applied sciences

Article

Rail Surface Defect Detection Based on An Improved YOLOv5s

Hui Luo, Lianming Cai * and Chenbiao Li

School of Information Engineering, East China Jiaotong University, Nanchang 330013, China; 1207@ecjtu.edu.cn (H.L.)
* Correspondence: 2020068085400024@ecjtu.edu.cn

Abstract: As the operational time of the railway increases, rail surfaces undergo irreversible defects. Once the defects occur, it is easy for them to develop rapidly, which seriously threatens the safe operation of trains. Therefore, the accurate and rapid detection of rail surface defects is very important. However, in the detection of rail surface defects, there are problems, such as low contrast between defects and the background, large scale differences, and insufficient training samples. Therefore, we propose a rail surface defect detection method based on an improved YOLOv5s in this paper. Firstly, the sample dataset of rail surface defect images was augmented with flip transformations, random cropping, and brightness transformations. Next, a Conv2D and Dilated Convolution(CDConv) module was designed to reduce the amount of network computation. In addition, the Swin Transformer was combined with the Backbone and Neck ends to improve the C3 module of the original network. Then, the global attention mechanism (GAM) was introduced into PANet to form a new prediction head, namely Swin transformer and GAM Prediction Head (SGPH). Finally, we used the Soft-SIoUNMS loss to replace the original CIoU loss, which accelerates the convergence speed of the algorithm and reduces regression errors. The experimental results show that the improved YOLOv5s detection algorithm reaches 96.9% in the average precision of rail surface defect detection, offering the accurate and rapid detection of rail surface defects, which has certain engineering application value.

Keywords: rail surface defect detection; YOLOv5s; data augmentation; CDConv; SGPH; Soft-SIoUNMS

check for updates

Citation: Luo, H.; Cai, L.; Li, C. Rail Surface Defect Detection Based on An Improved YOLOv5s. *Appl. Sci.* **2023**, *13*, 7330. https://doi.org/10.3390/app13127330

Academic Editor: Stéfano Frizzo Stefenon

Received: 5 June 2023
Revised: 18 June 2023
Accepted: 18 June 2023
Published: 20 June 2023

1. Introduction

As an important carrier of train operations, rail is also an important part of railroad system maintenance. As rail service time increases, the rail surface undergoes irreversible defects, and once these defects are produced, it is easy for them to develop quickly; thus, seriously threatening the safe operation of trains. If railway staff fail to identify defects and provide an early warning, it can lead to severe incidents, such as broken rails and other serious phenomena, which seriously threatens the safety of train operations, potentially leading to casualties and economic losses [1]. Until recently, the detection of rail surface defects was mainly dependent on manual detection by railway staff. Although manual detection has the advantages of being simple and low cost, it also has the disadvantages of low detection efficiency, high detection omission rate, and poor real-time performance [2]. Subsequently, non-destructive testing technologies were widely applied to railway systems, with common detection techniques including three-dimensional detection, eddy current detection, and acoustic wave detection [3]. Ye et al. [4] proposed a novel 3D perceptual system based on low-cost 2D laser sensors, using 3D reconstruction technology to model rail surface defect features in 3D through Matlab, and achieved good results in the laboratory and actual railroad track tests. Wang et al. [5] used magnetic flux leakage (MFL) technology to study the changes in magnetization and permeability of rail surface defects with dynamic magnetization time under high-speed motion, realizing damage detection at speeds of up to 200 km/h. Zhang et al. [6] achieved the rapid detection of rail surface defects by studying the application of laser ultrasound technology. Firstly, laser ultrasound signals were excited on both sides of the rail. Then, filtering, image alignment, and other algorithms were combined to obtain a rail surface defect image. In addition, a finite element model was established to simulate the propagation process of laser ultrasound signals in the rail, obtaining detection signals containing surface defect information to identify defect detection. However, the above methods are susceptible to interference from various external signals, which makes it difficult to process the collected signals effectively, resulting in missed detection, slow detection speed, and poor real-time performance.

In recent years, machine vision technology has been increasingly used for rail surface defect detection by virtue of its unique advantages, such as high detection accuracy, high speed, and non-contact operation [7]. Machine vision methods for rail surface defect detection can be divided into two approaches: traditional machine learning based detection methods and deep learning based detection methods. Traditional machine learning based detection methods are used to automatically detect rail surface defects using cameras. These methods require manually designed features or predefined features involving human analysis of rail surface defect images, and then proposing the corresponding feature learning algorithm for classification. Dubey et al. [8] and Yuan et al. [9] both manually extracted features for defect detection by analyzing the boundary of rail surface defects. Li et al. [10] proposed a real-time visual inspection system (VIS) for discrete surface defects, which to a certain extent solves the problems of uneven illumination and rail surface reflection characteristics. In addition, He et al. [11] used the inverse P-M (Perona-Malik) diffusion algorithm to detect rail surface defects, which improved the detection accuracy by differencing the original image from the diffusion image, highlighting the defects in the differential image, and removing the noise using a filtering algorithm based on edge features and the area of the defects. These methods have some effect on rail surface defect detection; however, their common disadvantage is that detection speed and detection accuracy are generally low. Some injuries, including linear injuries, cracks, and micro injuries, are difficult to detect and distinguish. Later, He et al. [12] proposed an algorithm for rail surface defect detection based on background difference, which improved the detection accuracy. Tian et al. [13] improved the Sobel algorithm for rail surface defect detection to compensate for the lack of sensitivity of the Sobel algorithm in the X and Y directions, which improved detection accuracy by 10%. Wang et al. [14] used the color characteristics of rail surface defects, which are very different from normal rust marks, within image blocks and handled different shapes of defect information very well. Liu et al. [15] performed gray value correction, binarization processing, and noise filtering on images to achieve fast and accurate detection of rail surface defects. While traditional machine learning detection methods have promoted the development of rail surface defect detection technology to a certain extent, they cannot extract the defect features sufficiently, with detection accuracy of small size targets being particularly low. In recent years, deep learning has developed rapidly, and convolutional neural networks (CNNs) have become the obvious choice for rail surface defect detection due to their unique feature representation advantages and modeling capabilities. Deep learning based detection methods are mainly divided into two groups: one-stage detection algorithms and two-stage detection algorithms.

Two-stage object detection algorithms are mainly region-based R-CNN series algorithms [16–18], which have high accuracy but slow speed. Faghih-Roohi et al. [19] proposed the use of Deep Convolutional Neural Network (DCNN) to classify images of rail surface areas with defects into six categories and designed three sizes of convolutional neural networks. The experimental results found that the convolutional neural network with greater depth is better at classifying images of rail surface defects, but the computation time is correspondingly longer. Gibert et al. [20] proposed a method to improve object detection performance by combining multiple detectors in a multi-task learning framework, which achieved high accuracy in the detection of railroad sleeper and fastener defects. Shang et al. [21] focused images on rail surface defects by cropping them, and then localized the defects using conventional methods, before feeding them into an improved CNN network for classification. Yu et al. [22] proposed a method to train a network using migration learning based on Faster R-CNN to achieve two types of rail surface defect detection. Jin et al. [23] proposed a modified version of a Gaussian mixture model for faster segmentation of rail surface defects. The robustness of the model was improved using Faster R-CNN for precise localization of the defects on the output of the segmentation model, resulting in more accurate detection results. Although the above two-stage object detection algorithms achieved higher levels of accuracy in rail surface defect detection, they both have the disadvantage of slow detection speed.

Single-stage object detection algorithms, as the name suggests, perform object detection stage in just one stage. The YOLO series [24–27] and SSD [28] are both representatives of single-stage object detection models. However, the SSD algorithm is poorer for small size target recognition, mainly because small size targets are mostly trained with lower level anchors, and the lower level features are not nonlinear enough to be trained with sufficient accuracy. Meanwhile, the YOLO series uses the idea of regression, which can directly complete the classification and localization of targets, and has the advantage of fast speed, so it is widely used in real-time detection systems. Song et al. [29] improved the loss function of the YOLOv3 algorithm for the characteristics of rail surface defects, and the experimental results showed that a 97% recognition rate of defects. Wei et al. [30] proposed various improved models based on the YOLOv3 network and lightweighted the models to obtain high detection accuracy. Zhao et al. [31] improved YOLOv5 for small size targets, enhancing the feature extraction capability of small size targets by adding a micro-scale detection layer and adjusting the confidence loss function of the detection layer to improve detection accuracy. Zhu et al. [32] proposed the TPH-YOLOv5 network with an additional layer of micro-detection combined with the CBAM attention mechanism, to replace the original prediction head with a transformer prediction head, to offer good dense target detection.

Due to the fast running speed of trains, the focus of our research is biased towards the real-time detection of rail surface defects, which very much requires the detection speed of the network; thus, we choose a single-stage target detection algorithm. YOLOv5 is currently one of the best performing models in the YOLO series and while it is constantly being updated, we choose to use version 6.0. In this paper, we propose a rail surface injury detection method based on an improved version of YOLOv5s, which was tested and compared with ablation experiments. The results show that the improved YOLOv5s detection algorithm achieves 96.7% of the average accuracy mean value mAP_0.5 for rail surface defect detection. The main contributions of this paper can be summarized as follows:

1. Due to insufficient rail surface defect image samples, the sample dataset was augmented with flip transformation, random cropping, and brightness transformation.

2. The CDConv module was designed to replace the original first layer convolution module of the backbone network of YOLOv5s. It reduces the number of network parameters and computations while improving detection speed and accuracy.

3. We improved the C3 module of the original network by combining the Swin Transformer at the tail end of the Backbone and the tail end of the Neck, providing better access to global information and richer contextual information.

4. The introduction of the GAM into the PANet makes the network pay more attention to injury-dense area features and forms a new prediction head, SGPH, which improves the detection accuracy of the network.

5. The original CIoU was replaced with the Soft-SIoUNMS loss function, which speeds up the convergence of the algorithm and reduces regression errors.

2. Related Works

2.1. YOLOv5s Network Model Structure

As an open source neural network, the YOLOv5 network achieves high performance in terms of detection speed and accuracy. With different network depths and feature map widths, the YOLOv5 network has four versions: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. We experimented with all four deep YOLOv5 networks and found that the YOLOv5s network performed the best. Furthermore, YOLOv5s is the smallest model. Since the release of YOLOv5s, the network has been updated to version 6.2, and our experiments showed further improvement based on the model using version 6.0. In particular, the model of YOLOv5s mainly consists of an input (Input), backbone network (Backbone), neck (Neck), and output (Output). This network model is shown in Figure 1.

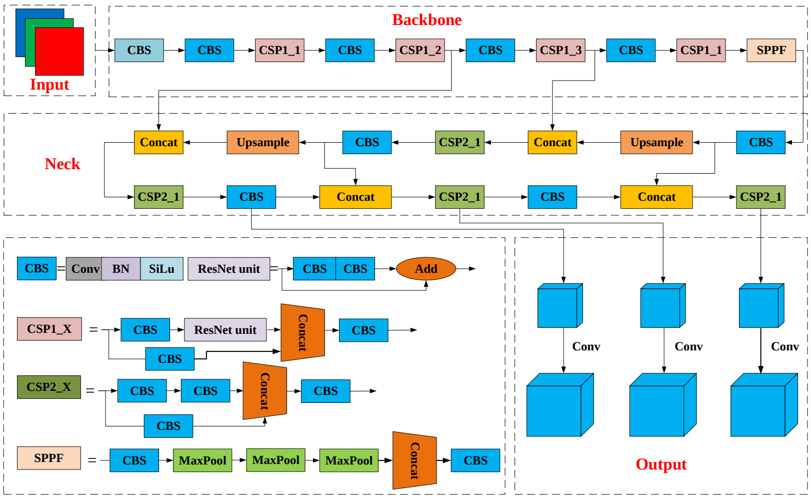


Figure 1. Structure diagram of YOLOv5s.

2.2. Data Enhancement

In this paper, the performance of the proposed method was verified using the publicly available Rail Surface Defect Dataset (RSDD). The dataset was collected and made public by Beijing Jiaotong University and consists of 195 challenging rail surface defect images, of which 67 are 160 pixels × 1000 pixels in size and 128 are 55 pixels × 1250 pixels in size. In the experiment, 347 rail surface defect images with a size of 160 pixels × 250 pixels and at least one defect were obtained after segmentation and image resizing, and an initial rail surface defect image sample dataset was established. Then, based on the initial established dataset, the dataset was expanded using flip transformation, random cropping, and brightness transformation. Finally, 3086 rail surface defect images were generated, of which we randomly selected 70% as the training set, 10% as the validation set, and 20% as the test set. The initial rail surface defect images used in this paper are shown in Figure 2, and the images generated using data enhancement methods are shown in Figure 3.

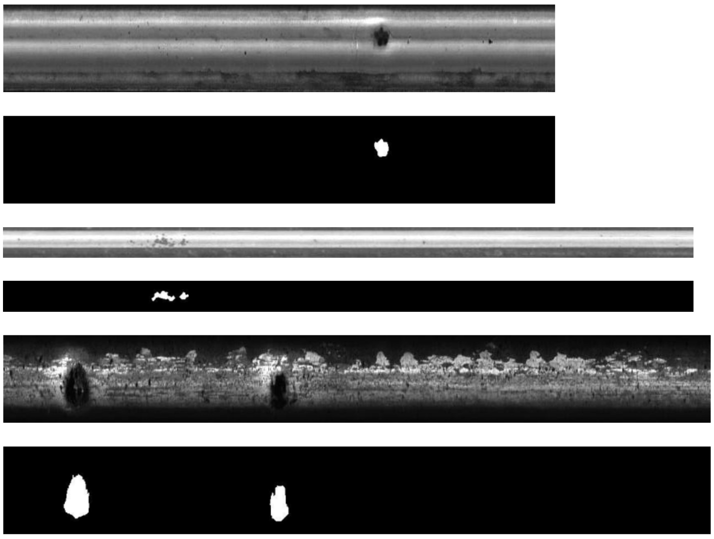


Figure 2. Initial dataset of rail surface defect images.

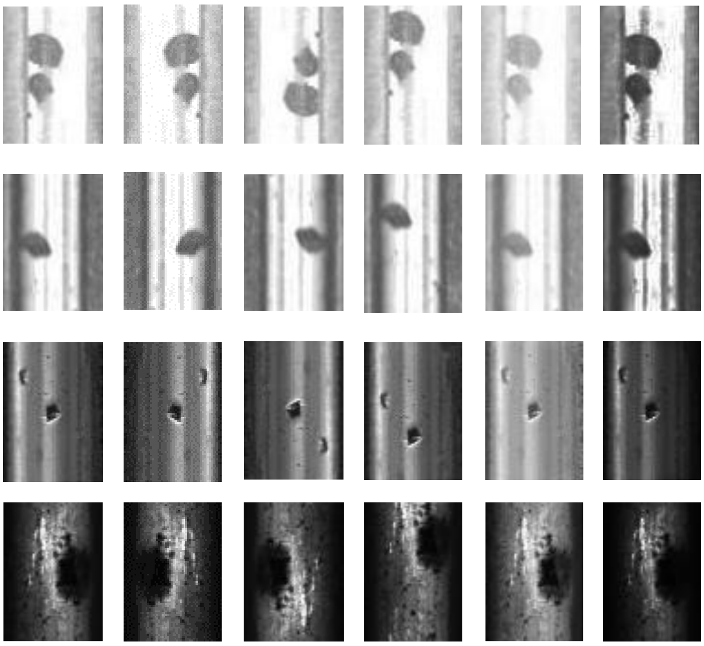


Figure 3. Examples of rail surface defect image data enhancement.

3. Methodology

Rail surface defects are often small targets and the original YOLOv5s model cannot fully meet the detection needs, leading to low detection accuracy and leakage detection. Therefore, this paper proposes to improve the YOLOv5s network model in several ways. Firstly, the backbone network is modified and the CDConv module replaces the original first layer convolution module of the Backbone network. Secondly, the original convolution prediction head is replaced with the Swin Transformer attention prediction head, and the GAM is introduced into the PANet to make the network focus on injury-dense region features. Finally, the original CIoU is replaced with the Soft-SIoUNMS loss function. Our improved network structure is shown in Figure 4.

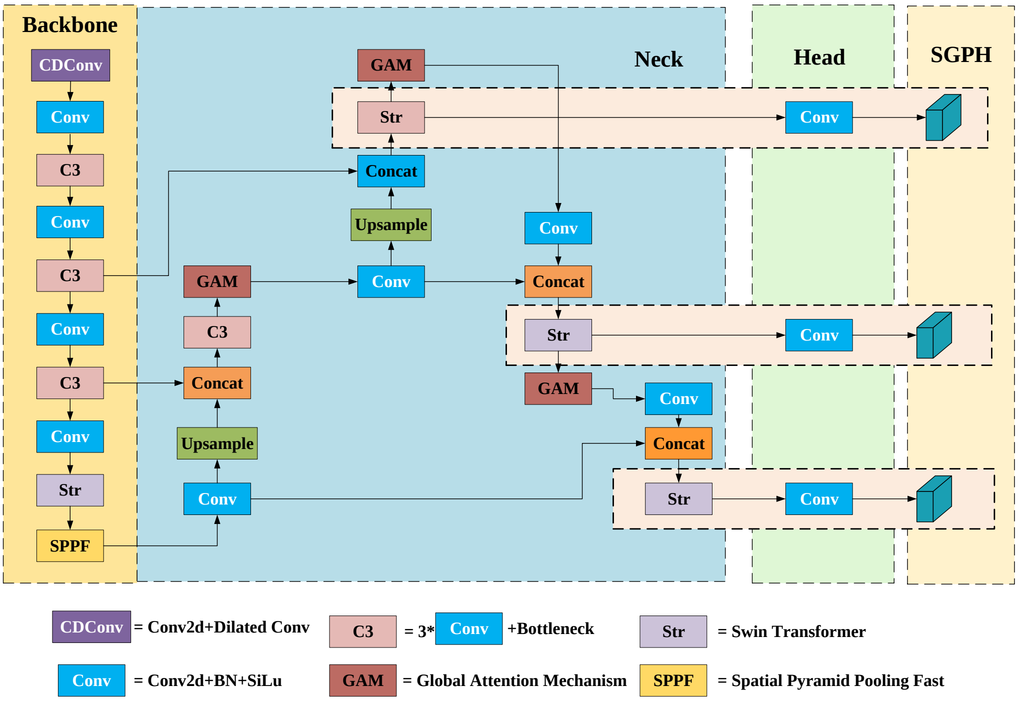


Figure 4. Diagram of the improved network structure.

3.1. CDConv

The traditional convolutional module of YOLOv5s performs Conv2D convolution, BN, and SiLU activation function operations on the input feature map, which greatly affect the speed. The CDConv designed by us changes the single channel into two channels: one channel represents the original Conv2D convolution, while the other channel represents the dilated convolution. The original 6 × 6 convolution is replaced with 3 × 3 convolution with the same step size; the padding of the Conv2D convolution is removed; and the padding of the dilated convolution is set to 2, so that the size of the output image remains the same as that of the Conv2D convolution. Finally, the two sides perform an Add operation to output the image feature map. The CDConv convolution module is shown in Figure 5.

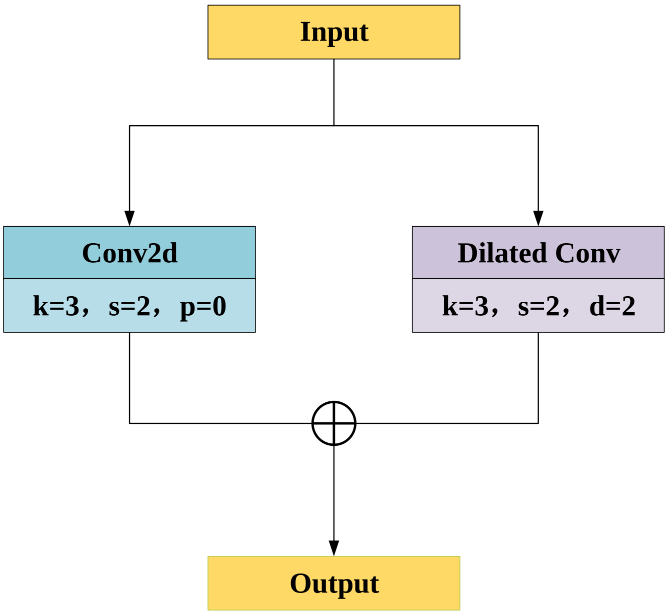


Figure 5. Structure diagram of CDConv.

3.2. C3STR Module

Due to the problem of small target sizes in the dataset, the detection of small object location information in the object detection layer of the downsampling of the original YOLOv5s network is limited. This makes the detected target not obvious and the detection accuracy is low. Therefore, background features should be reduced and target features should be highlighted in the detection process; using self-attention is a good method to achieve this. The transformer [33] is a method used in the field of natural language processing, and the Swin Transformer [34] is a modified network of the transformer. In this paper, we integrate the Swin Transformer module into the improved YOLOv5s network by combining the C3 module of the original network into the Backbone tail and the Neck tail. This is performed in order to improve the localization of rail surface defects by exploiting both the high-resolution spatial information of the CNN features and the global semantic information encoded by the Swin Transformer. The Swin Transformer Block (STB) consists of two subunits in pairs: the window-based multi-head self-attention (W-MSA) module and the sliding window-based multi-head self-attention (SW-MSA) module. Each subunit is composed of a layer normalization (LN), an attention module, a normalization layer, and a multi-layer perceptron (MLP). The computational procedure of the Swin Transformer Block is shown in Figure 6.

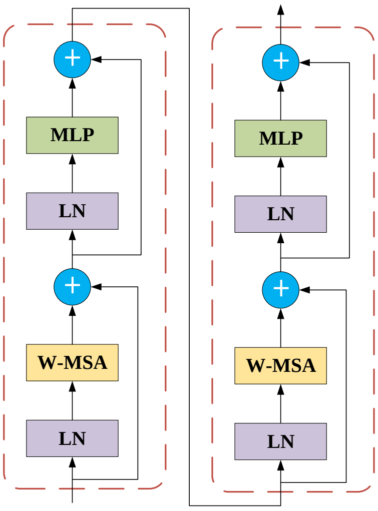


Figure 6. Structure diagram of the Swin Transformer Block.

3.3. SGPH

In this paper, we add the global attention mechanism (GAM) [35] to PANet to form a new prediction head called SGPH, based on the C3STR module. The GAM reduces the network information and amplifies the global dimensional interaction features. This attention mechanism is based on the sequential channel-space attention mechanism in convolutional block attention module (CBAM) [36] with optimized improvement of its submodules. CBAM uses sequential embedding of channel attention and space attention to improve the accuracy of the model, but does not consider the interaction between the two dimensions, which may lead to the separation of feature vectors and, thus, the loss of cross-dimensional information. To solve this problem, the GAM uses three-dimensional alignment to extract features in all three dimensions in the channel attention submodule, as well as grouped convolution with channel mixing wash to reduce the number of parameters in the spatial attention submodule. Pooling operations are removed from both submodules of the GAM, and this approach can effectively extract the interactions between feature vectors to improve the detection performance of the model. This procedure is shown in Equations (1) and (2). The structure of the GAM is shown in Figure 7.

$$F_2 = M_c(F_1) \otimes F_1 \tag{1}$$

$$F_3 = M_s(F_2) \otimes F_2 \tag{2}$$

where F_1 represents the input feature, F_2 represents the intermediate state, F_3 represents the output feature, M_c is the channel attention module, and M_s is the spatial attention module.

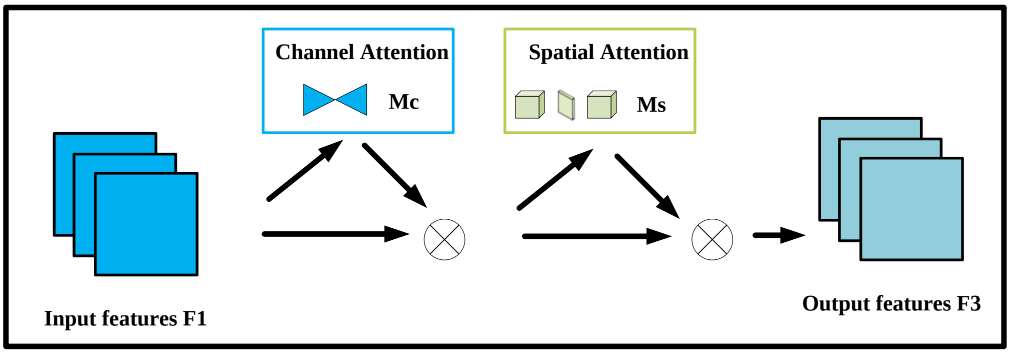


Figure 7. The structure of the GAM.

3.4. Soft-SIoUNMS

A loss function is a method to measure the accuracy of model prediction results. The original YOLOv5s network uses CIoU Loss as a loss function of the bounding box. Many IoU loss functions, such as GIoU, DIoU, and CIoU, do not take into account the direction between the real frame and the predicted frame. The predicted frame may have a large position deviation in the training process and eventually produce a worse model, resulting in slow convergence speed. To solve this problem, SIoU Loss [37] was used in this paper to replace CIoU Loss. The SIoU redefines the correlation loss function by introducing the vector angle between the real box and the prediction box. Considering that rail surface defects are characterized by a small target scale and high density, the YOLOv5s network may cause missed detection, thus reducing the recall rate of the model. In this paper, Soft-NMS [38] was combined to further improve the YOLOv5s network. Furthermore, Soft-SIoUNMS was combined with SIoU Loss and Soft-NMS, used as the loss function of border regression. The SIoU calculation method is shown in Equation (3) and the Soft-NMS calculation method is shown in Equation (4).

$$\begin{cases} SIoU = 1 - IoU + \frac{\Delta+\Omega}{2} \\ \Omega = \sum\limits_{t=w,h} \left(1 - e^{-w_i}\right)^{\theta} \\ \Delta = \sum\limits_{t=x,y} \left(1 - e^{-\gamma\rho_t}\right) \end{cases} \tag{3}$$

$$S_i = S_i e^{-\frac{iou(M,b_i)^2}{\sigma}}, \forall b_i \notin D \tag{4}$$

where Ω represents the shape cost, Δ represents the redefined distance cost after considering the angle cost, b_i represents the i-th anchor box, and S_i represents the score of the i-th anchor box.

4. Experiments

4.1. Experimental Environment

The experimental environment in this paper was built on a server workstation with the following hardware configuration: 3.5 GHz Inter i7-7800XCPU, NVIDIA GeForce RTX 2080 graphics card, and 32 Gb memory. The server configuration uses Ubuntu version 16.04, CUDNN version 7.6, and CUDA Toolkit version 10.1. The deep learning framework platform is Pytorch version 1.8.

4.2. Evaluation Indicators

In order to verify the effectiveness of the model, we used three indicators: precision, recall, and mean average precision (mAP) for comprehensive evaluation. Their calculation is shown in Equations (5)–(7). Precision is used to measure the ability to predict positive samples and predict them correctly, recall is used to measure the ability to detect all positive samples, and mean average precision is calculated by averaging the average precision and is used to evaluate the overall detection performance of the model.

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

$$mAP = \frac{\sum_{c-1}^{c} AP}{C} \tag{7}$$

4.3. Analysis of Experimental Results

4.3.1. Ablation Experiments

In this paper, ablation experiments of the Backbone, Head, and loss function were carried out in turn. The experimental results are shown in Table 1.

Table 1. Comparison of results of ablation experiments.

Model Name	P	R	mAP@0.5
YOLOv5s	0.965	0.933	0.954
YOLOv5s + CDConv (Method1)	0.968	0.936	0.961
YOLOv5s + CDConv + C3STR (Method2)	0.951	0.936	0.963
YOLOv5s + CDConv + C3STR + CBAM (Method3)	0.963	0.938	0.964
YOLOv5s + CDConv + C3STR + GAM (Method4)	0.967	0.940	0.965
YOLOv5s + CDConv + C3STR + GAM + Soft-SIoUNMS (Ours)	0.971	0.945	0.969

The first row of Table 1 shows the unimproved YOLOv5s model, where the mean average precision of rail surface defects is 95.4%. Then, the trunk and header were improved, respectively. After the addition of the CDConv module and C3STR in the second and third lines, the mAP was significantly improved, reaching 96.1% and 96.3%, respectively. The reason is that the CDConv module greatly reduces the calculation amount of network parameters, and the model detection speed is improved; however, some information may be lost. The Swin Transformer combined with the C3 module completed the global information and richer context information. Then, the attention mechanisms, CBAM and GAM, were added to the Head for comparison. It was found that the network model with the GAM achieved better detection accuracy. Although both CBAM and GAM are mixed attention mechanisms, CBAM does not consider the interaction between the two dimensions, which may lead to the separation of feature vectors and the loss of inter-dimensional information. Therefore, we chose the GAM for its superior performance. Finally, the Soft-SIoUNMS loss function, which is a combination of SIoU and Soft-NMS, was added, providing a better effect than a single loss function. The experimental results showed that the mAP of our proposed method reached 96.9%, achieving the best detection results. The detection diagram in Figure 8 corresponds to the different methods used in the ablation experiments in Table 1. In order to better demonstrate the benefits of the improved model on the rail surface defect dataset, each improvement point will be validated separately.

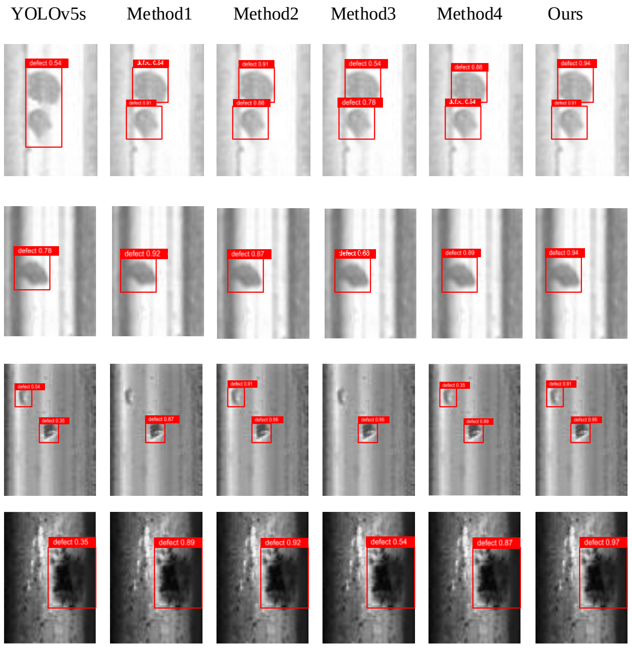


Figure 8. The different methods used in ablation experiments.

4.3.2. Improved Model Comparison Experiments

In order to prove the superiority of our proposed method in rail surface defect detection, we will verify each improvement point in turn, as follows:

- The number of C3STRs

Table 2 represents the influence of successively adding different amounts of C3STRs to the network model on the experimental results.

Table 2. The influence of the number of C3STRs on the model.

Model	Number of C3STRs	P	R	mAP@0.5
YOLOv5s + CDConv + C3STR	1	0.943	0.933	0.950
YOLOv5s + CDConv + C3STR	2	0.944	0.935	0.948
YOLOv5s + CDConv + C3STR	3	0.951	0.936	0.951
YOLOv5s + CDConv + C3STR	4	0.951	0.936	0.963
YOLOv5s + CDConv + C3STR	5	0.949	0.934	0.953

Due to the excellent performance of the CDConv, we directly improved the network by adding the CDConv and sequentially comparing the detection results of different numbers of C3STRs added to the network. We sequentially tested adding 1–5 C3STRs to the Backbone and Head, and the experimental results showed that although each addition of a C3STR module would increase the computational load of the network in turn, the detection accuracy gradually increased, reaching the highest detection accuracy with the addition of four modules. At the fifth point, the detection accuracy began to decrease, and the network model was already very large. Taking this into account, we chose to add four C3STR improved models in this paper.

- Attention mechanism

Table 3 shows the influence of adding CBAM, ECA, CA, SE, and GAM to the network model on the experimental results.

Table 3. Comparison of the results of different attention mechanisms.

Model	P	R	mAP@0.5
YOLOv5s + CDConv + C3STRs + CBAM	0.963	0.938	0.964
YOLOv5s + CDConv + C3STRs + ECA	0.964	0.935	0.961
YOLOv5s + CDConv + C3STRs + CA	0.951	0.936	0.963
YOLOv5s + CDConv + C3STRs + SE	0.962	0.938	0.960
YOLOv5s + CDConv + C3STRs + GAM	0.967	0.940	0.965

After confirming the optimal number of C3STRs, this paper further improved the model by adding different attention mechanisms to the Head to make the network more focused on the injury concentration area. In Table 3, the experimental comparison of multiple attention mechanisms in this paper shows that mixed attention achieves the highest detection accuracy than single-channel attention mechanisms, among which GAM has a higher comprehensive performance than CBAM.

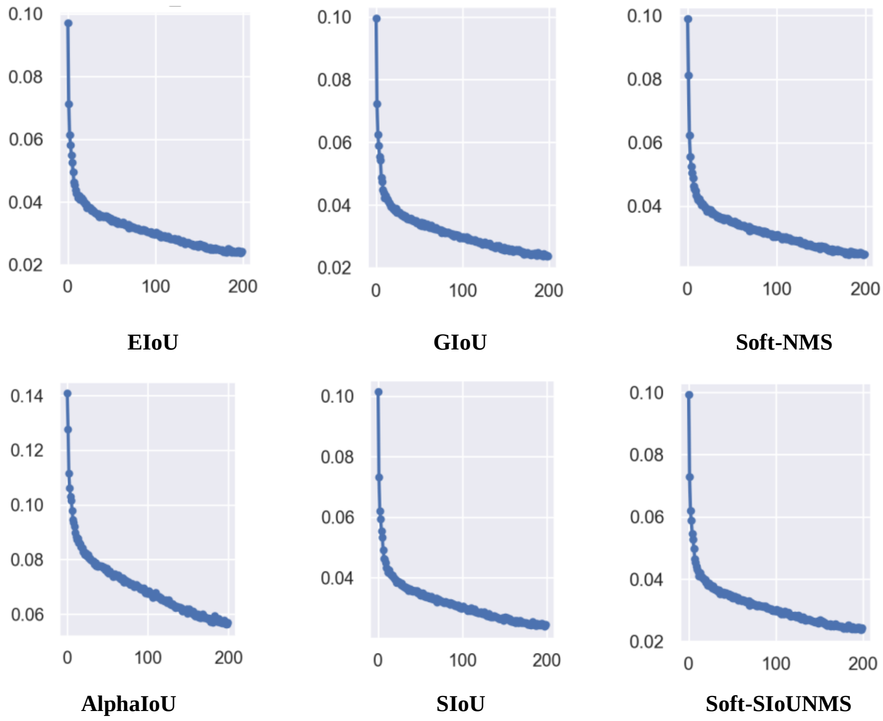
- Loss function

Table 4 shows the effects of adding different loss functions to the network model on the experimental results.

This paper compares several mainstream loss functions, including the latest AlphaIoU and SIoU, and finds that the combination of SIoU and Soft-NMS to make the function, Soft-SIoUNMS, achieves the best detection effect through experiments. The comparison of defect detection loss values of different loss functions is shown in Figure 9. In order to more intuitively reflect the difference between the improved loss and the original loss function, we compare them. Figure 10 is a comparison of the original CIoU and the improved Soft-SIoUNMS.

Table 4. Comparison of the results of different loss functions.

Model	P	R	mAP@0.5
YOLOv5s + CDConv + C3STRs + EIoU	0.965	0.933	0.954
YOLOv5s + CDConv + C3STRs + GIoU	0.968	0.936	0.961
YOLOv5s + CDConv + C3STRs + AlphaIoU	0.951	0.936	0.963
YOLOv5s + CDConv + C3STRs + SIOU	0.963	0.938	0.964
YOLOv5s + CDConv + C3STRs + Soft-NMS	0.965	0.938	0.962
YOLOv5s + CDConv + C3STRs + Soft-SIoUNMS	0.972	0.950	0.965



EIoU GIoU Soft-NMS

AlphaIoU SIoU Soft-SIoUNMS

Figure 9. Comparison of defect detection loss values of different loss functions.

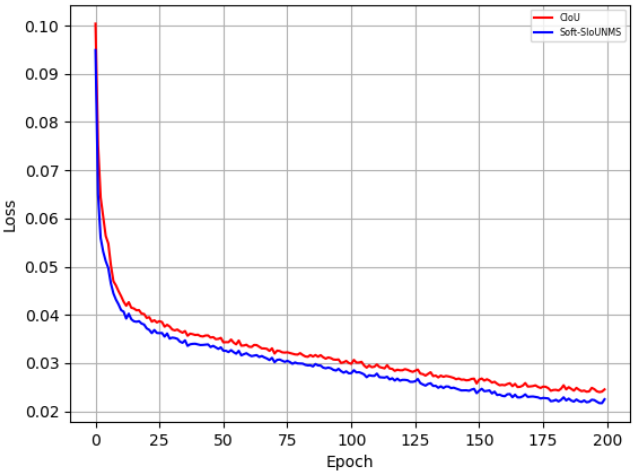


Figure 10. Comparison of the original CIoU and the improved Soft-SIoUNMS.

4.3.3. Performance Comparison Experiments

In order to fully evaluate the improved YOLOv5s detection algorithm in this paper, six algorithms were selected for experimental comparison, including the unimproved YOLOv5s, the earlier YOLOv4, and the updated YOLOX. Meanwhile, Faster-RCNN with the highest two-stage accuracy, and a lightweight network, EfficientDet, were compared. The mean average precision (mAP) was selected as the evaluation index of different detection algorithms, which can scientifically and reasonably evaluate the target detection and positioning ability of different detection algorithms. The experimental comparison results are shown in Table 5.

Table 5. The results of comparison experiments.

Model Name	P	R	mAP@0.5	Times/ms
YOLOv4	92.05	78.74	90.37	38.4
YOLOv5s	96.55	93.34	95.43	16.3
YOLOX	96.95	91.62	95.84	22.7
Faster-RCNN	94.39	74.61	94.43	160.3
EfficientDet	96.49	71.20	90.81	54.1
Ours	97.13	94.53	96.90	29.2

In this paper, the performance of the proposed algorithm and some other classical algorithms for rail surface defect detection were compared experimentally. The experimental results are shown in Table 5, where it can be seen that the improved YOLOv5s algorithm proposed in this paper achieved the highest detection accuracy. Compared with YOLOv4 [27], the unimproved YOLOv5s algorithm, YOLOX [39], Faster R-CNN [18], and EfficientDet, the improved YOLOv5S algorithm achieved the highest detection accuracy, with mAP increased by 6.53, 1.47, 1.06, 2.47, and 6.09 percentage points, respectively.

The prediction results in Figure 11 show the detection effects of YOLOv4, the unimproved YOLOv5s, YOLOX, Faster R-CNN, EfficientDet, and the improved YOLOv5s of this paper. It can be seen from the figure that YOLOv5s and Faster R-CNN both have missed detections, while YOLOv4 has inaccurate positioning. The improved YOLOv5s algorithm proposed in this paper detects all the defects in the obtained images, and compared with the unimproved YOLOv5s algorithm, the accuracy of the predicted border is higher. Thus, a better detection effect is achieved.

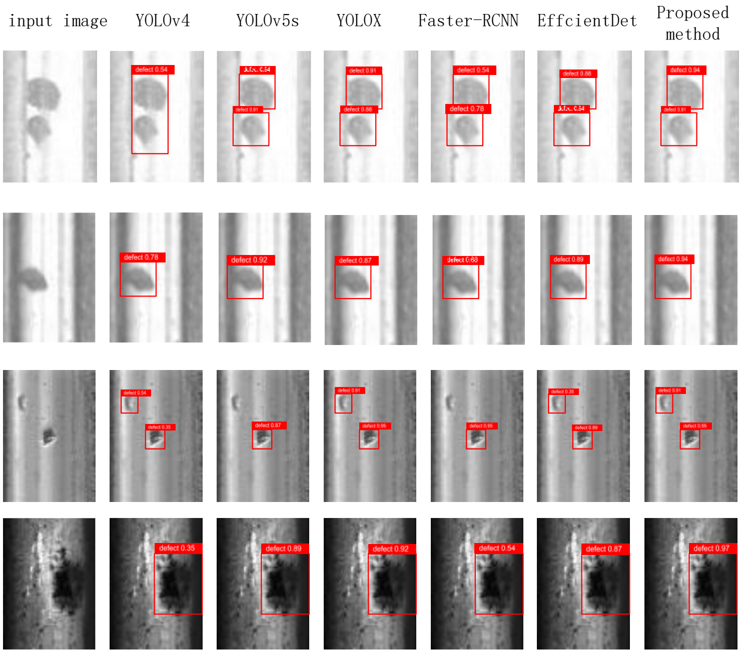


Figure 11. Comparison of the Detection Effect of Each Algorithm.

5. Conclusions

In this paper, the problems of low contrast between defects and the background, the large scale difference, and insufficient training samples in the detection of rail surface defects are studied. To address these problems, the YOLOv5s method was improved and the following research was carried out. Firstly, in view of the problem of low contrast between defects and the background as well as insufficient training samples, the rail surface defect image dataset was augmented with flip transformation, random cropping, brightness transformation, and a generative adversarial network. Then, the CDConv was added into the backbone network to reduce the number of network parameters and the number of calculations, while improving detection speed and accuracy. Targeting the problem of large scale differences in defects, the Swin Transformer module was combined with the Backbone and Neck end to improve the C3 module of the original network. Furthermore, the GAM was introduced in PANet to form a new detection head, SGPH, to expand the global dimension interaction across the spatial channel dimension. It enables the network to obtain better global information and rich context information, while improving the ability of the model to capture different dimensions of information. Finally, the Soft-SIoUNMS loss function replaced the original CIoU loss function, which accelerated the model convergence speed and achieved different levels of bounding box regression accuracy. The experimental results show that the mAP of our improved YOLOv5s rail surface defect detection method reached 96.9% with a detection time of 29.2 ms. This meets the requirements for accurate, real-time rail surface defect detection. Future research will study rail surface defect detection in complex environments and improve on the lightweight nature of the network model while maintaining detection accuracy.

Author Contributions: Conceptualization, L.C. and C.L.; formal analysis, L.C. and H.L.; investigation, L.C. and H.L.; methodology, L.C.; project administration, H.L.; validation, C.L.; writing—original draft, L.C. and H.L.; writing—review and editing, L.C. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the National Natural Science Foundation of China (Grant No. 62262021).

Data Availability Statement: The sample image data of rail surface defects used in this study are available from the website: http://icn.bjtu.edu.cn/Visint/resources/RSDDs.aspx, accessed on 1 December 2017.

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

References

1. Jessop, C.; Ahlström, J.; Hammar, L.; Fæster, S.; Danielsen, H.K. 3D characterization of rolling contact fatigue crack networks. *Wear* **2016**, *15*, 392–400. [CrossRef]
2. Papaelias, M.P.; Lugg, M. Detection and evaluation of rail surface defects using alternating current field measurement techniques. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2012**, *226*, 530–541. [CrossRef]
3. Jiang, Y.; Wang, H.; Tian, G.; Chen, S.; Zhao, J.; Liu, Q.; Hu, P. Non-contact ultrasonic detection of rail surface defects in different depths. In Proceedings of the 2018 IEEE Far East NDT New Technology & Application Forum (FENDT), Xiamen, China, 6–8 July 2018; pp. 46–49.
4. Ye, J.; Edward, S.; Clive, R. Use of a 3D model to improve the performance of laser-based railway track inspection. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* **2019**, *233*, 337–355. [CrossRef]
5. Wang, P.; Gao, Y.; Tian, G.; Wang, H. Velocity effect analysis of dynamic magnetization in high speed magnetic flux leakage inspection. *NDT E Int.* **2014**, *6*, 7–12. [CrossRef]
6. Zhang, Y.; Luo, L.; Zhang, Y.; Gao, X.; Long, J. Interlaced scanning by laser ultrasonic for defects imaging of train rail surface. In Proceedings of the Eleventh International Conference on Information Optics and Photonics (CIOP 2019), Xi'an, China, 6–9 August 2019; SPIE: Bellingham, WA, USA, 2019; pp. 189–199.
7. Kou, L. A review of research on detection and evaluation of the rail surface defects. *Acta Polytech. Hung.* **2022**, *19*, 167–186. [CrossRef]
8. Dubey, A.; Jaffery, Z. Maximally Stable Extremal Region Marking-Based Railway Track Surface Defect Sensing. *IEEE Sens. J.* **2016**, *16*, 9047–9052. [CrossRef]

9. Yuan, X.C.; Wu, L.S.; Chen, H.W. Rail image segmentation based on Otsu threshold method. *Opt. Precis. Eng.* **2016**, *24*, 1772–1781. [CrossRef]
10. Li, Q.; Ren, S. A Real-Time Visual Inspection System for Discrete Surface Defects of Rail Heads. *IEEE Trans. Instrum. Meas.* **2012**, *61*, 2189–2199. [CrossRef]
11. He, Z.; Wang, Y.N.; Mao, J.; Yin, F. Research on inverse P-M diffusion-based rail surface defect detection. *Acta Autom. Sin.* **2014**, *40*, 1667–1679.
12. He, Z.; Wang, Y.; Liu, J.; Yin, F. Background differencing-based high-speed rail surface defect image segmentation. *Chin. J. Sci. Instrum.* **2016**, *37*, 640–649.
13. Tian, S.; Kong, J.Y.; Wang, X.D. Improved Sobel algorithm for defect detection of rail surfaces with enhanced efficiency and accuracy. *J. Central South Univ.* **2016**, *23*, 2867–2875.
14. Wang, H.; Wang, M.; Zhang, H. Vision saliency detection of rail surface defects based on PCA model and color features. *Process. Autom. Instrum.* **2017**, *38*, 73–76.
15. Liu, Q.Q.; Zhou, H.Y.; Wang, X.Z. Research on Rail Surface Defect Detection Method Based on Gray Equalization Model Combined with Gabor Filter. *Surf. Technol.* **2018**, *19*, 745–755.
16. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
17. Girshick, R. Fast R-CNN. *Comput. Sci.* **2015**, *3*, 1440–1448.
18. Ren, S.; He, K.; 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]
19. Faghih-Roohi, S.; Hajizadeh, S.; Nunez, A.; Babuska, R.; De Schutter, B. Deep Convolutional Neural Networks for Detection of Rail Surface Defects. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 2584–2589.
20. Gibert, X.; Patel, V.M.; Chellappa, R. Deep Multitask Learning for Railway Track Inspection. *IEEE Trans. Intell. Transp. Syst.* **2016**, *18*, 153–164. [CrossRef]
21. Shang, L.; Yang, Q.; Wang, J.; Li, S.; Lei, W. Detection of rail surface defects based on CNN image recognition and classification. In Proceedings of the 2018 20th International Conference on Advanced Communication Technology (ICACT), Chuncheon, Republic of Korea, 11–14 February 2018; pp. 45–51.
22. Yu, C.; Deng, H.G.; Feng, Y.X. Effects of Faster Region-based Convolutional Neural Network on the Detection Efficiency of Rail Defects under Machine Vision. In Proceedings of the 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 12–14 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1377–1380.
23. Jin, X.; Wang, Y.; Zhang, H.; Zhong, H.; Liu, L.; Wu, Q.J.; Yang, Y. DM-RIS: Deep Multi model Rail Inspection System with Improved MRF-GMM and CNN. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 1051–1065. [CrossRef]
24. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; Volume 8, pp. 779–788.
25. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; Volume 3, pp. 6517–6525.
26. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 2767–2773.
27. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
28. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. *SSD: Single Shot Multi Box Detector. European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; Volume 12, pp. 21–37.
29. Song, Y.N.; Zhang, G.H.; Liu, L.; Zhang, H. Rail surface defect detection method based on YOLOv3 deep learning networks. In Proceedings of the CAC 2018: Proceedings of the 2018 Chinese Automation Congress, Xi'an, China, 30 November–2 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1563–1568.
30. Wei, X.; Wei, D.; Suo, D.; Jia, L.; Li, Y. Multi-Target Defect Identification for Railway Track Line Based on Image Processing and Improved YOLOv3 Model. *IEEE Access* **2020**, *8*, 61973–61988. [CrossRef]
31. Zhao, J.; Zhang, X.; Yan, J.; Qiu, X.; Yao, X.; Tian, Y.; Zhu, Y.; Cao, W. A Wheat Spike Detection Method in UAV Images Based on Improved YOLOv5. *Remote Sens.* **2021**, *13*, 3095. [CrossRef]
32. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 October 2021; pp. 2778–2788.
33. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
34. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical vision Transformer using shifted windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; pp. 9992–10002.

35. Liu, Y.; Shao, Z.; Hoffmann, N. Global Attention Mechanism: Retain Information to Enhance Channel Spatial Interactions. *Comput. Sci.* **2021**, *26*, 1–5.
36. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
37. Gevorgyan, Z. SIoU loss: More powerful learning for bounding box regression. *arXiv* **2022**, arXiv:2205.12740.
38. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS—Improving Object Detection With One Line of Code. In Proceedings of the 2017 IEEE International Conference on Computer Vision ICCV, Venice, Italy, 22–29 October 2017; Volume 5, pp. 5562–5570.
39. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *Comput. Sci.* **2021**, *13*, 1–7.

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.