformation operations in place of convolution operations, thereby assisting the network in maintaining a high level of representational capacity. When considering the same number of input and output feature maps, fewer parameters are required by the Ghost module compared to conventional convolutional modules.

All pixels in the row and column of a certain pixel position in DFC directly participate in the calculation of point attention, and all pixel positions in this area also indirectly participate in the calculation of point attention through decoupling operations, thus avoiding reorganization and recombination and other computationally intensive operations.

The input is the original feature $Z \in R^{W \times H \times C}$, and it is regarded as a $H \times WToken$; $Z \in \{Z_{11}, Z_{12}, \ldots, Z_{HW}\}$ is the generated feature map. The width, height, and number of channels of the input feature map are represented by W, H, and C, respectively, where *Token* is represented a single element in the input sequence. The attention map generated by the fully connected layer is generated by Equation (1). The long-range correlations along both horizontal and vertical directions are then captured separately.

$$\begin{gathered} a_{hw} = \sum_{h',w'} F_{hw,h'w'} \otimes Z_{h'w'} \\ h = 1, 2, \ldots, H, w = 1, 2, \ldots, W \end{gathered} \tag{1}$$

where $\otimes$ is represented matrix multiplication; F is the weight coefficient in the fully connected layer; h, w are the width and height of the feature map, and $\{a_{11}, a_{12}, \ldots, a_{HW}\}$ is the generated attention map.

Equation (2) is used to capture aggregated features in the horizontal direction. The fully connected layer feature F is decoupled into F^H along the horizontal direction.

$$\begin{gathered} a'_{hw} = \sum_{h'=1}^{H} F^H_{h,h'w} \otimes Z_{h'w} \\ h = 1, 2, \ldots, H, w = 1, 2, \ldots, W \end{gathered} \tag{2}$$

Equation (3) is used to capture aggregated features in the vertical direction. The fully connected layer feature F is decoupled into F^W along the vertical direction.

$$a_{hw} = \sum_{w'=1}^{W} F^W_{w,hw'} \otimes a'_{hw'} \quad h = 1, 2, \ldots, H, w = 1, 2, \ldots, W \tag{3}$$

General formulations for DFC attention are represented by Equations (2) and (3), which cluster pixels along the horizontal and vertical directions, respectively. By decoupling operations, convolution can be conveniently executed, thereby eliminating the time-consuming operations of tensor reshaping and transposing that impact the actual inference speed. Operated in parallel with the first Ghost module, the DFC attention branch serves to enhance extended features. Following this, the enhanced features are fed into the second Ghost module to generate output features. Within this branch, it is possible to capture long-distance dependencies among pixels situated at distinct spatial positions. Possessing fewer parameters, this branch contributes to the model's augmented capacity for feature extraction.

3.2. Feature Fusion Network Improvement

Feature fusion networks play a key role in the integration of multi-scale features. In the original feature fusion network, a deep residual structure is adopted. As the depth of the network increases, a surplus of redundant features will be generated. These features contain complex backgrounds, such as coal dust and strong light, which may adversely affect the detection of drilling rigs. Therefore, an optimized method is needed to deal with the challenges brought by the existence of redundant features and complex backgrounds, and to ensure accurate and efficient detection of rigs.

ELAN is replaced by the FasterNet module as the main component of the feature fusion network. Figure 4a shows the components of FasterNet, which consists of a Pconv module and two Convs. This combination can extract spatial features more efficiently by reducing redundant computation and memory access. However, due to the special operating environment of coal mines, neural networks can easily extract useless features, such as coal dust, strong light, and miners who are working. In order to improve the Fasternet module's ability to express rigs, we introduced an improved CA attention mechanism on the left, which makes the network more focused on extracting rig features in complex environments, while reducing the extraction of redundant features and improving the accuracy of detecting rigs. Figure 4b is the overall module structure diagram of FCA.

The reduced computational load achieved by the FasterNet module can be attributed to the clever design of the PConv. Unlike the conventional convolution approach that applies a comprehensive convolution to the feature map, PConv exclusively executes convolution on a section of the input channels, leaving the other channels unaltered. To leverage the entirety of channel information, two 1×1 pointwise convolutions are subsequently employed following the PConv to comprehensively extract features from all channels.

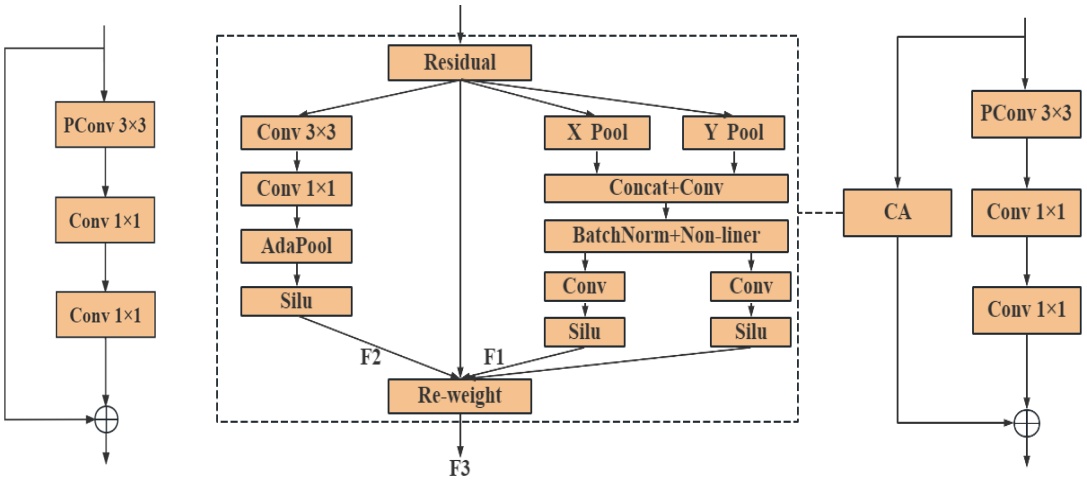
For the input feature map $D \in R^{c \times m \times n}$, use c convolution kernels $k \times k$ for convolution operation and m, n are the height and width of the input feature map. The amount of parameters to calculate the ordinary convolution is generated by Equation (4).

$$Flops_{conv} = m \times n \times k^2 \times c^2 \tag{4}$$

For PConv, it only convolutes some features. When the ratio is set to $1/4$ of the original channel. The amount of parameters to calculate the PConv is generated by Equation (5).

$$Flops_{ponv} = m \times n \times k^2 \times \left(\frac{1}{4}c\right)^2 \tag{5}$$

The amount of parameters generated by PConv is only 1/16 of the amount of ordinary convolution parameters. For the remaining channels, two 1×1 convolution kernels are used for convolution.



(**a**) Fasternet (**b**) FCA

Figure 4. FCA module: (**a**) represents the lightweight convolution module in FasterNet; (**b**) represents the improved FCA module, where we embedded the improved CA attention mechanism module on the left, and zoom in on the display; F2 represents the depth separability introduced convolution branch.

In order to enhance the ability of the device to deal with complex backgrounds, such as coal dust, an improved CA attention mechanism is integrated in the original FasterNet, and a deep convolution branch is added to better extract context information (F_2). This improved attention layer consists of three branches: deep convolution branch, residual branch and attention branch (F_1). Within the attention branch, feature extraction is performed on the feature maps in the horizontal and vertical directions, then the coordinate matrix is derived and the similarity is calculated to obtain the attention matrix (F_1) by Equations (6)–(8). By incorporating the improved CA attention mechanism, the network obtains improved feature representation and enhanced discriminative ability, resulting in enhanced detection performance in coal dust and similar challenging environments.

$$Z_c^h(h) = \frac{1}{W}\sum_{i=0}^{W-1} x_c(h,i) \tag{6}$$

$$Z_c^w(w) = \frac{1}{H}\sum_{j=0}^{H-1} x_c(j,w) \tag{7}$$

$$F_1 = \sigma(c^{1\times1}(\sigma(c^{1\times1}([z^h, z^w])))) \tag{8}$$

where Z^h, Z^w are feature maps representing the vertical and horizontal directions; (i, j) is the coordinate position; σ is the Relu activation function; $c^{1\times1}$ is 1×1 convolution kernel; F_1 is the attention branch output; $x_c(h,i)$ is tensor of channel c at height h; and $x_c(j,w)$ is tensor of channel c at width w.

In the deep convolution branch, the importance of channels is initially learned using 3×3 convolution, and upon output the information relationship between channels is captured using 1×1 convolution. Through the Silu activation function, only one multiplication and one calculation of the Sigmoid function are involved, so the amount of

parameters is small, and the sensitivity of the network to the characteristics of the rig is enhanced after using the global adaptive pooling in Equation (9).

$$F_2 = AdaPool(\partial(c^{1\times1}(c^{3\times3}(x)))) \tag{9}$$

where F_2 is the deep convolution branch; ∂ is the Silu activation function; $c^{3\times3}$ is 3×3 convolution kernel; and $AdaPool(\cdot)$ is the global pooling operation.

After matrix multiplication of the attention branch and the depth convolution branch, they are added to the original residual branch x to obtain the final output F_3 in Equation (10).

$$F_3 = x + F_1 \times F_2 \tag{10}$$

Less computation is expended by the above method to enhance the expressive capability of the drilling rig in the presence of a coal dust background, thereby meeting the accuracy requirements of detection while reducing weight.

3.3. More Precise Loss Function

The position loss function is represented the ratio between the predicted box A of the rig chuck and the manually marked real box B by Equation (11).

$$IOULoss = 1 - \frac{|A \cap B|}{|A \cup B|} \tag{11}$$

The smaller the ratio of the two, the more accurate the position prediction of the drill chuck. However, the *IOU* is sensitive to the configuration deviation of the detection frame with a small object, and the anchor frame may be flipped, which leads to difficulty in network convergence and reduces the detection performance.

A new metric, NWD [23] (Normalized Wasserstein Distance), is introduced to replace the *IOU* metric for measuring the similarity between two bounding boxes. First, denote the bounding box $R = (C_x, C_y, W, H)$,(C_x, C_y) is the center coordinates, W, H are width and height, respectively. Then, calculate the ellipse equation inscribed in the bounding box of the drilling rig, which can be expressed as Equation (12).

$$\frac{(x-\mu_x)^2}{\xi^2{}_x} + \frac{(y-\mu_y)^2}{\xi^2{}_y} = 1 \tag{12}$$

where (μ_x, μ_y) is the coordinates of the center point of the ellipse; and (ξ_x, ξ_y) is the length of the semi-axis of the ellipse. Equation (13) is the probability density function of the two-dimensional Gaussian distribution.

$$f(\mathrm{x}|\mu, \sum) = \frac{\exp(-\frac{1}{2}(\mathrm{x}-\mu)^{\mathrm{T}} \sum^{-1}(\mathrm{x}-\mu))}{2\pi|\sum|^{\frac{1}{2}}} \tag{13}$$

where x is coordinate (x, y)—it is a two-dimensional vector; μ is the mean vector; and $\sum$ is the covariance matrix of Gaussian distribution

$$(\mathrm{x}-\mu)^{\mathrm{T}} \sum^{-1} (\mathrm{x}-\mu) = 1 \tag{14}$$

When the conditions of Equation (14) are satisfied, the ellipse in Equation (12) will be the density profile of the two-dimensional Gaussian distribution with

$$\mu = \begin{bmatrix} C_x \\ C_y \end{bmatrix}, \sum = \begin{bmatrix} \frac{W^2}{4} & 0 \\ 0 & \frac{H^2}{4} \end{bmatrix}$$

Therefore, we can convert the target box and detection box into the similarity between two Gaussian distributions, and Wasserstein distance between target box A and detection box B is calculated by Equation (15).

$$W_2^2(N_A, N_B) = \|\mu_a - \mu_b\|_2^2 + \left\|\sum_a^{1/2} - \sum_b^{1/2}\right\|_F^2 \tag{15}$$

where $\|\cdot\|_F$ is Frobenius norm; μ is the mean vector; and $\sum$ is the covariance matrix. W_2^2 is the degree of similarity between two distributions.

But W_2^2 is a distance measure, which cannot directly measure the similarity between the target frame and the candidate frame, and is further converted into the normalized *NWD* distance in Equation (16).

$$NWD(N_A, N_B) = \exp\left(-\frac{\sqrt{W_2^2(N_A, N_B)}}{C}\right) \tag{16}$$

Aiming at the target of underground drilling rigs in coal mines, in order to reduce the loss of coordinates caused by the vibration of drilling rigs, the *NWD* measurement method is used to effectively replace the *IOU* measurement method.

The similarity between the target frame and the detection frame is effectively measured by this measurement method, contributing to the enhancement of the positioning accuracy of the drilling rig.

4. Data Sources and Evaluation Indicators

4.1. Data Source

The dataset utilized in this experiment originates from a coal mine situated in Xianyang City, Shanxi Province, China. Collect real-time operation videos of automatic drilling rigs and fully hydraulic deep-hole drilling rigs at coal mine gas extraction sites. The collected videos were segmented, resulting in a processed dataset comprising a total of 3714 images that encompass diverse scenes. Given the variations in models and sizes among different drilling rigs, while the chucks of each drilling rig are black cylinders, this experiment involved the labeling of the drilling rig chucks. The LabelImg labeling tool was employed to annotate the images, and subsequently, the dataset was partitioned into training, validation, and test datasets in a ratio of 4:1:1. Each dataset consisted of 2475, 619, and 620 images, respectively.

4.2. Evaluation Indicators

Several evaluation metrics are employed in this paper, which include the mean Average Precision (*mAP*) with an Intersection over Union (IoU) threshold of 0.5, model weight measured in megabytes (MB), detection speed quantified in frames per second (*FPS*), and model parameters. The calculation of map is related to *precision* and *recall*.

$$precision : p = \frac{TP}{TP + FP} \tag{17}$$

$$recall : r = \frac{TP}{TP + FN} \tag{18}$$

$$mAP\ @\ 0.5 = \int_0^1 p(r)\mathrm{d}r \tag{19}$$

where TP is the number of positive cases that are correctly classified; FP is the number of negative cases that are misclassified; FN is the number of positive cases that are misclassified; $p(r)$ is a two-dimensional curve function of the precision rate on the recall rate; and the *mAP* value is the area of the two-dimensional graph formed by calculating the accuracy rate and the recall rate.

The model weight is the size of the weight file generated during the training process; the model parameter quantity indicates the parameter quantity required for model training. The detection speed is expressed as the maximum number of frames that can be detected within 1 s; *FPS* is the number of frames transmitted per second in the screen. The more frames per second, the smoother the action displayed on the screen. Equation (19) is the calculation process of *FPS*.

$$FPS = \frac{FrameNum}{ElapsedTime} \tag{20}$$

where *FrameNum* is the total number of frames of a video, and *ElapsedTime* is the time consumed to run the video.

5. Experiments and Discussion

5.1. Model Training

To prevent overfitting, various data augmentation techniques were employed to enhance the input images before model training. These techniques include random flipping, color transformations, rotation, contrast adjustment, and mosaic. The experiments were conducted on a Windows 11 operating system with an AMD Ryzen 5000 CPU (Lenovo Ryzen 7 5800H, Beijing, China) and an NVIDIA GeForce RTX 3080 Laptop GPU. The implementation utilized PyTorch 1.8.1, CUDA 11.1, and Python 3.8.13 as the deep learning framework. The learning rate was set to 0.001, with a momentum value of 0.938. The input image size was 640 × 640, and the training process spanned 150 epochs, allowing for comprehensive model optimization and learning.

5.2. Ablation Experiment

In all the conducted ablation experiments, the number of training iterations and hyperparameters remained constant, while the input image resolution of the drill chuck was set to 640 × 640 pixels. The self-built coal mine drill pipe dataset was employed for conducting the experiments, enabling the assessment of the functionality and effectiveness of each module. The experimental findings and results are presented in Table 1, providing valuable insights into the performance of the proposed approach.

Table 1. Compared the effectiveness of different modules on the original YoloV7 algorithm.

Model Type	Map@0.5/%	FPS	Model Weight/MB	Parameter Amount/(10^6)
YoloV7	96.4%	47	71.8	9.319
+Ghost	95.8%	70	65.5	8.478
+FCA	98.2%	62	50.8	7.835
+NWD	97.3%	47	71.8	9.319
+Ghost + FCA + NWD	99.5%	80	54.3	6.994

It is shown in Table 1 above that after the changes of different modules on YOLOv7, the accuracy and speed have been improved to varying degrees.

(i) In order to reduce the parameters of the network, the ELAN module in the original backbone network was replaced with the GhostNetV2 module, the weight of the model was reduced by 6.3 M, the number of parameters was reduced by 8.41×10^5, and the inference speed was increased by 23 frames/s.

(ii) In order to improve the expression ability of drilling rigs in complex backgrounds, such as coal dust and strong light, after using the lightweight FCA with integrated attention, the average accuracy is increased by 1.8%, the weight of the model is reduced by 21 M, and the number of parameters is reduced by 1.484×10^6. The inference speed increased by 15 frames/s.

(iii) For the target of the downhole drilling rig, the positioning accuracy improved. After using the NWD loss, the accuracy increased by 0.9%.

5.3. Comparative Experiment

The experimental data is real and effective, and it comes from the real drilling site environment of different coal mines. The same drilling rig data sets are used in the comparative experiments. In order to verify the effectiveness of the algorithm proposed in this paper, it is verified on the self-built drilling rig dataset. At the same time, the current mainstream target detection algorithm is compared in Figure 5 and Table 2. The model has the same depth and width. The resolution of the input image is 640 × 640.

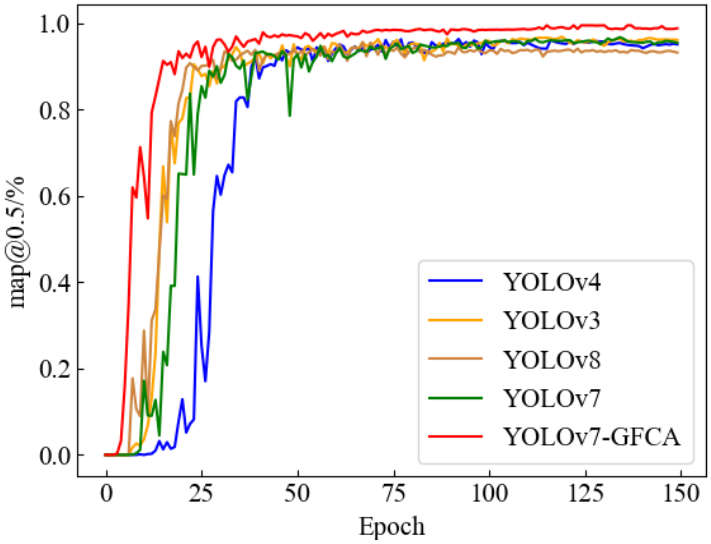


Figure 5. Comparison between the YOLOv7-GFCA algorithm and the current mainstream target detection models.

Table 2. Comparison between YOLOV7-GFCA and mainstream algorithms.

Model Type	Map@0.5/%	FPS	Model Weight/MB	Parameter Amount/(10^6)
YOLOv3	96.2%	41	71.4	9.308
YOLOv4	95.4%	35	70.2	9.115
YOLOv8	93.8%	45	70.2	9.124
YOLOv7	96.4%	47	71.8	9.319
YOLOv7-GFCA	99.5%	80	54.3	6.994

Both the convergence speed and accuracy of YOLOV7-GFCA are better than the original YOLOV7 algorithm. The identification of drilling rigs has achieved an accuracy rate of 99.5%. Compared with other network models, the parameter quantities were reduced to 2.314×10^6, 2.121×10^6, 2.13×10^6, and 2.325×10^6, respectively, and the model volume reduced by 17.1 M, 15.9 M, 15.9 M, 17.5 M. The inference speed increased by 39 frames/s, 45 frames/s, 35 frames/s, and 33 frames/s, respectively. The results of the simulation under the different scenes is shown in Figure 6.

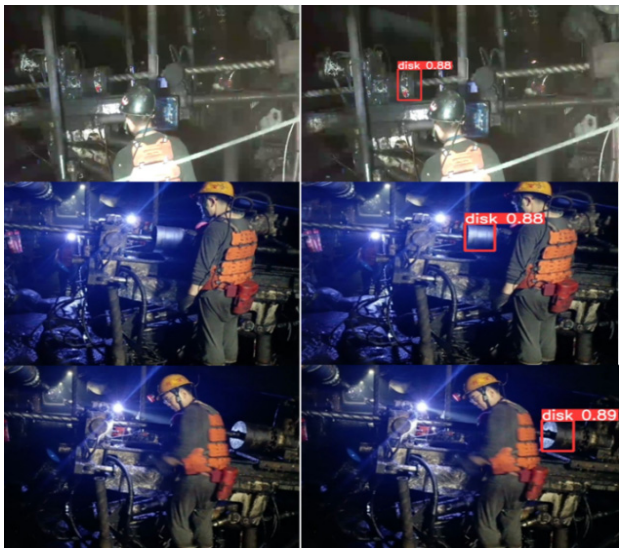


Figure 6. YOLOv7-GFCA model detection results of drilling rigs.

5.4. Counting the Number of Drilling Rigs

To facilitate the measurement of drilling depth, the utilization of YOLOV7-GFCA is incorporated in the proposed method for drilling rig target detection. The movement trajectory of the drilling rig is captured, thereby enabling the calculation of the number of drill pipes driven by the rig and, ultimately, the determination of the drilling depth.

YOLOV7-GFCA is employed to perform real-time detection on the working video of the drilling rig. The movement trajectory of the drilling rig is captured, and the position coordinate information of the drilling rig is stored in chronological order within a CSV file. Subsequent to this, mean filtering and normalization are applied to the position coordinate information of the drilling rig. The coordinate information of the detection frame vibration, caused by the drilling rig's movement, is filtered. The count of the maximum vertices in the waveform is used to determine the number of drill pipes. Given that the length of each drill pipe used in the coal mine is consistent, and the drilling rig executes a reciprocating motion along a designated straight line, the entry of a drill pipe is inferred. Accordingly, the drilling depth is accurately computed based on the drill pipe count. The filtering process of the drilling rig's position coordinate information is visually depicted in Figure 7. The vertical coordinate corresponds to the drilling rig's coordinate information acquired via YOLOV7-GFCA, while the abscissa denotes the real-time working video frame number of the drilling rig.

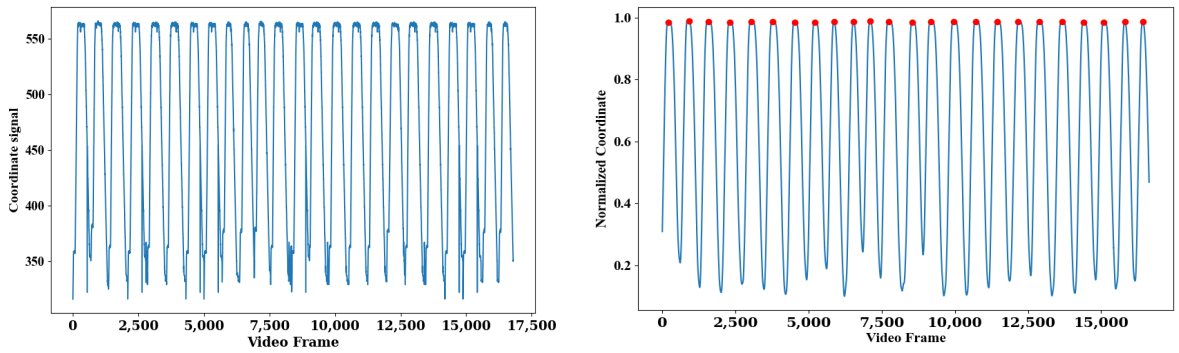


Figure 7. The process of processing the coordinate signal of the drilling rig: (**a**) the original left signal is obtained by the YOLOv7-GFCA target detection model; (**b**) coordinate signal after mean filtering and normalization processing. The red dot are the number of wave crests, which is also the number of drill pipes.

In order to verify the application effect of the drill pipe counting method proposed in this paper in the coal mine drill pipe counting task, the manual counting method and the improved algorithm were tested using the complete drilling rig working video recorded in the fully mechanized mining face of the coal mine. See Table 3 for statistical results.

Table 3. Comparison between the manual counting method and the algorithm before and after improvement.

Method	Actual Number	Detect Number	Actual Total Depth/(m)	Detect Total Depth/(m)	Accuracy/(%)
Manual Counting	500	484	750	726	96.8
YOLOV7	500	487	750	730.5	97.4
YOLOV7-GFCA	500	499	750	748.5	99.8

It is shown in Table 3 that the manual counting method is easily affected by subjective factors. When the number of drill pipes is large, the accuracy may decrease with time, resulting in a low drill pipe counting accuracy, and based on the image detection method, the counting is more stable. Among them, the counting effect based on the YOLOV7-GFCA method is the best, reaching 99.8%, which is higher than the manual counting method and the original algorithm, and the drill pipe counting function can be well realized within the allowable range of error.

6. Conclusions

For the gas drainage environment, the drill pipe counting speed is slow and the accuracy is low, and the traditional image processing algorithm is easy to lose the drilling rig target and other problems may arise, therefore, the YOLOV7-GFCA model detection drilling rig was proposed:

(i) Redundant features in the coal mine environment are effectively eliminated through the incorporation of the GhostNetV2 network, resulting in the acceleration of reasoning speed and the creation of a lightweight model. By embedding the lightweight FCA module into YOLOv7, the salience of the drilling rig is augmented, particularly within complex backgrounds. The feature expression capability of the rig is consequently improved, leading to heightened precision in detection outcomes. The conventional IOU measurement method is replaced by the NWD measurement method. This substitution mitigates the effects of drilling rig vibration, drill sticking, and other pertinent factors, consequently further enhancing the accuracy of drilling rig positioning.

(ii) The exceptional performance of YOLOV7-GFCA is demonstrated by experimental results, showcasing a remarkable detection accuracy of 99.5% at a detection speed of 80 FPS. This surpasses the performance of existing mainstream target detection algorithms. Upon the application of the YOLOV7-GFCA model to real-time dynamic videos of drilling rig working faces, the accuracy of drill pipe counting is impressively elevated to 99.8% through the filtration of motion trajectories. Furthermore, accurate determination of drilling depth is achieved by the proposed method. Compelling validation for the practicality and feasibility of the approach in real-world scenarios is provided by these results.

(iii) Owing to the distinctive challenges inherent in the coal mine environment, YOLOv7-GFCA might encounter limitations in the effective detection of drilling rig targets under specific circumstances. For instance, instances where underground workers directly shine searchlights at the camera or obstruct the camera for extended periods can potentially compromise the performance of the detection function. These situations introduce unique challenges that necessitate further investigation and potential adjustments to the detection algorithm, aiming to enhance robustness in such demanding conditions.

Author Contributions: Conceptualization, T.C. and L.D.; methodology, T.C. and X.S.; software, T.C.; validation, T.C., L.D. and X.S.; formal analysis, T.C.; investigation, T.C.; resources, L.D.; data curation, L.D.; writing—original draft preparation, T.C.; writing—review and editing, X.S.; visualization, T.C.; supervision, X.S.; project administration, L.D.; funding acquisition, L.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data source is explained in the article.

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

References

1. Zheng, Y.; Lu, Q.; Chen, A.; Liu, Y.; Ren, X. Rapid Classification and Quantification of Coal by Using Laser-Induced Breakdown Spectroscopy and Machine Learning. *Appl. Sci.* **2023**, *13*, 8158. [CrossRef]
2. Zhao, H.; Wang, W. Studying the Favorable Zone for Pressure-Relief Gas Extraction by Combining Numerical Investigation and On-Site Application. *Appl. Sci.* **2023**, *13*, 5045. [CrossRef]
3. Dong, L.; Wang, J.; She, X. Drill counting method based on improved Camshift algorithm. *Coal Mine Automat.* **2015**, *018*, 71–76. [CrossRef]
4. Dong, L.; Peng, Y.; Fu, L. Circular Harris corner detection algorithm based on Sobel edge detection. *J. Xian Univ.* **2019**, *39*, 374–380. [CrossRef]
5. Gao, R.; Hao, L.; Liu, B. Research on underground drill pipe counting method based on improved ResNet network. *Coal Mine Automat.* **2020**, *46*, 32–37. [CrossRef]
6. Du, J.; Dang, M.; Qiao, L. Drill pipe counting method based on improved spatial-temporal graph convolution neural network. *Coal Mine Automat.* **2023**, *49*, 90–98. [CrossRef]
7. Sun, X.; Wu, P.; Hoi, S.C. Face detection using deep learning: An improved faster RCNN approach. *Neurocomputing* **2018**, *299*, 42–50. [CrossRef]
8. López-Barrios, J.D.; Escobedo Cabello, J.A.; Gómez-Espinosa, A.; Montoya-Cavero, L.-E. Green Sweet Pepper Fruit and Peduncle Detection Using Mask R-CNN in Greenhouses. *Appl. Sci.* **2023**, *13*, 6296. [CrossRef]
9. Haji Mohd, M.N.; Mohd Asaari, M.S.; Lay Ping, O.; Rosdi, B.A. Vision-Based Hand Detection and Tracking Using Fusion of Kernelized Correlation Filter and Single-Shot Detection. *Appl. Sci.* **2023**, *13*, 7433. [CrossRef]
10. Sportelli, M.; Apolo-Apolo, O.E.; Fontanelli, M.; Frasconi, C.; Raffaelli, M.; Peruzzi, A.; Perez-Ruiz, M. Evaluation of YOLO Object Detectors for Weed Detection in Different Turfgrass Scenarios. *Appl. Sci.* **2023**, *13*, 8502. [CrossRef]
11. 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. [CrossRef]
12. Wang, Y.; Guo, W.; Zhao, S.; Xue, B.; Zhang, W.; Xing, Z. A Big Coal Block Alarm Detection Method for Scraper Conveyor Based on YOLO-BS. *Sensors* **2022**, *22*, 9052. [CrossRef] [PubMed]
13. Jo, B.W.; Khan, R.M.A. An event reporting and early-warning safety system based on the internet of things for underground coal mines: A case study. *Appl. Sci.* **2017**, *7*, 925. [CrossRef]
14. Yu, Y.; Zhao, J.; Yi, C.; Zhang, X.; Huang, C.; Zhu, W. Drill-Rep: Repetition counting for automatic shot hole depth recognition based on combined deep learning-based model. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106302. [CrossRef]
15. Tan, T.; Changfang, G.; Guohua, Z.; Wenhua, J. Research and application of downhole drilling depth based on computer vision technique. *Process. Saf. Environ.* **2023**, *174*, 531–547. [CrossRef]
16. Wu, D.; Jiang, S.; Zhao, E.; Liu, Y.; Zhu, H.; Wang, W.; Wang, R. Detection of Camellia oleifera fruit in complex scenes by using YOLOv7 and data augmentation. *Appl. Sci.* **2022**, *12*, 11318. [CrossRef]
17. Fu, Y.; Lu, Y.; Ni, R. Chinese Lip-Reading Research Based on ShuffleNet and CBAM. *Appl. Sci.* **2023**, *13*, 1106. [CrossRef]
18. Chen, W.; Wang, X.; Yan, B.; Chen, J.; Jiang, T.; Sun, J. Gas Plume Target Detection in Multibeam Water Column Image Using Deep Residual Aggregation Structure and Attention Mechanism. *Remote Sens.* **2023**, *15*, 2896. [CrossRef]
19. Chen, J.; Kao, S.H.; He, H.; Zhuo, W.; Wen, S.; Lee, C.H.; Chan, S.H.G. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023.
20. Jiang, K.; Xie, T.; Yan, R.; Yan, R.; Wen, X.; Li, D.; Jiang, H.; Jiang, N.; Feng, L.; Duan, X.; et al. An attention mechanism-improved YOLOv7 object detection algorithm for hemp duck count estimation. *Agriculture* **2022**, *12*, 1659. [CrossRef]
21. Lei, Y.; Pan, D.; Feng, Z.; Qian, J. Lightweight Human Ear Recognition Based on Attention Mechanism and Feature Fusion. *Appl. Sci.* **2023**, *13*, 8441. [CrossRef]

22. Mishra, Z.; Wang, Z.; Sadda, S.R.; Hu, Z. Using Ensemble OCT-Derived Features beyond Intensity Features for Enhanced Stargardt Atrophy Prediction with Deep Learning. *Appl. Sci.* **2023**, *13*, 8555. [CrossRef]
23. Wang, J.; Xu, C.; Yang, W.; Yu, L. A normalized Gaussian Wasserstein distance for tiny object detection. *arXiv* **2021**. [CrossRef]