engineering proceedings

MDPI

Proceeding Paper

GEB-YOLO: Optimized YOLOv7 Model for Surface Defect Detection on Aluminum Profiles †

Zihao Xu *, Jinran Hu, Xingyi Xiao and Yujian Xu

College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou 325035, China; jnranny@163.com (J.H.); wzuxxy@163.com (X.X.); wzuxyj@163.com (Y.X.)

* Correspondence: zhxu@stu.wzu.edu.cn

† Presented at the 4th International Conference on Advances in Mechanical Engineering (ICAME-24), Islamabad, Pakistan, 8 August 2024.

Abstract: In recent years, achieving high-precision and high-speed target detection of surface defects on aluminum profiles to meet the requirements of industrial applications has been challenging. In this paper, the GEB-YOLO is proposed based on the YOLOv7 algorithm. First, the global attention mechanism (GAM) is introduced, highlighting defect features. Second, the Explicit Visual Center Block (EVCBlock) is integrated into the network for key information extraction. Meanwhile, the BiFPN network structure is adopted to enhance feature fusion. The ablation experiments have demonstrated that the defect detection accuracy of the GEB-YOLO model is improved by 6.3%, and the speed is increased by 15% compared to the YOLOv7 model.

Keywords: target detection; surface defects; GEB-YOLO; attention mechanism; BiFPN

1. Introduction

Aluminum profiles are an important material widely used in industrial production. However, during the production and transportation of aluminum profiles, various surface defects, such as oxidation, deformation, scratches, etc., often occur, which seriously affect the overall quality and aesthetics of the product. For metal parts, detecting surface defects is significant for quality control. Detecting metal surface defects in an early stage can prevent losses in subsequent processing, testing, and other links, reducing production costs.

Traditional metal surface defect detection still uses manual inspection methods, which have the problems of low efficiency, high labor cost consumption, and inability to meet the needs of mass production. In addition, there are inspection technologies such as turbine inspection, optical inspection, and magnetic particle inspection, all of which have their own shortcomings. Ramirez-Pacheco et al. [1] used the eddy current testing technique with a giant magneto-resistive (GMR) sensor to identify surface defects on aluminum plates; however, the technique has low sensitivity for detecting small defects on metal surfaces. Additionally, optical inspection is not applicable in all environments due to its susceptibility to lighting conditions and reflective surfaces, and magnetic particle inspection is ineffective in detecting non-ferromagnetic materials [2].

Deep learning-based target detection algorithms are becoming more prevalent in metal surface defect detection. Single-stage and two-stage target detection algorithms are the two main types of target detection algorithms [3]. The two-stage algorithm evolves from R-CNN [4] (regions with convolutional neural network features), Fast-RCNN [5] to Faster-RCNN [6], with an additional region generation stage than the one-stage algorithm. Wei et al. [7] used Faster R-CNN for defect detection of track fasteners, which has high detection accuracy, but the model parameters are large. Moreover, the track features are relatively easy to classify, so it is difficult to be applied to complex surface defect problems. Hu et al. [8] proposed a new network architecture of Faster RCNN combined with ResNet50

check for updates

Citation: Xu, Z.; Hu, J.; Xiao, X.; Xu, Y. GEB-YOLO: Optimized YOLOv7 Model for Surface Defect Detection on Aluminum Profiles. *Eng. Proc.* **2024**, *75*, 28. https://doi.org/10.3390/engproc2024075028

Academic Editors: Muhammad Mahabat Khan, Muhammad Irfan, Mohammad Javed Hyder and Manzar Masud

Published: 25 September 2024

Copyright: © 2024 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/).

for detecting defects in printed circuit boards (PCBs), but its model parameters are too large, and the detection time is slow.

One-stage algorithms, on the other hand, perform bounding box regression and target categorization immediately on the image, with faster processing time and lower complexity. Its development has gone through YOLO (You Only Look Once) [9] to SSD (Single Shot Multibox Detector) [10] to YOLOv2 [11], YOLOv3 [12], etc. The YOLO series has been continuously updated and optimized to exhibit increasing efficiency and accuracy. Yuan et al. [13] used the YOLO model combined with Ghost convolution to achieve steel surface defects, which realizes lightweight and improves the response speed, but the accuracy of the detection is far from practical application. Wang et al. [14] employed the C2f-DSConv module and MPDIoU loss function in their enhanced YOLOv8n algorithm for the detection of surface defects on industrial aluminum plates to improve the detection accuracy. However, the model is so complex that the detection speed is sacrificed. In order to detect surface defects on strip steel, Wang et al. [15] proposed a YOLOv7-based model. By introducing the ASPP_CA architecture and involution mechanism, this model realizes the improvement of inference speed and small target detection capability, but its accuracy still struggles to meet the requirements of metal defect detection.

Therefore, to address the issues of low accuracy and slow speed of the target detection model in the detection of metal surface defects, this paper proposes an optimized detection model, GEB-YOLO. The GEB-YOLO is designed to increase the efficiency of feature extraction as well as the accuracy of defect detection in complex situations.

2. Improved Surface Defect Detection Model

2.1. GEB-YOLO Model

Based on the YOLOv7 target detection model, this paper proposes an improved model called GEB-YOLO to handle the metal surface defect detection task more accurately and efficiently. Figure 1 depicts its structure. The following are the strategies for improvement: the two convolutional modules in Backbone are replaced by the Global attention mechanism (GAM) [16] modules to suppress the interference of irrelevant information and extract important information; the EVCBlock module is embedded in the Neck to improve the learning of key features; and a cross-scale connecting line is added based on the BiFPN structure in order to realize more feature fusion.

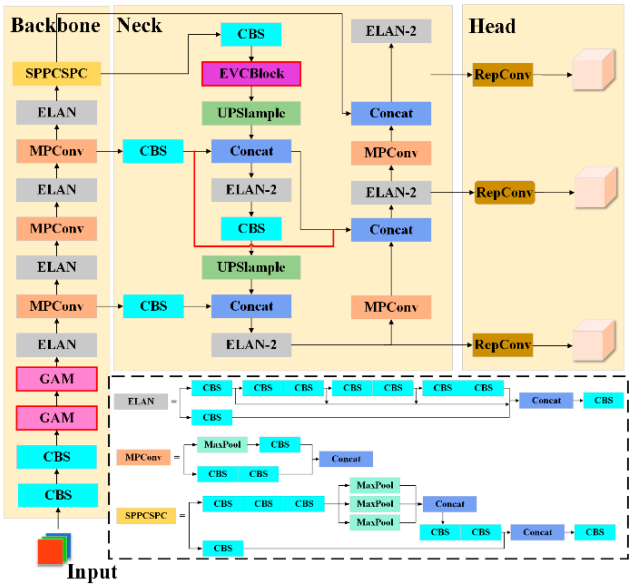


Figure 1. The GEB-YOLO Network Diagram.

2.2. YOLOv7 Model

GEB-YOLO is improved based on the YOLOv7 model. Three components make up the fundamental structure of YOLOv7: the backbone, neck, and head. It is shown below [3].

Backbone: The backbone mainly consists of the CBS and ELAN modules. The CBS consists of a convolutional layer, a normalization layer, and a Silu activation function layer. The ELAN employs a denser residual structure, which is an efficient layer aggregation network structure.

Neck: The neck consists of the SPPCSPC module and PAFPN structure composition in addition to the convolutional layer, aiming at capturing different scales of information about the target and realizing multi-scale information fusion.

Head: The head section uses RepConv to adjust the channel's output characteristics. The header contains several detection heads for detecting targets with different scales and aspect ratios.

2.3. GAM Module

The Global attention mechanism (GAM) module, shown in Figure 2, is a lightweight attention mechanism. The mechanism employs the channel attention submodule and the spatial attention submodule from the CBAM attention mechanism in order to extract relevant information in the desired part. The channel attention submodule stores three-dimensional data in a 3D layout and utilizes a multilayer perceptron to enhance the multidimensional channel-space correlation. Meanwhile, the spatial attention sub-module fuses spatial information through two convolutional layers. Ultimately, the purpose of minimizing information loss and enhancing global features is achieved, which significantly improves the performance of deep neural networks. It is worth mentioning that this module also helps to balance recognition speed and accuracy to achieve a superior level of recognition.

The realization process of the attention mechanism is expressed as Equations (1) and (2). The intermediate outcome F_2 and output feature F_3 are represented as follows according to the input feature F_1:

$$F_2 = M_C(F_1) \otimes F_1 \quad (1)$$

$$F_3 = M_S(F_2) \otimes F_2 \quad (2)$$

where M_C and M_S stand for channel and spatial attention map, respectively, and $\otimes$ indicates the multiplying of individual elements.

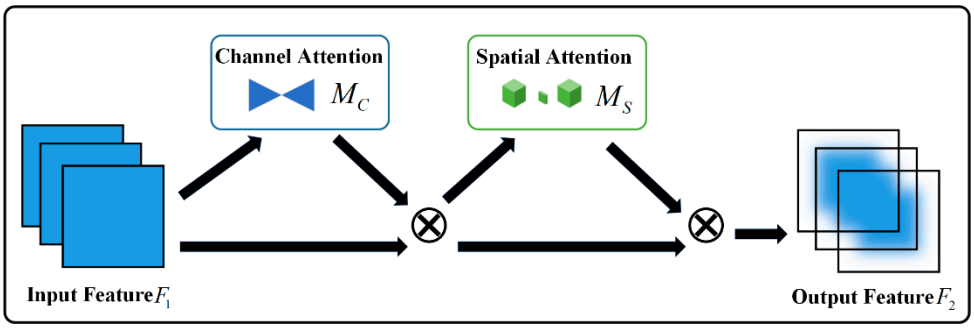


Figure 2. Structure of GAM module.

In this paper, the GAM module is integrated into the YOLOv7 backbone network, which facilitates the extraction of key defect features through global dimensional feature interactions, thus improving the accuracy of detection.

2.4. EVCBlock

As shown in Figure 3, EVCBlock (Explicit Visual Center Block) [17] is an enhanced convolutional vision block structure with two main parts. One of them is the Lightweight

MLP, which is mainly composed of a module for deep convolution and a module based on channel MLP. The deep convolution is processed in the module with deep convolutional layers for channel scaling and DropPath and finally for residual concatenation. Lightweight MLP improves the feature representation capability while using fewer parameters and computations than traditional MLP.

The other is Learnable Visual Center (LVC), which mainly consists of inherent Codebook and scaling factor. For the operation process of LVC, the encoded features are first encoded through a set of convolutional layers and then processed through a CBR block to input them into the Codebook. A set of scaling factors is used in the Codebook to continuously map the corresponding positional information so that complete information about each pixel point is computed. Subsequently, all the information is fused and fed to a fully connected layer and a 1×1 convolutional layer to predict key features. Ultimately, a good balance between efficiency and expressiveness is achieved by combining Lightweight MLP and LVC in parallel.

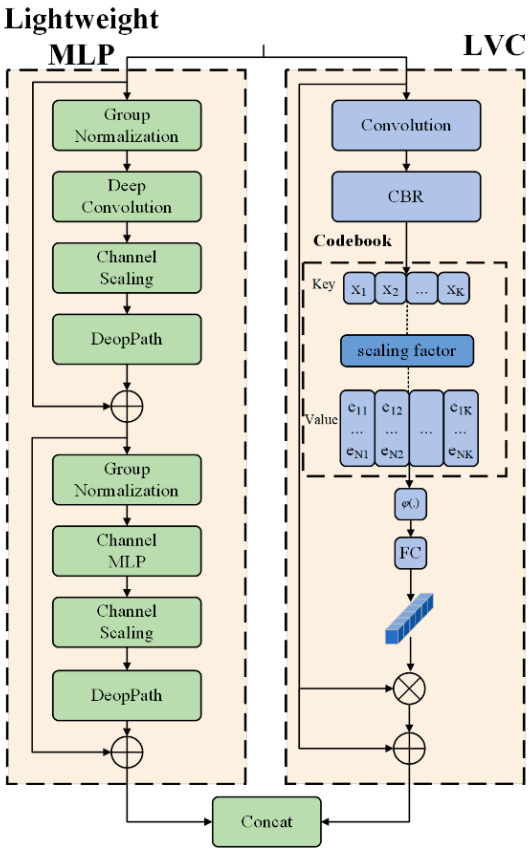


Figure 3. The structure of EVCBlock.

By embedding EVCBlock into the YOLO model, it can adaptively capture different defect characteristics and direct the model to focus on more important visual information, thus improving the overall performance.

3. BiFPN

The design concept of FPN utilizes top-down information transfer and lateral connections to achieve cross-layer feature fusion, thus constructing a feature pyramid structure. However, FPN can only realize top-down unidirectional information flow. For this reason, YOLOv7 introduces the path aggregation network (PAN), which employs bottom-up aggregated paths to improve the accuracy for targets of different sizes, as shown in Figure 4. However, this approach increases the parameters and computational complexity. Therefore, we choose the network structure of BiFPN to achieve path enhancement through effective

cross-scale interconnection and feature fusion. Specifically, in addition to the PAN's top-down and bottom-up sequential feature fusion characteristics, the BiFPN module improves on top of the PAN. The BiFPN is improved by removing intermediate nodes within the top and bottom edges and adding residual connections by skipping over the removed nodes. We follow the BiFPN structure and add an extra cross-scale connectivity channel across the input and output nodes in the middle layer to realize cross-scale connections [18]; we remove the intermediate nodes on both edges and keep only the nodes between C4 and P4 to fuse more features without increasing the computation. Eventually, more features are integrated by BiFPN, which can realize high-precision recognition of targets of different sizes.

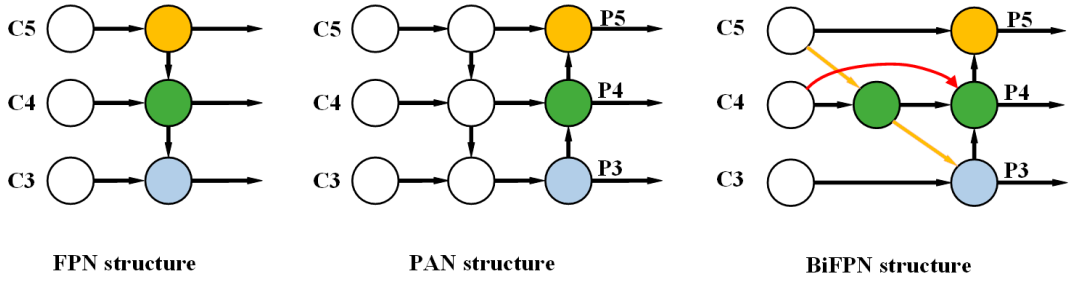


Figure 4. FPN, PAN, and BiFPN structures.

4. Experiment and Verification

4.1. Experimental Dataset and Environment

The data used in this paper is a dataset of industrial defects detection on the surface of an aluminum sheet captured using a Hikvision industrial camera (model MV-CS050 10GC-PRO, manufactured by Hangzhou Hikrobot Co., LTD. in Hangzhu, China). The dataset contains defect targets in four categories: fold, abrasion, dirt, and pinhole. It contains more than 1000 labeled images and is separated into training and validation sets at an 8:2 ratio. Table 1 lists the specific hardware and software environments used for model training.

Table 1. Experiment environment.

Configuration	Version
Operating System	Window 11
CPU	AMD Ryzen 9 5900X 12-Core Processor 3.70 GHz
GPU	NVIDIA GeForce GTX 3090
PyTorch	2.0.0
Cuda	11.7
Python	3.8

4.2. Evaluation Metrics

To validate the model performance, we choose Precision, Recall, *mAP*, and frames per second (FPS) as the experimental evaluation metrics [19]. Precision indicates the correctness of the model prediction, i.e., how many targets are correctly categorized. Recall is the number of targets successfully detected in the presence of positives. When Recall is 1, it means that there are no missed detections.

Precision and Recall are calculated as shown below:

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

where *TP* (true positives) indicates counts of correctly detected required targets, *FP* (false positives) indicates counts of undetected required targets, and *FN* (false negatives) indicates counts of incorrectly detected unwanted targets.

The *AP* value is calculated from the area under the precision and recall curves, which indicates the detection accuracy of each category, and its average value can be obtained as the mean average precision (*mAP*). The specific calculation is shown below:

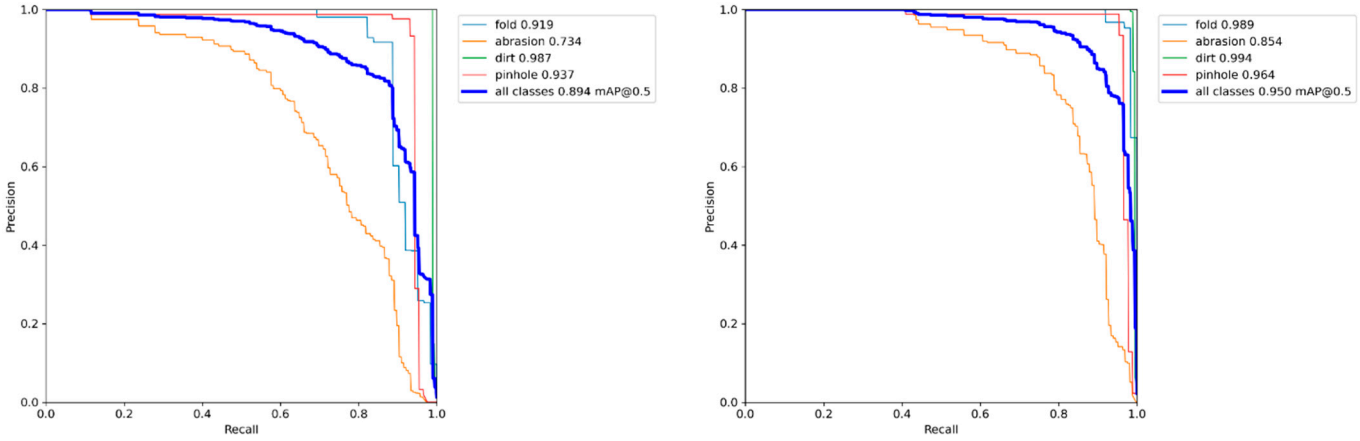
$$AP = \int_0^1 p(r)d(r) \tag{5}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{6}$$

where N is the total number of categories, and $N = 4$, in this paper.

4.3. Network Training

To compare the change in model performance before and after the improvement, we performed 150 rounds of training on the same dataset for both YOLOv7 and GEB-YOLO networks. The metrics stabilize when 100 rounds of experiments are conducted, and the model training normally converges. The final Precision-Recall curve is shown in Figure 5. As can be seen from the curves, the P-R curve of GEB-YOLO is closer to the upper right corner compared to the basic YOLOv7 model, indicating better overall performance. In addition, the numerical enhancement is larger for the features 'fold' and 'abrasion', while the values of 'dirt' and 'pinhole' also have slightly improved values, indicating that GEB-YOLO is more accurate in defect recognition.



(**a**) (**b**)

Figure 5. Precision-Recall curve comparison: (**a**) training results of YOLOv7; (**b**) training results of GEB-YOLO.

When the detection graphs of YOLOv7 are compared with those of the optimized model, the enhancement results of the optimized model can be observed more intuitively, as presented in Figure 6. It can be seen that YOLOv7 shows some missed defects, while GEB-YOLO can basically accurately recognize all defect features, including small scratches and inconspicuous folds.

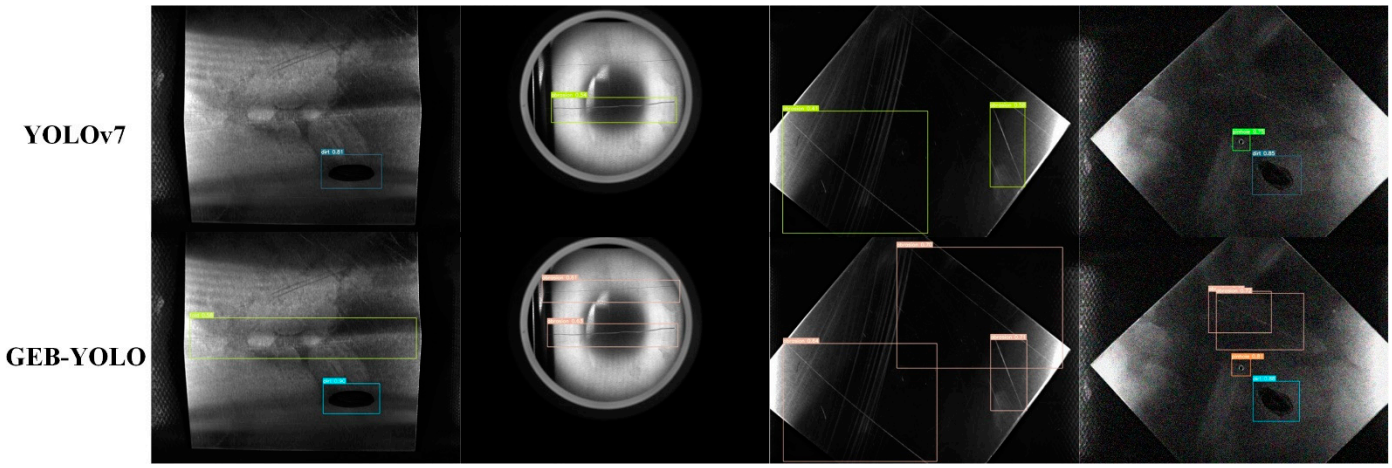


Figure 6. Comparison of detection graphs.

4.4. Ablation Experiments

In this study, the YOLOv7 algorithm was comprehensively improved, and the improvement effect was verified by ablation experiments (see Table 2 for details). The data showed that the YOLOv7 algorithm improved in terms of *mAP*, Precision, Recall, and FPS.

In addition, the implementation of the attention mechanism GAM and EVCBlock modules, respectively, improved the *mAP* by 0.9% and 0.4%, which suggests that GAM and EVCBlock play a positive role in the attention to defective features. After the combined introduction of GAM, EVCBlock, and BiFPN structures, the new model GEB-YOLO achieves an *mAP* of 95.03%, which is 6.3% higher than that before the improvement; its Precision exceeds 95%, and Recall exceeds 91%, which proves its strong detection ability.

The improved model also improves by 15% in terms of FPS, which shows that the integrated improvement strategy significantly improves detection speed.

Table 2. Results of ablation experiments.

	GAM	EVCBlock	BiFPN	mAP/%	Precision/%	Recall/%	FPS/fps
1				89.41	90.84	83.66	60
2	✓			90.21	90.81	84.60	65
3		✓		89.78	89.62	86.22	61
4			✓	92.96	92.99	88.16	60
5	✓	✓		90.77	91.48	87.56	67
6		✓	✓	92.15	93.27	87.66	63
7	✓		✓	94.15	95.18	91.09	67
8	✓	✓	✓	95.03	95.39	92.44	69

5. Conclusions and Future Work

This paper aims to address the problems of low accuracy and complex modeling of deep learning algorithms for metal surface defect detection. It performs a comprehensive improvement based on the YOLOv7 algorithm to meet the detection requirements.

By introducing GAM, EVCBlock, and BiFPN structures, the defect detection model of GEB-YOLO is proposed. The training and validation of the collected industrial defects dataset on the aluminum sheet surface yielded that the model's *mAP* reaches 95.03%, which is improved by 6.3% with respect to the original YOLOv7. Its Precision is over 95%, and its Recall is over 91%. By analyzing the detection graphs, the improved model of this paper can also accurately identify the defects that are not obvious. In conclusion, the new model GEB-YOLO achieves efficient detection of steel surface defects and higher precision feature extraction of aluminum surface defects.

However, the shortcoming is that this paper is only validated for surface defects of aluminum profiles, which does not prove the model's universality. The validation of other metal surface defects and different defect types must be carried out in future research.

Author Contributions: Conceptualization, Z.X.; methodology, Z.X.; software, Z.X.; validation, Z.X.; formal analysis, J.H.; investigation, X.X.; resources, Y.X.; writing, Z.X. and J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available in this manuscript.

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

References

1. Ramirez-Pacheco, E.; Espina-Hernandez, J.H.; Caleyo, F.; Hallen, J.M. Defect detection in aluminium with an eddy currents sensor. In Proceedings of the 2010 IEEE Electronics, Robotics and Automotive Mechanics Conference, Cuernavaca, Mexico, 28 September–1 October 2010.
2. Gupta, M.; Khan, M.A.; Butola, R.; Singari, R.M. Advances in applications of Non-Destructive Testing (NDT): A review. *Adv. Mater. Process. Technol.* **2022**, *8*, 2286–2307. [CrossRef]
3. Xu, Z.; Meng, Y.; Yin, Z.; Liu, B.; Zhang, Y.; Lin, M. Enhancing autonomous driving through intelligent navigation: A comprehensive improvement approach. *J. King Saud Univ.—Comput. Inf. Sci.* **2024**, *36*, 102108. [CrossRef]
4. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
5. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
6. 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.* **2015**, *39*, 1137–1149. [CrossRef] [PubMed]
7. Wei, X.; Yang, Z.; Liu, Y.; Wei, D.; Jia, L.; Li, Y. Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study. *Eng. Appl. Artif. Intell.* **2019**, *80*, 66–81. [CrossRef]
8. Hu, B.; Wang, J. Detection of PCB surface defects with improved faster-RCNN and feature pyramid network. *IEEE Access* **2020**, *8*, 108335–108345. [CrossRef]
9. 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, Las Vegas, NV, USA, 26 June–1 July 2016.
10. Liu, W.; Anguelov, D.; Erhan, D. Ssd: Single shot multibox detector. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.
11. 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 June 2017.
12. Farhadi, A.; Redmon, J. Yolov3: An incremental improvement. In *Computer Vision and Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2018.
13. Yuan, Z.; Ning, H.; Tang, X.; Yang, Z. GDCP-YOLO: Enhancing steel surface defect detection using lightweight machine learning approach. *Electronics* **2024**, *13*, 1388. [CrossRef]
14. Wang, L.; Zhang, G.; Wang, W.; Chen, J.; Jiang, X.; Yuan, H.; Huang, Z. A defect detection method for industrial aluminum sheet surface based on improved YOLOv8 algorithm. *Front. Phys.* **2024**, *12*, 1419998. [CrossRef]
15. Wang, Z.; Liu, W. Surface defect detection algorithm for strip steel based on improved yolov7 model. *IAENG Int. J. Comput. Sci.* **2024**, *51*, 308–316.
16. Liu, Z.; Li, L.; Fang, X.; Qi, W.; Shen, J.; Zhou, H.; Zhang, Y. Hard-rock tunnel lithology prediction with TBM construction big data using a global-attention-mechanism-based LSTM network. *Autom. Constr.* **2021**, *125*, 103647. [CrossRef]
17. Quan, Y.; Zhang, D.; Zhang, L.; Tang, J. Centralized feature pyramid for object detection. *IEEE Trans. Image Process.* **2023**, *32*, 4341–4354. [CrossRef] [PubMed]

18. He, J.; Wang, Y.; Wang, Y.; Li, R.; Zhang, D.; Zheng, Z. A lightweight road crack detection algorithm based on improved YOLOv7 model. *Signal Image Video Process.* **2024**, *18*, 847–860. [CrossRef]
19. Zhai, X.; Huang, Z.; Li, T.; Liu, H.; Wang, S. YOLO-Drone: An Optimized YOLOv8 Network for Tiny UAV Object Detection. *Electronics* **2023**, *12*, 3664. [CrossRef]

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