



Article Rail Surface Defect Detection Based on Image Enhancement and Improved YOLOX

Chunguang Zhang ^{1,2}, Donglin Xu¹, Lifang Zhang ^{1,*} and Wu Deng ^{2,3}

- ¹ School of Electronics and Information Engineering, Dalian Jiaotong University, Dalian 116028, China
- ² Traction Power State Key Laboratory, Southwest Jiaotong University, Chengdu 610031, China;
 - wdeng@cauc.edu.cn
- ³ School of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China
- * Correspondence: zlf214@163.com

Abstract: During the long and high-intensity railway use, all kinds of defects emerge, which often produce light to moderate damage on the surface, which adversely affects the stable operation of trains and even endangers the safety of travel. Currently, models for detecting rail surface defects are ineffective, and self-collected rail surface images have poor illumination and insufficient defect data. In light of the aforementioned problems, this article suggests an improved YOLOX and image enhancement method for detecting rail surface defects. First, a fusion image enhancement algorithm is used in the HSV space to process the surface image of the steel rail, highlighting defects and enhancing background contrast. Then, this paper uses a more efficient and faster BiFPN for feature fusion in the neck structure of YOLOX. In addition, it introduces the NAM attention mechanism to increase image feature expression capability. The experimental results show that the detection of rail surface defects using the algorithm improves the mAP of the YOLOX network by 2.42%. The computational volume of the improved network increases, but the detection speed can still reach 71.33 fps. In conclusion, the upgraded YOLOX model can detect rail surface flaws with accuracy and speed, fulfilling the demands of real-time detection. The lightweight deployment of rail surface defect detection terminals also has some benefits.

Keywords: image processing; rail surface defects; image enhancement; YOLOX

1. Introduction

In recent years, after continuous large-scale development, China's rail transportation industry has made remarkable achievements and stepped into a new development period. In the process of train operation, seamless steel rails play a guiding and supporting role, and pressure is continuously generated between wheels and rails; at the same time, high-strength and high-density fatigue wear will lead to different degrees of surface defects on the rails [1–4]. This will affect the smoothness of the track, the stability and the comfort of train operation, and become an important factor that restricts the transportation efficiency of rail trains and affects the safety of train operation [5–8]. The traditional rail surface defect detection method mainly comprises manual inspection; this way is subjective, inefficient and costly, and there are safety risks, which makes it difficult to meet operation and development needs [9–11].

To solve the above problems, the nondestructive testing and evaluation of rail surface defects is receiving increasing attention [12–14]. Common NDT techniques include ultrasonic inspection [15,16], electromagnetic inspection [17,18], and machine vision inspection [19–21]. Machine vision inspection is faster, more efficient, and completely non-invasive compared to other inspection techniques, making it more suited to the job of locating flaws on the rail surface. Yu et al. [18] suggested a coarse-to-fine model (CTFM) to discover flaws at several sizes, including sub-image level, region level, and pixel level.



Citation: Zhang, C.; Xu, D.; Zhang, L.; Deng, W. Rail Surface Defect Detection Based on Image Enhancement and Improved YOLOX. *Electronics* **2023**, *12*, 2672. https:// doi.org/10.3390/electronics121222672

Academic Editor: Dung-An Wang

Received: 29 May 2023 Revised: 13 June 2023 Accepted: 13 June 2023 Published: 14 June 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/). Zhang et al. [19] put out a novel line-level labeling-based finite-sample RSD detection technique based on line-level labeling, which classifies pixel lines by using defect images as sequence data. Then, to detect rapid track defects and common heavy track flaws, respectively, two detection methods, OC-IAN and OC-TD, were developed. Wang et al. [20] designed a new feature pyramid for multiscale fusion in the mask R-CNN network and used complete intersection over union (CIOU) to overcome the limitation of intersection over union (IOU) in some special cases, and the mean average precision (mAP) of rail surface defect detection reached 98.70%. Hu et al. [21] added the CA attention mechanism and adaptive spatial feature fusion (ASFF) to the YOLOX-Nano network, and the modified network mAP was improved by 18.75%. Feng et al. [22] organically combined MobileNet with the YOLOv3 network and combined it with experiments to demonstrate that the M2-Y3 network can achieve relatively optimal detection results. Zhang et al. [23] self-captured rail images and segmented rail surface defects using a multi-contextual information segmentation network (MCnet) with good performance. Jin et al. [24] proposed a DM-RIS model for segmentation of rail surface defect edges and also trained an improved faster RCNN to remove non-defects, which showed a high robustness. Li et al. [25] designed a network called WearNet that detects metal surface scratches through image classification. The network is quite light, and in the experiments, the network classification accuracy reached 94.16%. In addition, some other methods were also proposed in recent years [26–35].

According to the literature, real-time issues must be prioritized in realistic surface defect detection tasks [36–43]. The YOLO series [44–47] represents the one-stage target identification technique, which is faster than the two-stage approach but less accurate. The two-stage target detection algorithm, represented by the faster R-CNN [48] and cascade R-CNN [49], on the other hand, typically has worse algorithmic accuracy but greater speed. At the same time, due to the relatively complex environment of rail image acquisition and poor illumination, the direct implementation of the above detection methods to their application is not ideal.

This work initially suggests a fusion image enhancement technique to highlight the details of rail surface defects to address the issues raised above and accomplish effective rail surface defect identification. Second, based on YOLOX [50,51], BiFPN-Tiny is used as an enhanced feature extraction network to lower the training cost of the network, add NAM attention to the three effective feature layers of the backbone network, and further improve the detection accuracy of rail surface defects [52,53]. In the end, the enhanced model is evaluated and contrasted with conventional target detection networks, and is experimentally confirmed to satisfy the real-world requirements of rail surface defects.

2. Materials and Methods

2.1. Data Gathering and Preparation

2.1.1. Dataset Construction

Depending on the classification standard of injury and damage, and the study of the formation mechanism of rail surface defects, common defects on rails are as follows: (a) Dent: oval-shaped edge, only locally present; (b) Crush: obvious transverse depression; (c) Scratch: longitudinal slight long strip damage; (d) Slant: tear in the side of the rail surface; (e) Damage: parent material from rail surface displacement; (f) Unknown: cannot clearly determine whether it is damage, needs manual confirmation; (g) Dirt: paint or dirt covering the rail surface; (h) Gap: weld or obvious break between adjacent rails. The above eight types of defects are shown in Figure 1 below.

In this study, Dent, Crush, Scratch, Slant, Damage, and Unknown are collectively referred to as Defect. Based on this, the dataset in this study has three categories of Defect, Dirt, and Gap. The image data in this study were mainly obtained from self-collected track images, the RSDDs Dataset, and Dutch heavy rail track images provided by ProRail. After removing the low-quality images, there are 200 valid Defect images [54–56].



Figure 1. Examples of rail surface defects.

After completing image annotation, the dataset is relatively small. To make the dataset more diverse and avoid overfitting during the training process, this study uses horizontal flip, vertical flip, random brightness, random contrast, etc., to enhance the dataset and generate the corresponding label files at the same time. The enhanced sample images are shown in Figure 2. The training set, validation set, and test set in this study are divided into 1000 images in the augmented dataset in the ratio of 8:1:1.





(e) Random Contrast



2.1.2. Fusion Image Enhancement Algorithm

Self-captured rail surface images are susceptible to light and other environmental factors, resulting in low image contrast and blurred defect details that are difficult to detect; therefore, enhancing the quality of rail surface images and highlighting defect information is the key to rail surface defect detection. The image enhancement algorithm is shown in Figure 3.

The HSV color space is more in line with how the human eye perceives color than the RGB color space is. To avoid adjusting the three color channels separately in the RGB color space, the rail surface image is first converted to the HSV color space, and the S component and the V component are adjusted while keeping the H component constant. The ensuing color distortion brings the improved image closer to how the human eye sees things.



Figure 3. Fusion image enhancement algorithm.

According to Retinex theory, an image can be regarded as consisting of an irradiation component and a reflection component. The irradiation component describes information about the light source, the brightness of the shooting environment, etc. The reflection component reflects the surface information of the object and portrays the essential characteristics of the object. The relationship between the image and the irradiation and reflection components is shown in Equation (1).

$$I(x,y) = L(x,y) \cdot R(x,y)$$
⁽¹⁾

The basic idea of single-scale Retinex (SSR) is to suppress the irradiation component in the original image and retain the reflective properties that reflect the essential characteristics of the target. The SSR algorithm is calculated as follows:

$$r(x,y) = \log R(x,y) = \log \frac{I(x,y)}{L(x,y)} = \log I(x,y) - \log[G(x,y) * I(x,y)]$$
(2)

where r(x, y) is the output image and * denotes the convolution calculation. A Gaussian filter is generally used in SSR, which is prone to misjudgment and leads to halo artifacts in the processed image. In this study, a bilateral filter—which can significantly enhance the smoothness and continuity of the picture boundary—is utilized to address this issue rather than a Gaussian filter. The bilateral filter introduces the Gaussian standard deviation C on the basis of Gaussian filtering and takes into account gray-scale similarity and spatial distribution, which can achieve edge-preserving denoising. The following is the bilateral filter expression:

$$f(x,y) = \sum_{(i,j)\in N_{(x,y)}} \omega(i,j) \cdot I(i,j) / \sum_{(i,j)\in N_{(x,y)}} \omega(i,j)$$
(3)

In Equation (3), f(x, y) is the filtered image; N(x, y) is the neighborhood centered on pixel I(x, y); and ω is the weighting factor, which is defined as follows:

$$\omega(i,j) = \omega_s(i,j) \times \omega_r(i,j) \tag{4}$$

$$\omega_s(i,j) = \exp(-\frac{(x-i)^2 + (y-j)^2}{2\sigma_s^2})$$
(5)

$$\omega_r(i,j) = \exp(-\frac{|I(i,j) - I(x,y)|^2}{2\sigma_r^2})$$
(6)

In Equations (9) and (10), the weighting coefficients in the spatial and grayscale domains are represented as ω_s and ω_r , respectively. The grayscale and spatial domains' corresponding standard deviations are shown by σ_s and σ_r , respectively.

After SSR processing based on the bilateral filter, the V component image r(x, y) is subjected to the adaptive gamma transform, as indicated in Formula (7), to enhance the dark features of the image, correct the exposure of the image, and make the color of the image more natural.

$$G(x,y) = [r(x,y)]^{\gamma}$$
(7)

$$\gamma = [2 + l(x, y)]^{[2 \times l(x, y) - 1]}$$
(8)

G(x, y) is the processed image. γ controls the degree of scaling of the whole transformation. l(x, y) is positively correlated with γ . By using adaptive gamma correction, we can effectively avoid the generation of overly bright areas while further enhancing the brightness of darker areas.

After adaptive Gamma correction, the image is subsequently subjected to contrastlimited adaptive histogram equalization (CLAHE) for contrast stretching to increase the image's contrast, which establishes a threshold value for each region of the histogram and evenly distributes the pixels above the threshold to other gray levels of the histogram. This approach limits the amount of variation in the histogram, suppresses the noise introduced during Histogram Equalization, and enhances local contrast without affecting the overall contrast.

After processing the V-component of the image, the S-component of the image is low, and the color is not full enough and needs to be stretched appropriately. In this study, the saturation is adaptively adjusted according to the luminance so that the overall image color is more realistic. First, the image luminance component enhancement multiplier is calculated, which is defined as shown below:

Ŀ

$$3 = \bigvee_{V}^{\prime} \bigvee_{V}$$
(9)

$$\beta' = \frac{\beta - \beta_{\min}}{\beta_{\max} - \beta_{\min}} \tag{10}$$

$$S' = \left[\frac{1}{2} + \beta' \times \frac{\max(R, G, B) + \min(R, G, B) + 1}{2 \times \max(R, G, B) + 1}\right] \times S$$
(11)

In the Equations (9)–(11), β denotes the luminance ratio. *V* and *V'* denote the luminance components before and after enhancement, respectively. β' denotes the normalized luminance ratio. *S* and *S'* denote the saturation components before and after adjustment, respectively. In the original image, max(*R*, *G*, *B*), min(*R*, *G*, *B*), and mean(*R*, *G*, *B*) stand for the respective RGB channel's associated pixels' greatest, lowest, and average values, respectively.

After processing in HSV color space, the image is returned to the RGB color space for display. An example of the steel rail surface image processed by the image enhancement algorithm of this study is shown in Figure 4.



Figure 4. Example of image enhancement effect.

The YOLOX algorithm is a high-performance detector proposed by Kuangwei Technology Research Institute 2021 that achieves an AP beyond YOLOv3, YOLOv4, and YOLOv5 with competitive inference speed. Figure 5 depicts the entire network layout of YOLOX, and the significant advancements over the prior YOLO series are briefly discussed.



Figure 5. YOLOX network structure.

CSPDarkNet, which employs the focus structure rather than convolution to compress the width and height of the input image, is used as the backbone feature extraction network in YOLOX. A cross-stage local network CspLayer, as shown in Figure 6, is introduced in the network instead of the original residual structure. The SPPBottleneck is also introduced into the network to increase the perceptual field and feature extraction capability of the network through the maximum pooling of different pooling kernels for feature extraction. The problem of over-learning is effectively avoided by using the SiLU activation function, which is shown in Equation (12).

$$SiLU(x) = x \cdot sigmoid(x) = x \cdot \frac{1}{1 + e^{-x}}$$
(12)



Figure 6. CspLayer structure.

A lightweight decoupling header with a separation of the localization and classification processes is introduced on the prediction side of YOLOX, which is structured as follows. The branch output of the predicted category is in the form $H \times W \times C$, the branch output of the predicted location is in the form of $H \times W \times 4$, and the branch output of the predicted IoU score is in the form of $H \times W \times 1$. The feature map's dimensions are H and W, and c stands for the target's category count. By using a lightweight decoupling head, the network convergence speed of YOLOX is improved. YOLOX decoupled head is shown in Figure 7.



Figure 7. YOLOX Decoupled Head.

In addition, YOLOX is an unanchored frame target detection algorithm that does not require an abundance of preset anchor frames, and the hyperparameter adjustment is much smaller than other algorithms in the YOLO series. In the practice detection mission, the positive and negative samples of YOLOX are not evenly distributed, so a preliminary screening of the prediction frame is performed first, which requires the feature points of the prediction frame to be within the real frame of the target object and the distance from the center of the target object to be within a certain range. Simple optimal transport assignment (SimOTA) is used to automatically select the feature maps to be matched after the initial screening is finished.

2.3. Improved the YOLOX Model

Figure 8 illustrates two important ways in which the model in this study has improved. The NAM attention mechanism is introduced to increase the feature expression capability of the image after the three feature layers of the backbone network have been extracted. This allows the network to concentrate more on the track surface to be measured and ignore the extraneous background information.

2.3.1. Principle of Improved Multi-Scale Feature Fusion Network

The neck structure of YOLOX extracts the target features by layered abstraction. The shallow layer of the network contains clear image location information, while the deep layer contains more semantic information about the image. YOLOX uses PANet as the feature fusion structure, which organically fuses the shallow and deep information of the network. However, the computing method for PANet is made more challenging by the network's intricate topology. In this study, we borrow the weighted bidirectional feature pyramid BiFPN for multi-feature fusion to improve the inference speed of the algorithm.



Figure 8. Improved YOLOX network structure.

To combine the feature information of various scales in the backbone network, BiFPN uses the same bi-directional channels of up- and down-sampling as PANet. Figure 9a depicts the structure of the BiFPN network, where the blue arrows indicate down-sampling and the red arrows indicate up-sampling. The input feature layer in the first column is half the length and width of the previous layer from top to bottom. The feature layers in the second column form the top-down pathway, accepting feature information from this level and the previous level, splicing and performing convolutional fusion, and then passing it into the next level and the bottom-up pathway. The feature layer in the third column constitutes the bottom-up pathway, which transmits the incoming features to the next layer or for prediction after convolutional fusion.



Figure 9. BiFPN and BiFPN Tiny network structure.

Since the fused feature layers have different resolutions and different importance to the output features, BiFPN adds learnable weights to all the inputs and continuously adjusts the network. Fast normalized fusion, as shown in the Equation (13), is used in BiFPN, which is less computationally intensive and has a similar accuracy compared with the Softmax function-based fusion approach.

$$OUT = \sum_{i} \frac{\omega_i}{\varepsilon + \sum_{i} \omega_j} \cdot IN_i$$
(13)

In Equation (13), IN_i and OUT denote the input and output features, respectively. The normalization and ReLU functions are used to make the weights $\omega_i \in [0, 1]$. ε is minimal and to keep the values stable. The backbone part of the network in this study extracts three effective feature layers, so the BiFPN is simplified (denoted as BiFPN-Tiny) to reduce computation while adapting the network. Figure 9b depicts the BiFPN-Tiny network structure.

2.3.2. Theorem of Incorporating NAM Attention Mechanism

Relying solely on the spontaneous transmission of feature information from the network, the lack of filtering of key information may lead to the neglect of small targets, which is difficult to apply in the case of complex backgrounds on the track surface and the presence of small defects. NAM is a lightweight and efficient attention mechanism that can effectively avoid the use of fully connected and convolutional layers to improve computational efficiency while maintaining similar performance to other attention mechanisms. By revamping the channel attention submodule (CAM) and the spatial attention submodule (SAM), NAM uses the modular integration strategy of CBAM and can be incorporated at the end of the residual structure. Figure 10 illustrates its construction.



Figure 10. CAM and SAM in the NAM module.

By figuring out the factor of scaling in batch normalization, the size of each channel change is reflected in NAM. The channel change is more dramatic, and the information it provides is proportionally richer and more significant the bigger the scaling factor. This is how the scaling factor is determined:

$$B_{out} = BN(B_{in}) = \gamma \frac{B_{in} - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} + \beta$$
(14)

where μ_B and σ_B are the mean and standard deviation, respectively. In CAM, γ denotes the scaling factor of each channel. In SAM, λ is the scaling factor and the weights of γ and λ are shown in the Equation (15).

$$W_{\gamma} = \gamma_i / \sum_{j=0} \gamma_j; W_{\lambda} = \lambda_i / \sum_{j=0} \lambda_j$$
(15)

In Figure 10, M_c and M_s are the output features of CAM and SAM. Their calculations are as follows:

$$M_c = sigmoid(W_{\gamma}(BN(F_1))) \tag{16}$$

$$M_s = sigmoid(W_{\lambda}(BN_s(F_2))) \tag{17}$$

The attention mechanism can be coupled with any feature layer as a plug-and-play module. The three effective feature layers recovered from the backbone section and the upand down-sampling in the multi-scale feature fusion network are employed first, and then seven NAMs are added in this study.

2.4. Model Evaluation Methods

2.4.1. Environmental Setup of the Experiment

To verify the effectiveness of the algorithm in this paper, we used the Windows 10 operating system. The deep learning framework was Python3.9 and Pytorch1.10.1, and the CPU was a dual-way Intel Xeon Silver 4310 @2.10 GHz, and the RAM was 32 G DDR4. The graphics card was the Nvidia GeForce RTX3080 10 G and we used the CUDA11.3 and cuDNN8.2.0 accelerated computing architecture.

The size of the input rail surface defect image was set to 640×640 , and the batch size was set to 16. Mixed precision training was used in this study's training method to conserve video memory. Gradient descent was performed using Adam's algorithm. The initial learning rate was set to 0.001. By integrating the data volume of this study's dataset, the number of iterative rounds of epochs was set to 300.

2.4.2. Model Evaluation Criteria

Recall, precision, and mean average precision (mAP) were utilized as the metrics to assess the detection effectiveness of the algorithm in this study, helping to further evaluate the efficacy and viability of the suggested model. Higher values imply greater detection performance of the model. mAP shows the comprehensive performance of the model in detecting all categories. Additionally, frames per second (fps) was used as the evaluation parameter to gauge how quickly the model could recognize objects. The following is the calculation for the evaluation metrics:

$$R = \frac{TP}{TP + FN} \tag{18}$$

$$P = \frac{TP}{TP + FP} \tag{19}$$

$$AP = \int_0^1 P(R)dR \tag{20}$$

$$mAP = \frac{1}{n} \sum_{i=1}^{n} AP_i$$
(21)

$$fps = \frac{n}{t} \tag{22}$$

3. Results

3.1. Contrasting Various Module Combination Patterns

In this study, the PAFPN of the neck structure is swapped out for the BiFPN multiscale feature fusion network using YOLOX as the baseline model. Additionally, the up-sampling and down-sampling operations of BiFPN, as well as the three output feature layers of CSPDarknet, now include the NAM attention mechanism. The model assessment metrics are compared under the same training settings as stated in Table 1.

Table 1. Comparison of the detection effects of various improvement schemes.

Scheme	Average Accuracy (%)	Average Recall (%)	mAP (%)	fps
YOLOX	87.70	85.40	90.78	78.40
YOLOX + Image Enhancement	91.35	89.33	91.95	70.26
YOLOX + Image Enhancement + BiFPN	93.47	90.61	92.87	71.69
YOLOX + Image Enhancement + BiFPN + NAM	94.56	91.71	93.20	71.33

The detection results for YOLOX and the algorithm used in this paper are shown in Figures 11 and 12, respectively. It is clear from the comparison that there is a significant

difference between the actual detection effect predicted by the previous model and the improved model in this work. YOLOX displays a missed detection in the detection results of Figure 11a,b, and it is challenging to precisely detect the tiny Defects; so, the detection results are unsatisfactory. YOLOX displays false detection and repeated detection in the findings provided in Figure 11c,d, which negatively impact the repair method and defect statistics during field inspection. Due to the darkness of the image itself and the findings displayed in column (e) of Figure 11, YOLOX was unable to detect the subtle dirt, and the fine defect was also not quantified. In contrast, the improved model presented in this study has a strong detection effect, no miss detection, no false detection, and can reliably detect minor defects, making it appropriate for the task of detecting rail surface defects.



Figure 11. Detection results of YOLOX (**a**) Defect, (**b**) Defect + Gap, (**c**) Defect + Dirt, (**d**) Gap + Dirt, and (**e**) Defect + Gap + Dirt.



Figure 12. Detection results of the algorithm of this paper (**a**) Defect, (**b**) Defect + Gap, (**c**) Defect + Dirt, (**d**) Gap + Dirt, and (**e**) Defect + Gap + Dirt.

3.2. Comprehensive Performance Comparison of Different Network Models

Considering the real-time requirements in industrial practice and to more thoroughly assess the efficacy of the algorithms in this study, six algorithms, faster R-CNN, SSD, YOLOv4 [57], YOLOv5-s, and YOLOv7-Tiny [45], YOLOX-Nano were selected for comparison. Among them, Resnet50 was chosen as the backbone network for faster R-CNN, and VGG16 was chosen as the backbone network for SSD. We used the same experimental environment as shown in Section 2.4.1.

The experimental comparison results between the algorithms in this chapter and other models are shown in Table 2. It is clear from the table that the algorithms in this chapter produced the best results for detecting all three categories of rail problems. The mAP@0.5 of this paper's algorithm is substantially higher than the classical faster R-CNN, SSD algorithm. Moreover, compared with YOLO series YOLOv4, YOLOv5, and YOLOv7-Tiny, the mAP@0.5 is improved by 5.43%, 3.34%, and 11.19%, respectively, and the accuracy and recall rate of all three defects are higher than the comparison algorithms. Therefore, the network model suggested in this paper is more accurate for the identification of rail surface defects. In terms of detection speed, the algorithm in this paper substantially outperforms the two-stage algorithm faster-RCNN, improving 19.83 fps and 6.5 fps over SSD and YOLOv4, respectively. Even if it is slower than YOLOv5, YOLOv7-Tiny, and YOLOX, the fps can be kept above 70, which can match the demand for the real-time detection of rail surface flaws with the use of appropriate hardware. In conclusion, the upgraded YOLOX performs better all around.

Table 2. Performance evaluation of different models' detecting abilities.

Model	P (%)			R (%)			mAP@0.5	
	Defect	Dirt	Gap	Defect	Dirt	Gap	- (%)	ips
Faster R-CNN	77.80	62.73	72.86	75.59	66.67	75.00	68.39	13.43
SSD	79.47	73.00	82.51	65.00	63.33	72.58	75.13	53.50
YOLOv4	87.70	85.71	82.14	79.95	80.00	75.00	87.77	66.83
YOLOv5	87.72	87.30	85.17	73.53	86.67	82.13	89.86	78.65
YOLOv7-Tiny	82.24	88.38	86.77	87.77	72.26	89.95	82.01	95.32
YOLOX	83.24	83.67	88.65	85.44	86.67	84.11	90.78	80.40
Algorithm of this paper	94.75	95.06	93.86	91.68	90.70	92.75	93.20	73.33

In the algorithm of this paper, the darker rail surface image is enhanced, which can highlight the details of the rail surface defects and make the edges of the defects clearer. At the same time, the algorithm uses YOLOX as the baseline model, in which BiFPN replaces the original PANet, so that the network has efficient cross-scale connections and feature reuse is more absolute rather than average. The addition of the NAM attention mechanism can suppress unimportant pixels and make the network more efficient. Combining the above improvement strategies, the algorithm in this paper has achieved better detection results.

4. Conclusions

This research suggests an image enhancement and enhanced YOLOX-based rail surface defect detection method to increase the detection efficiency and accuracy of rail surface defects. To highlight defect information, the fusion image enhancement algorithm in the HSV color space is used. To improve the algorithm's inference speed, the PAFPN in the YOLOX network is replaced with the BiFPN depth feature fusion, and the NAM attention mechanism is added to improve the ability to characterize images.

The algorithm presented in this paper has superior accuracy, recall, and confidence in rail surface defect identification when compared to the enhanced YOLOX model and other comparable models. The enhanced model performs well in terms of localization

13 of 15

accuracy and detection accuracy, and it can satisfy practical requirements for real-time defect detection for rail surface defects, providing a trustworthy and precise detection technique to guarantee the security of rail tracks as well as the safe operation and upkeep of high-speed railroads.

In the future, we will devote ourselves to collecting more images of rail surface defects, enriching our dataset, and making the types of defects more detailed. At the same time, it is also an important research direction to deploy our algorithm in embedded devices and apply it in industrial practice. In terms of algorithms, we will continue to pay attention to the latest developments in surface defect detection and strive to improve detection efficiency. Combining vision transformers with YOLO series is the focus of our next work.

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

Funding: This research was funded by Liaoning Provincial Transportation Technology Project under Grant 202244, and the Open Project Program of the Traction Power State Key Laboratory of Southwest Jiaotong University under Grant TPL2203.

Data Availability Statement: All data, models, and code generated or used during the study appear in the submitted article.

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

Abbreviations

BiFPN	Bidirectional Feature Pyramid Network
NAM	Normalization-based Attention Module
SSD	Single-Shot Multi-Box Detector
YOLO	You Only Look Once
CSP	Cross Stage Partial Connections
CAM	Channel Attention Module
SAM	Spatial Attention Module
Р	Precision
R	Recall
AP	average precision
mAP	mean average precision
fps	frames per second
PAFPN	Path Aggregation Feature Pyramid Network
TP	number of true predictions
FP	number of positive samples of false predictions
FN	number of negative samples of false predictions

References

- 1. Mićić, M.; Brajović, L.; Lazarević, L.; Popović, Z. Inspection of RCF rail defects–Review of NDT methods. *Mech. Syst. Signal Process.* **2023**, *182*, 109568. [CrossRef]
- Song, Y.; Zhao, G.; Zhang, B.; Chen, H.; Deng, W.Q.; Deng, Q. An enhanced distributed differential evolution algorithm for portfolio optimization problems. *Eng. Appl. Artif. Intell.* 2023, 121, 106004. [CrossRef]
- 3. Li, M.; Zhang, J.; Song, J.; Li, Z.; Lu, S. A clinical-oriented non severe depression diagnosis method based on cognitive behavior of emotional conflict. *IEEE Trans. Comput. Soc. Syst.* 2022, 10, 131–141. [CrossRef]
- 4. Yu, Y.; Tang, K.; Liu, Y. A fine-tuning based approach for daily activity recognition between smart homes. *Appl. Sci.* **2023**, *13*, 5706. [CrossRef]
- 5. Huang, C.; Zhou, X.; Ran, X.; Wang, J.; Chen, H.; Deng, W. Adaptive cylinder vector particle swarm optimization with differential evolution for UAV path planning. *Eng. Appl. Artif. Intell.* **2023**, *121*, 105942. [CrossRef]
- Cai, J.; Ding, S.; Zhang, Q.; Liu, R.; Zeng, D.; Zhou, L. Broken ice circumferential crack estimation via image techniques. *Ocean.* Eng. 2022, 259, 111735. [CrossRef]
- 7. Zhou, X.; Cai, X.; Zhang, H.; Zhang, Z.; Jin, T.; Chen, H.; Deng, W. Multi-strategy competitive-cooperative co-evolutionary algorithm and its application. *Inf. Sci.* 2023, *635*, 328–344. [CrossRef]

- Sun, Q.; Zhang, M.; Zhou, L.; Garme, K.; Burman, M. A machine learning-based method for prediction of ship performance in ice: Part, I. ice resistance. *Mar. Struct.* 2022, *83*, 103181. [CrossRef]
- Duan, Z.; Song, P.; Yang, C.; Deng, L.; Jiang, Y.; Deng, F.; Jiang, X.; Chen, Y.; Yang, G.; Ma, Y.; et al. The impact of hyperglycaemic crisis episodes on long-term outcomes for inpatients presenting with acute organ injury: A prospective, multicentre follow-up study. *Front. Endocrinol.* 2022, 13, 1057089. [CrossRef]
- 10. Ren, Z.; Zhen, X.; Jiang, L.; Gao, Z.; Li, Y.; Shi, W. Underactuated control and analysis of single blade installation using a jackup installation vessel and active tugger line force control. *Mar. Struct.* **2023**, *88*, 103338. [CrossRef]
- 11. Xie, C.; Zhou, L.; Ding, S.; Liu, R.; Zheng, S. Experimental and numerical investigation on self-propulsion performance of polar merchant ship in brash ice channel. *Ocean. Eng.* 2023, 269, 113424. [CrossRef]
- 12. Li, M.; Zhang, W.; Hu, B.; Kang, J.; Wang, Y.; Lu, S. Automatic assessment of depression and anxiety through encoding pupil-wave from HCI in VR scenes. *ACM Trans. Multimid. Comput. Commun. Appl.* **2022**. [CrossRef]
- Chen, T.; Song, P.; He, M.; Rui, S.; Duan, X.; Ma, Y.; Armstrong, D.G.; Deng, W. Sphingosine-1-phosphate derived from PRP-Exos promotes angiogenesis in diabetic wound healing via the S1PR1/AKT/FN1 signalling pathway. *Burn. Trauma* 2023, *11*, tkad003. [CrossRef] [PubMed]
- 14. Yue, G.; Cui, X.; Zhang, K.; An, D. Guided wave propagation for monitoring the rail base. *Math. Probl. Eng.* **2020**, 2020, 4756574. [CrossRef]
- 15. Yuan, M.; Tse, P.W.; Xuan, W.; Xu, W. Extraction of least-dispersive ultrasonic guided wave mode in rail track based on floquet-bloch theory. *Shock. Vib.* **2021**, 2021, 6685450. [CrossRef]
- 16. Jia, Y.; Zhang, S.; Wang, P.; Ji, K. A method for detecting surface defects in railhead by magnetic flux leakage. *Appl. Sci.* **2021**, *11*, 9489. [CrossRef]
- 17. Gao, B.; Bai, L.; Woo, W.L.; Tian, G.Y.; Cheng, Y. Automatic Defect Identification of Eddy Current Pulsed Thermography Using Single Channel Blind Source Separation. *IEEE Trans. Instrum. Meas.* **2014**, *63*, 913–922. [CrossRef]
- Yu, H.; Li, Q.; Tan, Y.; Gan, J.; Wang, J.; Geng, Y.; Jia, L. A coarse-to-fine model for rail surface defect detection. *IEEE Trans. Instrum. Meas.* 2018, 68, 656–666. [CrossRef]
- 19. Zhang, D.; Song, K.; Wang, Q.; He, Y.; Wen, X.; Yan, Y. Two deep learning networks for rail surface defect inspection of limited samples with line-level label. *IEEE Trans. Ind. Inform.* 2020, *17*, 6731–6741. [CrossRef]
- 20. Wang, H.; Li, M.; Wan, Z. Rail surface defect detection based on improved Mask R-CNN. *Comput. Electr. Eng.* **2022**, 102, 108269. [CrossRef]
- 21. Hu, J.; Qiao, P.; Lv, H.; Yang, L.; Ouyang, A.; He, Y.; Liu, Y. High speed railway fastener defect detection by using improved YoLoX-Nano Model. *Sensors* 2022, 22, 8399. [CrossRef] [PubMed]
- 22. Feng, J.H.; Yuan, H.; Hu, Y.Q.; Lin, J.; Liu, S.W.; Luo, X. Research on deep learning method for rail surface defect detection. *IET Electr. Syst. Transp.* **2020**, *10*, 436–442. [CrossRef]
- 23. Zhang, D.; Song, K.; Xu, J.; He, Y.; Niu, M.; Yan, Y. MCnet: Multiple context information segmentation network of no-service rail surface defects. *IEEE Trans. Instrum. Meas.* 2020, *70*, 5004309. [CrossRef]
- 24. Jin, X.; Wang, Y.; Zhang, H.; Zhong, H.; Liu, L.; Wu, Q.M.J.; Yang, Y. DM-RIS: Deep multimodel rail inspection system with improved MRF-GMM and CNN. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 1051–1065. [CrossRef]
- 25. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
- Xu, J.; Zhao, Y.; Chen, H.; Deng, W. ABC-GSPBFT: PBFT with grouping score mechanism and optimized consensus process for flight operation data-sharing. *Inf. Sci.* 2023, 624, 110–127. [CrossRef]
- 27. Jin, T.; Zhu, Y.; Shu, Y.; Cao, J.; Yan, H.; Jiang, D. Uncertain optimal control problem with the first hitting time objective and application to a portfolio selection model. *J. Intell. Fuzzy Syst.* **2022**, *44*, 1585–1599. [CrossRef]
- 28. Deng, W.; Zhang, L.; Zhou, X.; Zhou, Y.; Sun, Y.; Zhu, W.; Chen, H.; Deng, W.; Chen, H.; Zhao, H. Multi-strategy particle swarm and ant colony hybrid optimization for airport taxiway planning problem. *Inf. Sci.* **2022**, *612*, 576–593. [CrossRef]
- 29. Yu, C.; Liu, C.; Yu, H.; Song, M.; Chang, C.-I. Unsupervised Domain Adaptation with Dense-Based Compaction for Hyperspectral Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2021, 14, 12287–12299. [CrossRef]
- 30. Jin, T.; Yang, X. Monotonicity theorem for the uncertain fractional differential equation and application to uncertain financial market. *Math. Comput. Simul.* 2021, 190, 203–221. [CrossRef]
- Yu, C.; Gong, B.; Song, M.; Zhao, E.; Chang, C.-I. Multiview Calibrated Prototype Learning for Few-shot Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2022, 60, 5544713. [CrossRef]
- Bi, J.; Zhou, G.; Zhou, Y.; Luo, Q.; Deng, W. Artificial Electric Field Algorithm with Greedy State Transition Strategy for Spherical Multiple Traveling Salesmen Problem. *Int. J. Comput. Intell. Syst.* 2022, 15, 5. [CrossRef]
- Huang, C.; Zhou, X.B.; Ran, X.J.; Liu, Y.; Deng, W.Q.; Deng, W. Co-evolutionary competitive swarm optimizer with three-phase for large-scale complex optimization problem. *Inf. Sci.* 2023, 619, 2–18. [CrossRef]
- Wei, Y.; Zhou, Y.; Luo, Q.; Deng, W. Optimal reactive power dispatch using an improved slime mould algorithm. *Energy Rep.* 2021, 7, 8742–8759. [CrossRef]
- 35. Jin, T.; Gao, S.; Xia, H.; Ding, H. Reliability analysis for the fractional-order circuit system subject to the uncertain random fractional-order model with Caputo type. *J. Adv. Res.* **2021**, *32*, 15–26. [CrossRef] [PubMed]

- 36. Wu, E.Q.; Zhou, M.; Hu, D.; Zhu, L.; Tang, Z.; Qiu, X.-Y.; Deng, P.-Y.; Zhu, L.-M.; Ren, H. Self-paced dynamic infinite mixture model for fatigue evaluation of pilots' brain. *IEEE Trans. Cybern.* **2020**, *52*, 5623–5638. [CrossRef] [PubMed]
- 37. Deng, W.; Xu, J.J.; Gao, X.Z.; Zhao, H.M. An enhanced MSIQDE algorithm with novel multiple strategies for global optimiza-tion problems. *IEEE Trans. Syst. Man Cybern. Syst.* 2022, 52, 1578–1587. [CrossRef]
- 38. Wu, J.; Wang, Z.; Hu, Y.; Tao, S.; Dong, J. Runoff forecasting using convolutional neural networks and optimized bi-directional long short-term memory. *Water Resour. Manag.* **2023**, *37*, 937–953. [CrossRef]
- 39. Deng, W.; Shang, S.F.; Cai, X.; Zhao, H.; Zhou, Y.; Chen, H.; Deng, W. Quantum differential evolution with cooperative coevolution framework and hybrid mutation strategy for large scale optimization. *Knowl.-Based Syst.* **2021**, 224, 107080. [CrossRef]
- 40. Deng, W.; Xu, J.; Song, Y.; Zhao, H. Differential evolution algorithm with wavelet basis function and optimal mutation strategy for complex optimization problem. *Appl. Soft Comput.* **2021**, 100, 106724. [CrossRef]
- Deng, W.; Liu, H.; Xu, J.; Zhao, H.; Song, Y. An improved quantum-inspired differential evolution algorithm for deep belief network. *IEEE Trans. Instrum. Meas.* 2020, 69, 7319–7327. [CrossRef]
- 42. Chen, M.; Shao, H.; Dou, H.; Li, W.; Liu, B. Data augmentation and intelligent fault diagnosis of planetary gearbox using ILoFGAN under extremely limited sample. *IEEE Trans. Reliab.* **2022**, 1–9. [CrossRef]
- Joseph, R.; Ali, F. Yolov3: An incremental improvement. arXiv 2018, arXiv:1804.02767.
- 44. Zheng, G.; Songtao, L.; Wang, F.; Zeming, L.; Jian, S. Yolox: Exceeding yolo series in 2021. arXiv 2021, arXiv:2107.08430.
- 45. Wang, C.-Y.; Alexey, B.; Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* 2022, arXiv:2207.02696.
- Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* 2015, 28. Available online: https://proceedings.neurips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed3 8046-Abstract.html (accessed on 12 June 2023). [CrossRef]
- Cai, Z.; Nuno, V. Cascade R-CNN: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
- 48. Wang, W.; Yuan, X.; Chen, Z.; Wu, X.; Gao, Z. Weak-light image enhancement method based on adaptive local gamma transform and color compensation. *J. Sens.* **2021**, 2021, 5563698. [CrossRef]
- 49. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
- 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 (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
- 51. Liu, Y.; Shao, Z.; Teng, Y.; Hoffmann, N. NAM: Normalization-based attention module. arXiv 2021, arXiv:2111.12419.
- Li, W.; Zhang, L.; Wu, C.; Cui, Z. A new lightweight deep neural network for surface scratch detection. *Int. J. Adv. Manuf. Technol.* 2022, 123, 1999–2015. [CrossRef]
- 53. Chen, Y.; Ding, Y.; Zhao, F.; Zhang, E.; Wu, Z.; Shao, L. Surface defect detection methods for industrial products: A review. *Appl. Sci.* **2021**, *11*, 7657. [CrossRef]
- 54. Park, J.; Woo, S.; Lee, J.-Y.; Kweon, I.S. Bam: Bottleneck attention module. arXiv 2018, arXiv:1807.06514.
- 55. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, C.A. SSD: Single Shot MultiBox Detector. *arXiv* 2015, arXiv:1512.02325.
- 56. Gan, J.; Li, Q.; Wang, J.; Yu, H. A hierarchical extractor-based visual rail surface inspection system. *IEEE Sens. J.* 2017, 17, 7935–7944. [CrossRef]
- 57. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. arXiv 2020, arXiv:2004.10934.

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.