Improving Tire Specification Character Recognition in the YOLOv5 Network

Qing Zhao, Honglei Wei * and Xianyi Zhai

School of Mechanical Engineering and Automation, Dalian Polytechnic University, Dalian 116034, China
* Correspondence: weihl2005@163.com

Abstract: The proposed method for tire specification character recognition based on the YOLOv5 network aimed to address the low efficiency and accuracy of the current character recognition methods. The approach involved making three major modifications to the YOLOv5 network to improve its generalization ability, computation speed, and optimization. The first modification involved changing the coupled head in YOLOv5 to a decoupled head, which could improve the network's generalization ability. The second modification proposed incorporating the C3-Faster module, which would replace some of the C3 modules in YOLOv5's backbone and head and improve the network's computation speed. Finally, the third modification proposed replacing YOLOv5's CIoU loss function with the WIoU loss function to optimize the network. Comparative experiments were conducted to validate the effectiveness of the proposed modifications. The C3-Faster module and the WIoU loss function were found to be effective, reducing the training time of the improved network and increasing the mAP by 3.7 percentage points in the ablation experiment. The experimental results demonstrated the effectiveness of the proposed method in improving the accuracy of tire specification character recognition and meeting practical application requirements. Overall, the proposed method showed promising results for improving the efficiency and accuracy of automotive tire specification character recognition, which has potential applications in various industries, including automotive manufacturing and tire production.

Keywords: YOLOv5; decoupling head; C3-Faster; WIoU; tire specification character recognition

1. Introduction

Automobile tires are an important part of automobile safety, and their design, manufacture, and use require strict standards and specifications [1]. The tire production process involves a large number of character marks, such as specifications, models, batches, and other information, which are of great significance for the quality control and traceability of tires. In the production of automobile tires, it is necessary to classify the tires according to the size of the specifications and different models. The traditional method of manually recognizing tire characters is time-consuming and inefficient and has a high error rate. Tire character recognition using machine vision is more effective than the traditional recognition method, but it has higher requirements for the recognition environment and harsher conditions. Therefore, it is of great practical significance to study an automatic tire identification method.

Currently, many achievements have been made in character recognition based on machine vision systems. Wang H. et al. [2] proposed a machine vision-based tire rubber surface character recognition method which implements character recognition through character localization, character segmentation, morphological processing, and template matching. Peng Q. et al. [3] proposed a steel plate billet spray marking character recognition system which uses an image sensor to capture images in real-time and the Baidu Paddle-OCR character recognition algorithm to achieve automatic character recognition. Chen S. et al. [4] proposed a multi-line character recognition method for clutch flywheels.

Firstly, the clutch flywheel image is preprocessed and the edge coordinates are extracted. Secondly, the circular ring area is located by using the DBSCAN clustering algorithm, and then character segmentation is performed using the pixel projection method. Finally, LBP (local binary patterns) and support vector machine are used for recognition. Bai S. et al. [5] proposed a machine tool information acquisition method which uses the projection transformation principle to determine the machine tool panel range, preprocesses the acquired image using filtering methods, and, finally, achieves character recognition through convolutional neural networks. Yang G. et al. [6] developed a chip character recognition system. Firstly, the character area is obtained by using the grayscale value projection method for character segmentation. Secondly, chip positioning is performed using shape matching, and then character recognition is achieved through BP (back propagation) neural networks. Traditional machine vision methods have higher requirements for image quality and have lower versatility. The recognition effect will also change when the environment changes.

With the development of deep learning technology, object detection and recognition have greatly improved. The YOLO (You Only Look Once) algorithm series is comprised of fast and accurate object detection algorithms [7]. Through improvements to YOLOv5's model structure, the introduction of new data augmentation techniques [8], higher detection speed, and better accuracy have been achieved. YOLOv5 technology has also been widely used. Gong P. et al. [9] proposed a steel stamp character recognition method, based on YOLOv5, which first expanded the dataset through image preprocessing (which was trained with YOLOv5) and then recognized characters using the trained model. However, a large amount of computing resources is required for training and running during training. Zhang J. et al. [10] proposed a vehicle and tank number detection and recognition method based on an improved YOLOv5 network which added attention mechanisms and GBN modules to enhance the feature extraction capabilities and improve detection speed. Adding attention mechanisms improved accuracy but increased the model's computational complexity, leading to overfitting. Laroca [11] proposed an end-to-end, efficient, and independent automatic LP (license plate) recognition system, based on the YOLO model, which included a unified LP detection and layout classification method. The system achieved a balance between accuracy and speed, but the CNNs (convolutional neural networks) used in this method required a large amount of labeled data for training and were less effective in some scenarios. Aduen [12] proposed YOLO-Z, an improved YOLOv5 method, for detecting small objects in autonomous driving, which simplified Pan-Net into FPN and replaced it with biFPN. However, this method's ability to process scale changes was limited, making it difficult to detect objects with large scale differences in the same image. Jiang L. et al. [13] proposed a method for traffic sign detection, based on the YOLOv5 network model, which used a balanced feature pyramid structure and a global context block to enhance feature fusion and feature extraction capabilities. However, in some cases, the balanced feature pyramid structure resulted in information loss and could lead to decreased detection performance when there were too few low-resolution features.

To improve the accuracy and precision of recognizing characters on automobile tires, this paper proposes an improved automobile tire character recognition model based on YOLOv5. The model added a decoupled head, replaced the C3 module with the C3-Faster module, and replaced the CIOU with WIOU in the original YOLOv5 network, which enhanced the accuracy and precision of the network in recognizing the characters on automobile tires.

2. Improving the YOLOv5 Network

YOLOv5 has made significant improvements compared to previous YOLO models and is currently one of the more advanced models in object detection [14]. In YOLOv5, new techniques have been used, such as adaptive training data augmentation [15], multi-scale training [16], and multi-scale prediction for initial detection layers [17], which make detection faster and more accurate. The YOLOv5 network model is shown in Figure 1. However, conducting character recognition on the surface of a tire is complex because the characters

are molded into the tire and blend into the complex background. Additionally, dust on the surface of a tire further complicates character recognition. To address these issues and strengthen the network's recognition capabilities, this paper proposes an improved model based on YOLOv5 for recognizing characters on the surfaces of car tires.
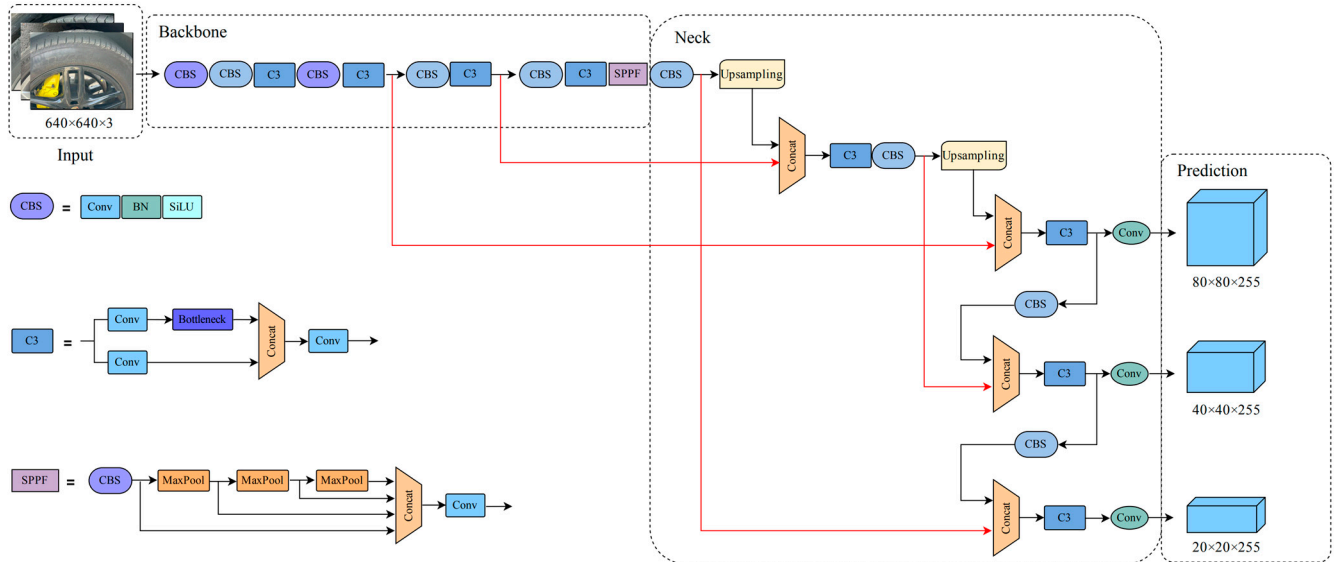

Figure 1. YOLOv5 network model.

This article proposes three main improvements to the network. Firstly, it proposes separating the feature extraction and output of detection and segmentation tasks to accelerate model convergence and improve model accuracy. Secondly, it proposes improving the FasterNet Block by introducing C3-Faster to reduce the number of convolution operations and increase the computational speed. Thirdly, it proposes using WIoU-Loss as the loss function to measure the similarity between the predicted bounding boxes and the actual annotations. The improved network model in this study is shown in Figure 2. In the figure, the green rectangular frame indicates the improved part.
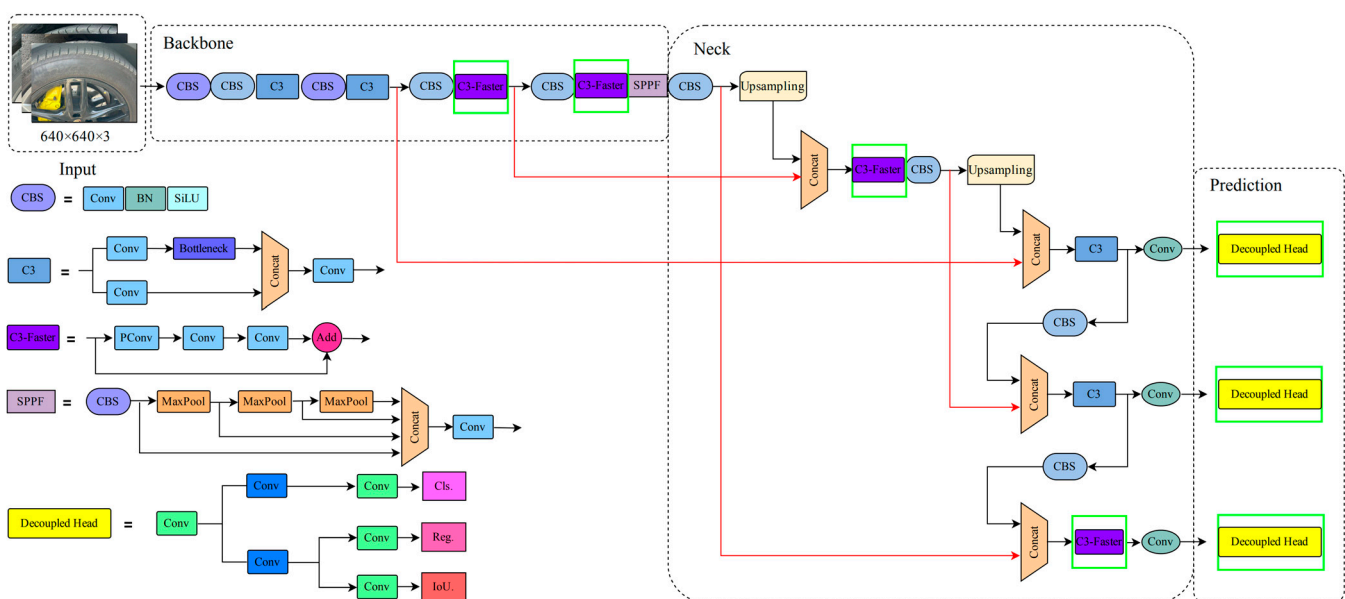

Figure 2. The improved network model.

2.1. The YOLOv5 Decoupled Head

A decoupling head [18] is a technique that separates convolutional layers from fully connected layers, reducing computational complexity and model size, thus improving model speed and efficiency. The traditional detection head in YOLOv5 is a coupling head [19] that typically consists of a fully connected layer that converts the feature map outputs by convolutional layers into a prediction vectors. This fully connected layer is usually trained together with the convolutional layers, requiring significant computational resources and longer training times. In this study, the decoupling head was added to YOLOv5, with the aim of separating the classification and localization branches to reduce computational resources and training time, as shown in Figure 3. First, a 1×1 convolution was applied to the output feature map to reduce the model's complexity while also reducing the number of input data channels and adjusting the feature map size. Then, the convolutionally processed result was split into two branches, each undergoing 3×3 convolutional processing. The first branch was further processed with a 1×1 convolution to obtain the classification branch while the second branch was split into two further branches, each undergoing 1×1 convolutional processing to obtain the target coordinate and confidence branches.
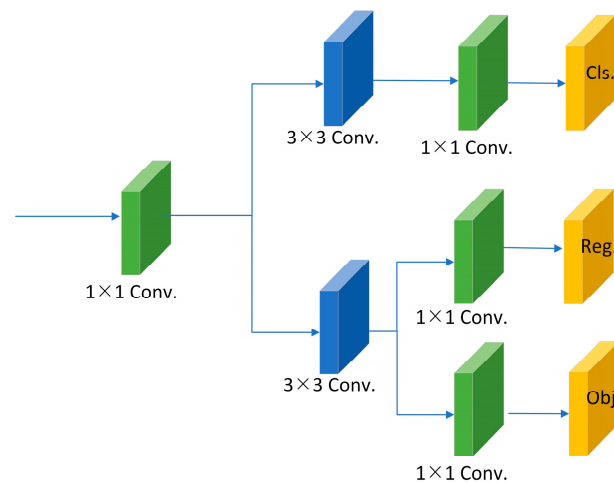

Figure 3. Decoupling head.

The decoupling head separated the classification and localization branches, reducing the number of parameters and calculations needed in the model and greatly speeding up the training and inference speed, improving the model's perception ability for target features of different scales and, thus, improving the model's robustness and accuracy while reducing the occurrence of overfitting.

2.2. Improving FasterNet Block by Proposing C3-Faster

In YOLOv5, the depth and receptive field of the network are increased through the C3 module [20], and the feature extraction ability of the network is improved. The C3 module is shown in Figure 4 and consists of three 3×3 Conv kernels and several bottleneck modules. Among them, the first one is a 1×1 Conv kernel with a step size of two, which halves the size of the feature map, reduces the number of parameters, and increases the receptive field of the network. The second and third modules have step sizes of 1×1 Conv kernels, and they retain more local information without changing the size of the feature map, further extract features, and increase the depth and receptive field of the network model. In the bottleneck module, the channel number of the image is reduced by half by a 1×1 Conv kernel with a step size of one, and then the number of channels of the image is doubled by a 1×1 Conv kernel with a step size of three. The number of channels of the image remains the same, the parameters of the network are reduced, and the depth is increased.
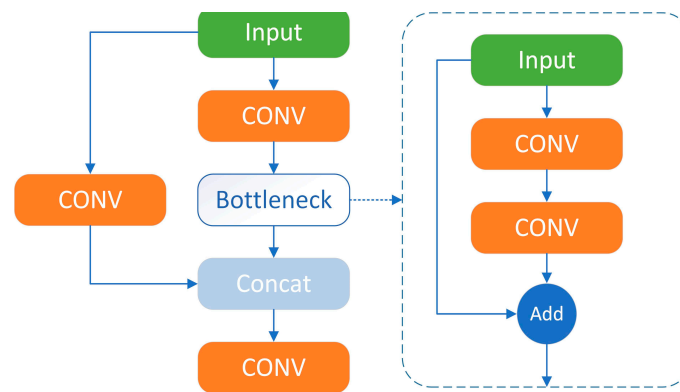
Figure 4. The C3 module.

When running the C3 module once, it requires five convolution operations, and generating too many parameters will consume too much memory, further limiting the operating efficiency of the model, prolonging the training time, and affecting the processing speed of the model. In order to further improve the speed and accuracy of the network model for tire character recognition, in this study, FasterNet Block [21] was improved and the C3-Faster module was proposed and added to the YOLOv5 network structure. The C3-Faster module is shown in Figure 5 and consists of one 3×3 PConv and two 1×1 Convs. First, the feature map was calculated by the first PConv, in which the PConv could reduce redundant information and memory usage in the calculation, and then it passed through two 1×1 Conv kernels in turn to obtain the effective information of the feature map, and finally, it output the effective information, after which we could proceed to the next step. In this study, the parameter volume using the C3 module was 6.25 m, and the parameter volume using the C3-Faster module was 4.57 m. Fewer parameters can reduce the memory footprint and computational cost of a model, reducing the risk of overfitting.
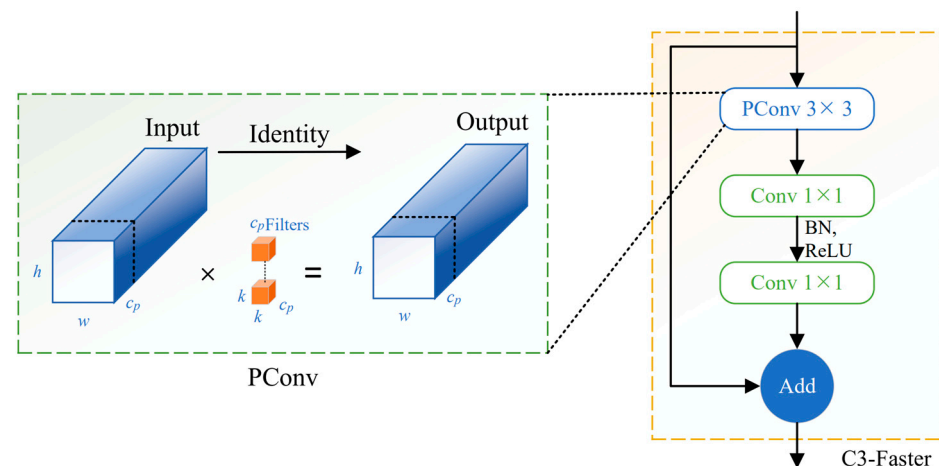

Figure 5. The C3-Faster module.

2.3. Improved Regression Loss Function

As a loss function, IoU-Loss is used to measure the similarity between predicted bounding boxes and actual annotations, with more emphasis on the overlap between the predicted results and the ground truth [22]. It is the most widely used metric for measuring the similarity between bounding boxes, but in tire logo recognition, the characters are relatively small, and using IoU may result in cases where the predicted bounding box and the ground truth do not intersect, resulting in the IoU being zero and unable to be optimized. GIoU [23] adds the minimum bounding box of the predicted and actual bounding boxes, which solves the problem of the IoU being zero, but when the predicted and actual bounding boxes have the same widths and heights and are on the same horizontal or vertical line,

GIoU degenerates into IoU. DIoU [24] adds a Euclidean distance between the center points of the two bounding boxes and a Euclidean distance between the two diagonal vertices of the minimum rectangle box based on GIoU, but DIoU degenerates into IoU when the two bounding boxes have the same center point but different aspect ratios. CIoU [25] is used as the loss function in YOLOv5. CIoU considers the consistency of the aspect ratio between the predicted and ground truth bounding boxes based on DIoU, and it adds a penalty term for the aspect ratio. However, due to the use of complex function calculations, CIoU consumes a lot of computing power in the calculation process, increasing the training time. WIoU [26] proposes a dynamic non-monotonic focus mechanism which uses "out-lierness" instead of IoU to evaluate the quality of anchor boxes, and it adopts a gradient gain allocation strategy to not only reduce the competitiveness of high-quality anchor boxes but also reduce the harmful gradients produced by low-quality anchor boxes, which enables WIoU to focus on low-quality anchor boxes and improve the overall performance of the detector.

There are three versions of WIoU, among which WIoUv1 constructs an attention-based bounding box loss and WIoUv2 and WIoUv3 are obtained by adding a gradient gain to the focus mechanism on the basis of v1. In Figure 6 [26], the green rectangle indicates the annotation box, the gray rectangle indicates the prediction box, and the blue rectangle indicates the minimum bounding box.
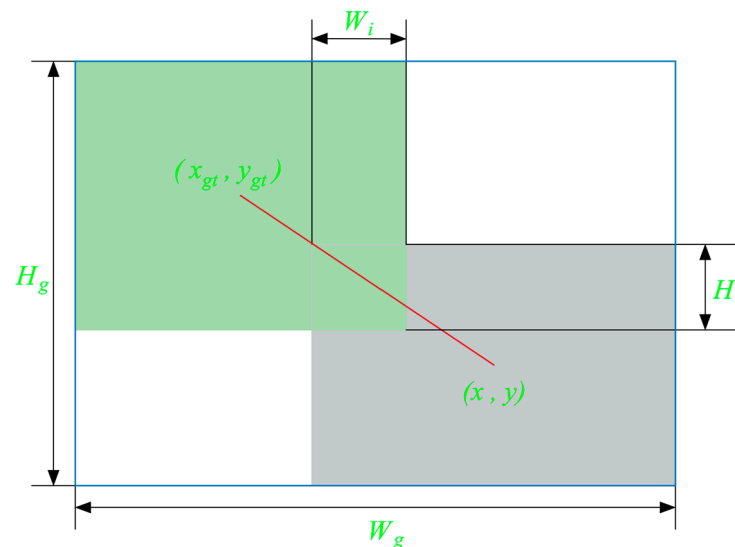
Figure 6. Schematic diagram of annotation box, prediction box, and minimum bounding box.

The calculation formula of the loss function L_{WIoUv1} [26] for *WIoUv1* is shown in Equations (1)–(3):

$$L_{IoU} = 1 - IoU = 1 - \frac{W_i H_i}{wh + w_{gt} h_{gt} - W_i H_i} \overline{IoU},$$

(1)

$$R_{WIoU} = exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{\left(W_g^2 + H_g^2\right)}\right), \text{and}$$

(2)

$$L_{WIoUv1} = R_{WIoU} L_{IoU}.$$

(3)

The calculation formula for the loss function L_{WIoUv2} [26] of *WIoUv2* is shown in Equation (4):

$$L_{WIoUv2} = \left(\frac{L^*_{IoU}}{\overline{L_{IoU}}}\right)^\gamma L_{WIoUv1}.$$

(4)

In Equation (4), L^*_{IoU} represents the monotonic attention coefficient and $\overline{L_{IoU}}$ is the mean value, which is normalized in the formula to keep the gradient gain at a high level.

The calculation formula for the loss function L_{WIoUv3} [26] of $WIoUv3$ is shown in Equations (5) and (6):

$$\beta = \frac{L^*_{IoU}}{\overline{L_{IoU}}} \in [0, +\infty), \text{and} \tag{5}$$

$$L_{WIoUv3} = rL_{WIoUv1}, r = \frac{\beta}{\delta\alpha^{\beta-\delta}}. \tag{6}$$

In Equations (5) and (6), β is the non-monotonic focusing coefficient and α and δ are hyperparameters. When $\beta = \delta$, $r = 1$, and when the outlier degree of the anchor box satisfies $\beta = C$ (C is a fixed value), the anchor box will obtain the highest gradient gain. The values of β and r are controlled by the hyperparameters α and δ. The relationship between the hyperparameters α and δ, the outlier β, and the gradient gain r is shown in Figure 7 [26].
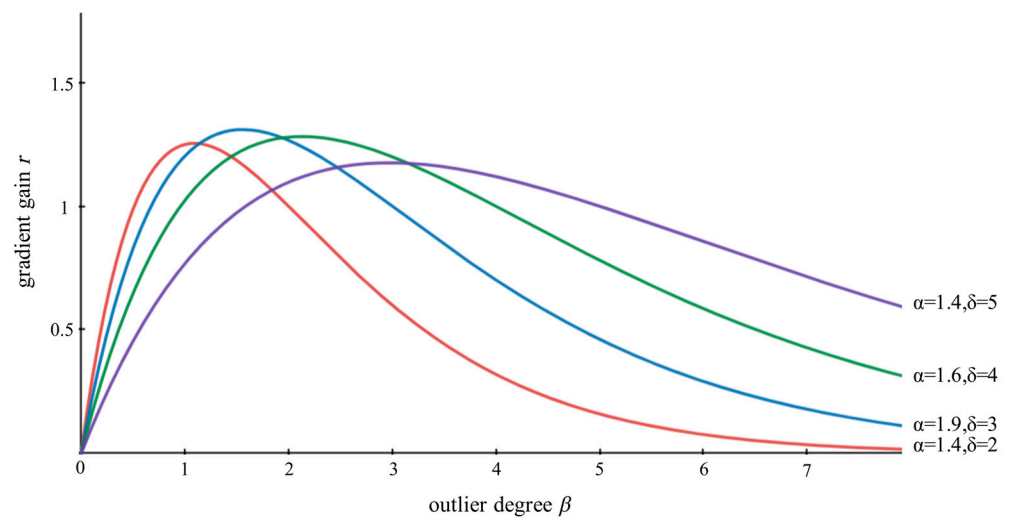
Figure 7. The relationship between the hyperparameters α and δ, the outlier β, and the gradient gain r.

3. Experiments

3.1. Model Training

The main tasks during model training are collecting and annotating the dataset and setting the training parameters of the network.

(1) Dataset processing: We randomly selected car tires in a parking lot for image collection, and a total of 1000 images were collected, including 800 training sets, 100 verification sets, and 100 test sets. We used the annotation tool Labelimg to annotate the dataset, export YOLO format annotation files, and prepare for the next step of training. The labeled sample of the dataset is shown in Figure 8.

Figure 8. Dataset labeled samples.

(2) Setting the network training parameters: The operating system used was Windows 11, the GPU was an NVIDIA GeForce RTX 3060, and the programming language was Python 3.9. The network training parameters are shown in Table 1, and the smaller yolov5s model was selected during training.

Table 1. Network training parameters.

Parameter	Value
number of images	1000
epochs	400
batch size	48
learning rate	0.01
weight decay	0.0005

3.2. Evaluation Index

The trained model needed to be evaluated for the accuracy of the detection. In this experiment, precision (*P*) and mean average precision (*mAP*) were used to evaluate the performance of the model, as follows:

$$Precision = \frac{TP}{TP + FP}, \text{and} \tag{7}$$

$$mAP = \frac{\sum\limits_{j=0}^{n} AP(j)}{n}. \tag{8}$$

Formula (7) calculates the precision, where *TP* is a true positive and *FP* is a false positive. Formula (8) calculates the *mAP*, where *AP(j)* is the average precision for the *j* defect class, with *j* representing the number of defect categories, i.e., $j = 0, 1, 2, \ldots, n$.

3.3. Comparative Experiment

To validate the effect of the improved C3-Faster on the training results of the yolov5s network model and find the optimal solution for training speed and accuracy, the C3-Faster was respectively replaced in the eight C3 modules of the backbone and neck. The dataset and network training parameters mentioned above were used for training, and the comparative experimental results of the different replacement positions of the C3-Faster module are shown in Table 2.

Table 2. Comparative experiment of the C3-Faster module with the different replacement positions.

Experimental Number	mAP (%)	Precision (%)	Training Time (h)	Inference Time (ms)
1	88.9	89.1	**7.112**	**31.3**
2	90.6	88.4	7.422	37.5
3	**93.4**	91.8	7.301	34.0
4	91.5	**93.3**	7.212	34.5
5	**95.1**	**93.7**	7.310	35.1
6	91.3	92.3	7.315	34.7
7	91.5	89.4	7.209	33.9
8	92.0	91.1	**7.123**	**33.8**
9	93.4	93.3	7.435	34.5

Experiments 1, 2, 3, and 4 in Table 2 respectively replaced C3-Faster in the four C3 modules of the backbone. Experiment 1 replaced the first C3 module, Experiment 2 replaced the second, and so on. Experiments 5, 6, 7, and 8 respectively replaced C3-Faster in the four C3 modules of the neck. Experiment 5 replaced the first C3 module, Experiment 6 replaced the second, and so on. Experiment 9 was the original YOLOv5 model without any modifications. From the experimental results, it could be seen that the training time of

the first four experiments was reduced compared to that of the original YOLOv5. When C3-Faster replaced the first C3 module, the training time was the lowest. However, the highest mAP was achieved when replacing the third C3 module, and the highest precision was achieved when replacing the fourth C3 module. To improve the training speed while maintaining the training accuracy, it was decided to replace the third and fourth C3 modules of the backbone with C3-Faster. Similarly, after analyzing experiments 5, 6, 7, and 8, it was decided to replace the first and fourth C3 modules of the neck with C3-Faster as they performed better.

In order to compare the differences between GIoU, DIoU, CIoU, WIoUv1, WIoUv2, and WIoUv3, in terms of helping to optimize model parameters and improve model accuracy, a control experiment was set up. The YOLOv5s model was used to conduct experiments on the six loss functions using the aforementioned dataset and training parameters. The results of the different loss function comparison experiments are shown in Table 3.

Table 3. Comparative experimental results of the different loss functions.

Network Architecture	mAP (%)	Precision (%)	Training Time (h)
YOLOv5s + GIoU	91.2	91.1	7.354
YOLO5s + DIoU	92.7	89.3	7.621
YOLO5s + CIoU	93.4	93.3	7.435
YOLO5s + WIoUv1	94.6	**92.9**	7.401
YOLO5s + WIoUv2	95.0	92.1	7.231
YOLO5s + WIoUv3	**96.0**	92.2	**7.136**

Analysis of the experimental results showed that using the WIoU loss function led to improvements in both the mAP and precision while also further reducing training time. Among the different variants of the WIoU loss function, the WIoUv1 loss function performed best in terms of precision, while the WIoUv3 loss function performed best in terms of mAP and training time. The difference in precision between the WIoUv1 and WIoUv3 loss functions was not significant. To achieve a lightweight network and improve the speed and accuracy of car tire character recognition, the WIoUv3 loss function was used instead of the original CIoU loss function.

3.4. Ablation Experiment

To validate the performance improvement brought by the decoupled head, C3-Faster, and WIOU in the YOLOv5 network, ablation experiments were conducted. A total of five experiments were set up, including the original YOLOv5s network, YOLOv5s with a decoupled head and C3-Faster, YOLOv5s with a decoupled head and WIOU, YOLOv5s with C3-Faster and WIOU, and YOLOv5s with a decoupled head, C3-Faster, and WIOU. These experiments were conducted on the same device with the same parameters. The results of the ablation experiments for the different modules of the improved YOLOv5s are shown in Table 4.

Table 4. Ablation experimental results of the different modules for improving YOLOv5s.

Network Architecture	mAP (%)	Precision (%)	Training Time (h)
YOLOv5s	93.4	93.3	7.435
YOLOv5s + Decoupled Head	95.6	93.9	7.510
YOLOv5s + Decoupled Head + C3-Faster	94.9	94.3	7.411
YOLOv5s + Decoupled Head + WIOU	93.8	95.0	7.510
YOLOv5s + C3-Faster + WIOU	95.5	94.4	**7.281**
YOLOv5s + Decoupled Head + C3-Faster + WIOU	**97.1**	**95.4**	7.341

On the car tire dataset, the improved YOLOv5s network outperformed the original network model in the ablation experiment. The mAP of the improved YOLOv5s network was 97.1%, and the precision was 95.4%. Compared with the original model, the training

time did not change much, but the other indicators were improved. The mAP increased by 3.7 percentage points, and the precision increased by 2.1 percentage points.

3.5. Comparison with Different Models

To further evaluate the model in this study, performance comparisons were conducted using the improved YOLOv5s model along with YOLOx, YOLOv7, YOLOv5s, and the largest model in YOLOv5, YOLOv5x. The comparative results of the different models are shown in Table 5.

Table 5. Comparison results of the different models.

| Network Architecture	mAP (%)	Precision (%)	Training Time (h)
YOLOx	90.1	84.0	8.011
YOLOv7	94.6	**95.6**	7.923
YOLOv5x	**97.3**	94.6	8.502
YOLOv5s	93.4	93.3	7.435
Improved YOLOv5s	97.1	95.4	**7.341**

From Table 5, it can be observed that among the tested models, YOLOv5x, which had the largest number of residual structures of the YOLOv5 models, exhibited enhanced feature extraction and fusion capabilities, resulting in improved detection accuracy and higher network precision. The test results showed the highest mAP value of 97.3% for YOLOv5x. However, it also required a longer training time. The YOLOv7 network demonstrated the highest precision value (95.6%), surpassing the improved YOLOv5s (95.3%). However, it fell behind in terms of mAP and training time compared to the improved YOLOv5s network. The improved YOLOv5s network had the shortest training time. Although its mAP and precision values were slightly lower than those of the YOLOv5x and YOLOv7 networks, its overall performance was better due to the shorter training time, making it more suitable for tire character recognition. Overall, the model proposed in this paper, after improvements, performed slightly poorer in terms of mAP and precision compared to the YOLOv5x and YOLOv7 networks. However, it was better than other networks in terms of training time, and its overall performance was better, which met the requirements of tire character recognition.

4. Method Validation

Further validation was performed using the original YOLOv5 network and the improved network proposed in this paper. The network models obtained from the ablation experiments were tested using the same set of images, and the test results are shown in Table 6. Among them, the improved YOLOv5 network increased the detection accuracy by 2.9%, the FPS (frame rate per second) increased by 5.3 f/s, the detection speed was faster, and the accuracy was improved. A comparison of the results is shown in Figure 9. In (a) and (c), the original YOLOv5 network was used, while in (b) and (d), the improved YOLOv5 network was used. By comparing the results, it was found that the improved network achieved higher confidence, improved accuracy, and better recognition performance when tested on the same set of images. This achieved a balance between speed and accuracy for car tire character recognition.

Table 6. Test results.

Network Architecture	Accuracy (%)	FPS (h)
YOLOv5s	94.4	30.4
Improved YOLOv5s	**97.3**	**35.7**
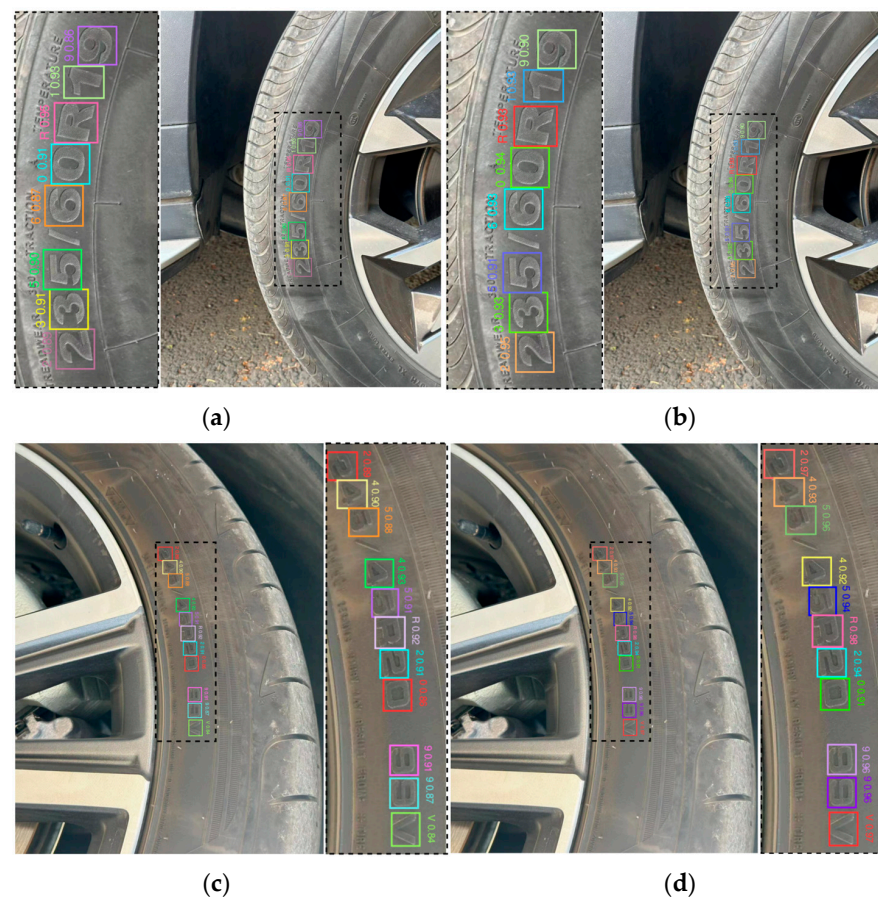
Figure 9. Comparison of the results. In (**a**,**c**), the original YOLOv5 network was used, while in (**b**,**d**), the improved YOLOv5 network was used.

5. Conclusions

This study aimed to improve the efficiency and accuracy of recognizing tire specification characters for automobiles. Based on the YOLOv5s network model, a decoupled head was used to separate the classification and positioning branches, which improved the robustness and accuracy of the model and reduced the occurrence of overfitting. The C3-Faster module was proposed by improving the FasterNet Block to replace some C3 modules in the original backbone and head, reducing the number of convolution operations, speeding up the calculation of the parameters, and further reducing memory usage. Finally, WIoU-Loss was introduced to improve the regression loss function, and a gradient gain allocation strategy was used to reduce the harmful gradients produced by low-quality anchor boxes and further improve the network's running speed.

Through the comparative experiments that replaced the different C3 modules with the C3-Faster module in different positions in the backbone, as well as the first and fourth C3 modules in the head with the C3-Faster module, and through the comparison of different loss functions, the WIoUv3 loss was used instead of the CIoU loss that was in the original YOLOv5. In the ablation experiments, the improved YOLOv5s network outperformed other network models, with a mAP improvement of 3.7 percentage points and a precision improvement of 2.1 percentage points. In the final method validation, the effectiveness of the proposed method was confirmed, which met the practical application requirements.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Qin, Z.H.; Song, Y.T.; Li, M.; Li, C.M.; Zhang, Z.W. Research on Quality problems and countermeasures of Zhang Zhengyin Automobile Tire. *Plast. Technol. Equip.* **2022**, *48*, 25–27.
2. Wang, H.N.; Zhang, X.Q.; Guo, Y.K.; Li, W.Q. Character recognition of tire rubber based on machine vision. *J. Electron. Meas. Instrum.* **2021**, *35*, 191–199.
3. Peng, Q.; Tu, L.F.; Song, W.; Don, G.; Yu, Z.Y. On-line character recognition system of steel plate blank spray label based on image sensor. *Instrum. Technol. Sens.* **2022**, *479*, 57–61.
4. Chen, S.X.; Liu, W.; Wan, S.X. Detection and recognition of multi-line characters on clutch flywheel based on machine vision. *Comb. Mach. Tool Autom. Process. Technol.* **2022**, *581*, 127–129.
5. Sun, C.J.; Jiang, L.C.; Chen, G.L.; Wang, H. Information Recognition Method of Industrial Machine Tool Based on Machine Vision. *Mech. Des. Res.* **2022**, *38*, 78–84.
6. Yang, G.H.; Tang, W.W.; Dai, Z.C.; Wei, J.L. Chip Character Recognition System Based on Machine Vision. *Electron. Meas. Technol.* **2022**, *45*, 105–110.
7. Zhao, Z.; Yang, X.; Zhou, Y.; Sun, Q.; Ge, Z.; Liu, D. Real-Time Detection of Particleboard Surface Defects Based on Improved YOLOV5 Target Detection. *Sci. Rep.* **2021**, *11*, 1–15. [CrossRef]
8. Thuan, D. Evolution of Yolo Algorithm and Yolov5: The State-of-the-Art Object Detention Algorithm. Available online: http://www.theseus.fi/handle/10024/452552 (accessed on 18 April 2023).
9. Gong, P.H. A stamped character recognition method based on YOLOv5 algorithm. *J. Ordnance Equip. Eng.* **2022**, *43*, 101–105.
10. Zhang, J.K.; Liang, Y.; Zhou, Y.H.; Chai, Y.F. Identification algorithm of ladle transport vehicle number and tank number for precise location. *Electron. Meas. Tech.* **2022**, *45*, 162–170.
11. Laroca, R.; Severo, E.; Zanlorensi, L.A.; Oliveira, L.S.; Gonçalves, G.R.; Schwartz, W.R.; Menotti, D. A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–10.
12. Benjumea, A.; Teeti, I.; Cuzzolin, F.; Bradley, A. YOLO-Z: Improving Small Object Detection in YOLOv5 for Autonomous Vehicles. Available online: https://arxiv.53yu.com/abs/2112.11798v4 (accessed on 18 April 2023).
13. Jiang, L.; Liu, H.; Zhu, H.; Zhang, G. Improved YOLO v5 with Balanced Feature Pyramid and Attention Module for Traffic Sign Detection. *MATEC Web Conf.* **2022**, *355*, 03023. [CrossRef]
14. Malta, A.; Mendes, M.; Farinha, T. Augmented Reality Maintenance Assistant Using YOLOv5. *Appl. Sci.* **2021**, *11*, 4758. [CrossRef]
15. Yao, J.; Fan, X.; Li, B.; Qin, W. Adverse Weather Target Detection Algorithm Based on Adaptive Color Levels and Improved YOLOv5. *Sensors* **2022**, *22*, 8577. [CrossRef] [PubMed]
16. Zhang, X.; Fan, H.; Zhu, H.; Huang, X.; Wu, T.; Zhou, H. Improvement of YOLOV5 Model Based on the Structure of Multiscale Domain Adaptive Network for Crowdscape. In Proceedings of the 2021 IEEE 7th International Conference on Cloud Computing and Intelligent Systems (CCIS), Xi'an, China, 7–8 November 2021; pp. 171–175.
17. Zhang, H.; Tian, M.; Shao, G.; Cheng, J.; Liu, J. Target Detection of Forward-Looking Sonar Image Based on Improved YOLOv5. *IEEE Access* **2022**, *10*, 18023–18034. [CrossRef]
18. Xie, C.H.; Wu, J.M.; Xu, H.Y. The small target detection algorithm of YOLOv5 UAV image is improved. *Comput. Eng. Appl.* **2023**, *59*, 198–206.
19. Wang, J.B.; Wu, Y.X. Improved YOLOv4-tiny helmet wearing detection algorithm. *Comput. Eng. Appl.* **2023**, *59*, 183–190.
20. Yuan, S.; Du, Y.; Liu, M.; Yue, S.; Li, B.; Zhang, H. YOLOv5-Ytiny: A Miniature Aggregate Detection and Classification Model. *Electronics* **2022**, *11*, 1743. [CrossRef]
21. Chen, J.; Kao, S.; He, H.; Zhuo, W.; Wen, S.; Lee, C.-H.; Chan, S.-H.G. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. *arXiv* **2023**, arXiv:2303.03667.
22. Guo, S.; Li, L.; Guo, T.; Cao, Y.; Li, Y. Research on Mask-Wearing Detection Algorithm Based on Improved YOLOv5. *Sensors* **2022**, *22*, 4933. [CrossRef]
23. Qian, X.; Zhang, N.; Wang, W. Smooth GIoU Loss for Oriented Object Detection in Remote Sensing Images. *Remote Sens.* **2023**, *15*, 1259. [CrossRef]

24. Kong, L.; Wang, J.; Zhao, P. YOLO-G: A Lightweight Network Model for Improving the Performance of Military Targets Detection. *IEEE Access* **2022**, *10*, 55546–55564. [CrossRef]
25. Gao, J.; Chen, Y.; Wei, Y.; Li, J. Detection of Specific Building in Remote Sensing Images Using a Novel YOLO-S-CIOU Model. Case: Gas Station Identification. *Sensors* **2021**, *21*, 1375. [CrossRef] [PubMed]
26. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. *arXiv* **2023**, arXiv:2301.10051.