Article

Automotive Parts Defect Detection Based on YOLOv7

Hao Huang [1] and Kai Zhu [2],*

1 School of Mechanical Engineering, Jiangsu University of Technology, Changzhou 213000, China; 2022655079@smail.jsut.edu.cn
2 School of Automobile and Traffic Engineering, Jiangsu University of Technology, Changzhou 213000, China
* Correspondence: fatkyo@jsut.edu.cn

Abstract: Various complex defects can occur on the surfaces of small automobile parts during manufacturing. Compared with other datasets, the auto parts defect dataset used in this paper has low detection accuracy due to various defects with large size differences, and traditional target detection algorithms have been proven to be ineffective, which often leads to missing detection or wrong identification. To address these issues, this paper introduces a defect detection algorithm based on YOLOv7. To enhance the detection of small objects and streamline the model, we incorporate the ECA attention mechanism into the network structure's backbone. Considering the small sizes of defect targets on automotive parts and the complexity of their backgrounds, we redesign the neck portion of the model. This redesign includes the integration of the BiFPN feature fusion module to enhance feature fusion, with the aim of minimizing missed detections and false alarms. Additionally, we employ the Alpha-IoU loss function in the prediction phase to enhance the model's accuracy, which is crucial for reducing false detection. The IoU loss function also boosts the model's efficiency at converging. The evaluation of this model utilized the Northeastern University steel dataset and a proprietary dataset and demonstrated that the average accuracy (mAP) of the MBEA-YOLOv7 detection network was 76.2% and 94.1%, respectively. These figures represent improvements of 5.7% and 4.7% over the original YOLOv7 network. Moreover, the detection speed for individual images ranges between 1–2 ms. This enhancement in detection accuracy for small targets does not compromise detection speed, fulfilling the requirements for real-time, dynamic inspection of defects.

Keywords: feature attention; automobile parts; YOLOv7

1. Introduction

At present, the detection of defects in automobile parts is receiving increasing attention [1], with the quality of these parts having a direct effect on the vehicle's overall quality. Low-quality components can cause significant economic losses and even pose risks to life and safety. The presence of ambiguous information and the rising complexity of defects, including minor ones, complicate the detection process [2]. Despite the fact that object detection methods based on deep learning have demonstrated superior performance in defect detection tasks in recent years [3–6], these methods often fall short when faced with complex and smaller targets. These challenges significantly impair the accuracy of detecting defects in automobile parts.

Addressing one of the challenges, this paper focuses on the issue of background noise [7,8]. The complexity of backgrounds in the detection process can lead to confusion due to the mistaking of noise for relevant information [9]. This confusion can cause missed or incorrect detections, reducing the accuracy when identifying defects. Moreover, defects that blend in with the color of the product present an additional challenge [10] that represents a significant barrier to effective detection.

This paper introduces an innovative approach to detecting defects in automotive parts. The main contributions of this study are outlined as follows:

(1) To enhance the detection accuracy of automobile part defects, we propose a new detection network (MEBA-YOLO). This network utilizes a unique fusion and attention mechanism built on YOLOv7.

(2) To achieve exceptional results for detecting defects in automobile parts, we introduce a model incorporating the AlphaIoU loss function. This function significantly increases accuracy for detecting complex and small defects, marking a significant advancement in the field.

(3) Our proposed method offers real-time defect detection on production lines, which aids with the immediate identification of defects in automobile parts. This contribution is crucial for enhancing vehicle safety.

In addition, we have confirmed the effectiveness of our method through comprehensive testing and benchmarking. The results from these tests are backed by solid evidence that indicates that our method outperforms others in terms of accuracy.

The remainder of this article is organized as follows: In Section 2, we explore our analysis of defect detection and the obstacles encountered when identifying small and complex defects. Section 3 details the method we propose. In Section 4, we share the results of our experiments and our analysis of these results. We conclude with a summary of our findings in Section 5.

2. Related Work

To enhance defect detection in automotive parts, extensive studies have been carried out on various methods for detecting object defects. This section offers an overview of some of the most effective strategies for defect detection. We especially emphasize methods based on deep learning [11–16], which have proven to be highly effective in this area.

2.1. Defect Detection

Networks for detecting surface defects with deep learning typically rely on target localization to fulfill their tasks. This method aligns with traditional approaches to defect detection by aiming to accurately identify the locations and types of defects. Currently, networks for detecting defects can generally be classified into two types based on their architecture:

(1) Two-stage networks, represented by Faster R-CNN (Region-CNN) [17];

(2) One-stage networks, represented by SSD (Single Shot Multibox Detector) [13] or YOLO (You Only Look Once) [12].

The first type, referred to as two-stage detection networks (e.g., Faster R-CNN), starts with generating feature maps from the input image using the backbone network. The region proposal network (RPN) calculates the confidence level of the anchor box and identifies the proposal region. Following ROI pooling, the feature maps from the proposal region are fed into the network for initial detection results, which are then enhanced to determine the accurate location and classification of defects. Cha et al. [18] pioneered the application of Faster R-CNN for the localization of defects on bridge surfaces by substituting the backbone network with ZFnet. They achieved a mean average accuracy (mAP) of 87.8% across five categories of bridge construction defects using a dataset of 2366 images measuring 500 × 375 pixels.

In 2020, Tao et al. [19] developed a two-stage Faster R-CNN framework specifically for identifying insulator defects during UAV power inspections. In the first stage, the framework targets the identification of insulator areas in natural environments. Following this, it focuses on detecting defects in these identified insulator regions. Similarly, He et al. [20] introduced an enhanced defect detection system based on Faster R-CNN for analyzing strip steel surfaces. This enhancement involved integrating multilevel feature maps from the backbone architecture into a comprehensive multiscale feature map. Their approach recorded an mAP of 82.3% on the NEU-DET defect detection dataset when utilizing a ResNet-50 backbone. This detection method, leveraging Faster R-CNN, has also found application across various defect detection areas, including tunnels [21], LCD

panel polarizer surfaces [22], thermal imaging for insulator defects [23], aluminum profile surfaces [24], and tire hubs [25].

The second type, single-stage detection networks, is divided into two categories: SSD and YOLO. These methodologies process the entire image as input and directly determine the bounding box's location and category in the output layer. Chen et al. [26] enhanced an SSD network for identifying defects in fasteners on contact network supports: specifically, by employing various feature map layers for detection purposes. Li et al. [27] devised a method based on MobileNetSSD for spotting defects on sealing surfaces of containers on filling production lines. They enhanced the SSD's backbone with MobileNet and streamlined the model's parameters. Zhang et al. [28] adopted the most recent YOLOv3 version for bridge surface defect detection and enhanced the original network with pre-training weights, batch renormalization, and focal loss to increase the detection accuracy.

2.2. Feature Fusion Strategy

The swift progress in computer networks has led to the proposal of several feature fusion methods [29–33]. These methods aim to merge the detailed location information from shallow feature maps with the comprehensive semantic insights from deeper layers. Such integration is designed to enhance the detection of small targets in complex settings.

However, the generation of shallow target features often lacks semantic depth and relies heavily on surrounding context. Li et al. [34] tackled this by incorporating the FPN concept into the SSD framework, thus creating a method for lightweight feature fusion. This approach combines different levels of feature maps to create feature pyramids, enhancing the use of small target feature information.

Shi et al. [35] introduced FFESSD (Single Shot Object Detection with Feature Enhancement and Fusion), which applies a shallow feature enhancement (SFE) module to enhance shallow semantic details and a deep feature enhancement (DFE) module to enrich deep feature mapping with additional input image details.

To advance the improvement of shallow feature details, Pengfei Zhao et al. [36] developed the feature enhancement module (FEM). They combined feature maps from the FEM with those from channel dimensionality reduction. However, this method of combining channels did not consider the interrelationship among channels. To address this oversight, they implemented the efficient channel attention module (ECAM) after merging operations to fully leverage the contextual information of the target features.

2.3. Challenges of Defect Detection in Automotive Parts

The challenge of accurately identifying defects in automotive parts in complex environments remains unresolved and demands further research. To overcome this, several advanced deep learning strategies [37–42] have been employed to enhance the accuracy of defect detection in automotive parts. These strategies include using synthetic automotive part datasets to enlarge training data and enhance method generalization, adjusting the core network or integrating new modules to handle various complex situations, and employing attention mechanisms or post-processing methods to minimize noise and emphasize important features.

The issue of inaccurate localization and categorization is particularly acute in the detection of small targets. When targets are small, image details can become indistinct, complicating the process of identifying object details [43]. This issue is a significant hurdle to the accurate detection of defects in automotive parts.

To achieve accurate identification of defects in automotive parts under complex backgrounds and challenging conditions, we introduce a novel multi-class target detection method named the MBEA-YOLOv7 network. This network incorporates the MBEA structure to enhance accuracy, and we validate its effectiveness specifically for the detection of automotive parts defects. This paper primarily focuses on this method.

3. Method

In the following section, we discuss the details of the MBEA-YOLO method. The rationale behind this method is its ability to outperform traditional classification methods.

Therefore, our proposed method seeks to enhance the accuracy of detection and categorization of defects in automotive parts. In doing so, it confirms the effectiveness and reliability of our approach for boosting the performance of defect detection. The architecture of MBEA-YOLO is illustrated in Figure 1. In the figure, we can see that in the neck part we have added BIFPN and ECA attention mechanisms. In the model, four BIPFN modules are added to improve the feature fusion ability of the model, and an ECA attention mechanism is added to enhance the detection ability for small target defects. Finally, for the loss function part of the model, the original CIoU is replaced with Alpha-IoU, which can improve the robustness of the model. These changes to the model are described in detail later.
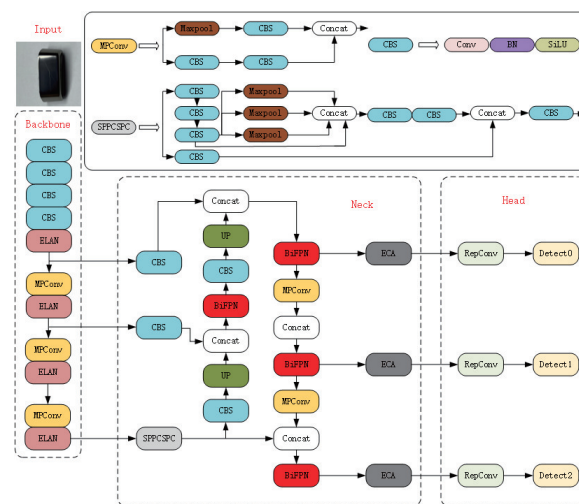

Figure 1. Structure of MBEA-YOLOv7.

3.1. BiFPN-Based Feature Fusion Network

As outlined in Section 2, to tackle the problem of imbalanced feature and semantic information, we use a feature fusion network to enhance our original network. The original version, YOLOv7, employs a PANet feature fusion network for extracting features at various levels. While it is capable of adaptive feature pooling and comprehensive fusion, it faces challenges in efficiently processing images of different resolutions due to its uniform approach to up-sampling and down-sampling in feature fusion.

In this paper, we enhance the bidirectional feature fusion pyramid with a bidirectional feature pyramid network (BiFPN). In the process of feature fusion of auto parts defect detection, different from PANet, PANet adopts a unified feature sampling method, which will lead to deviations of features of different sizes in the process of feature fusion, which will lead to the loss of small defect features. The BiFPN proposed in this paper can effectively solve this problem. According to the different targets of up-sampling and down-sampling, the features of auto parts defects can be effectively fused, combined, and transferred according to the different dimensions of defect features to adapt to the input of different auto parts defect features. Therefore, it can greatly enrich the details of auto parts defects, minimizes the occurrence of missing detection and false positives for small targets, and improves the overall accuracy of the model.

BiFPN enhances feature fusion at a higher level than PANet. It removes nodes with only one input edge (for example, the first node in Figure 2) because removing a node without feature fusion simplifies the network structure. In addition, it introduces additional edges that connect input and output nodes. This change not only simplifies the bidirectional network but also allows other features to be incorporated without significantly increasing

computational costs. In BiFPN, each bidirectional path (top-down and bottom-up) is considered to be a layer of the feature network and is replicated multiple times to achieve high-level feature fusion, and BiFPN gives different weights to the features of different inputs so that the fusions of different input features are differentiated.
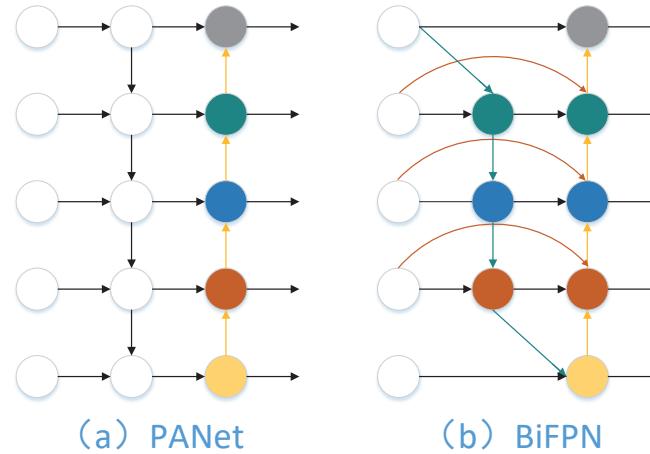

(a) PANet   (b) BiFPN

Figure 2. Structures of PANet and BiFPN.

Simultaneously, the integrated feature maps are easily discernible. Depending on the resolution of the input defective feature map, the contribution to the combined feature map also varies. Building upon the BiFPN concept, we have incorporated cross-scale connections into the feature fusion process to enhance model accuracy for anomaly detection. This approach is particularly beneficial for effectively integrating data across various scales and resolutions and thereby improves model detection accuracy.

3.2. Attention Mechanism

The role of the attention mechanism in deep learning models is critical. It isolates a small yet crucial portion of data from a large dataset and concentrates solely on this information. Considering the broad range of defect sizes in automotive parts, which results in inconsistent image defect scales, incorporating the attention mechanism enhances the model's ability to represent data. This adaptation allows the model to pay closer attention to defective areas, thus elevating the accuracy of defect identification. It presents an intelligent solution to the challenge presented due to the varying sizes of defects in automotive components.

The ECA attention mechanism, representing an advancement over SENet, is illustrated in Figure 3. SENet reduces dimensionality to manage nonlinear cross-channel interactions and minimize model complexity. This method, however, affects channel attention prediction and is insufficient in capturing the full spectrum of inter-channel dependencies. In contrast, the ECA attention module circumvents dimensionality reduction and employs one-dimensional convolution to effectively facilitate local cross-channel interactions. This process extracts the dependencies between channels, thereby efficiently capturing their interactions. When a neck module with an ECA attention mechanism is added, the defect features input into this module are convolution in one dimension so as to realize the interdependence of features of different auto parts and to cause effective interaction with context features, which can effectively enhance the recognition ability for objects with different sizes. The simplicity of the ECA mechanism in both concept and execution minimally affects the processing speed of the network, guaranteeing accuracy and efficiency when detecting automotive parts.

$$C = \phi(k) = 2^{(y \neq k - b)} \tag{1}$$

In terms of practical application, ECA first compresses the automotive parts image through global average pooling (GAP). This process averages the feature maps on each

channel of the auto parts image to produce a global feature vector through which the overall context information of the auto parts image is captured. The global feature vector is then subjected to a one-dimensional convolution using a kernel of size K. It is possible to connect the cross-channel information of defects of auto parts and understand the relationship between each channel, and the weight coefficient of each channel is determined by the sigmoid activation function. Through these coefficients, the defect characteristics of each channel can be adjusted. Finally, the weight of each channel is applied to the corresponding element of the feature mapping in the original uncompressed auto parts image. Generating a feature map of the final output thereby enhances the ability to extract detailed information about target defects of various sizes in automotive components, which is particularly important for some small defects.

$$k = \varphi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{2}$$
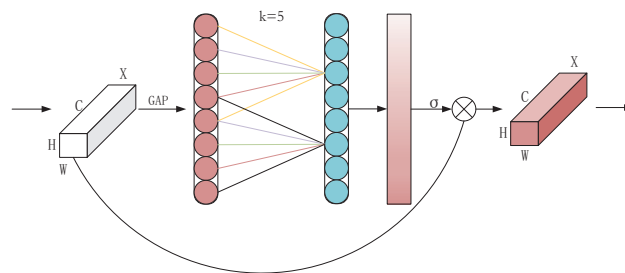

Figure 3. ECA structure diagram.

3.3. Loss Function

The original YOLOv7 algorithm employs the CIoU (complete intersection over union) loss function for prediction box regression calculations. The CIoU algorithm addresses the issue of bounding box aspect ratios. The loss function is an operation function used to measure the degree of difference between the predicted value and the real value of the model. The difference between the predicted value and the real value can be calculated through the loss function. The difference value can be backpropagated to update each parameter so as to make the model closer to the real value to achieve the purpose of learning and to improve the robustness of the model. CIoU operates on the principle illustrated in Figure 4:
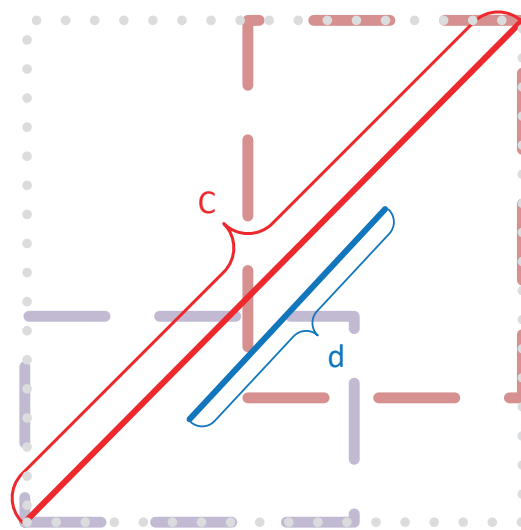

Figure 4. CIoU schematic diagram.

The formula for CIoU is as follows:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v \qquad (3)$$

IoU, or the intersection over union, measures the overlap between the predicted bounding box and the actual bounding box relative to their combined area. This metric calculates the ratio of the area of overlap to the total area encompassed by both boxes, providing a quantitative assessment of prediction accuracy. Its formula is:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \qquad (4)$$

where α represents the weighting function:

$$\alpha = \frac{v}{(1 - IoU) + v} \qquad (5)$$

where v is utilized to indicate the similarity of the aspect ratios.

$$v = \frac{4}{\pi^2}\left(\arctan\frac{\omega^{gt}}{h^{gt}} - \arctan\frac{\omega}{h}\right)^2 \qquad (6)$$

CIoU considers the overlapping area, the distance between centers of mass, and the aspect ratio of the predicted and actual boxes. However, for the defect features of auto parts, which have large aspect ratio differences, it cannot accurately capture the variance between width and height. This limitation hinders the effectiveness of the model at optimizing the similarity measure, and thus, the ability to improve the robustness of the model is limited. To overcome this challenge, this paper introduces Alpha-CIoU: a method aimed at enhancing the accuracy of bounding box regression and improving target detection.

Alpha-CIoU represents a uniform idempotentization of the existing loss based on IoU. It introduces a novel power IoU loss function that enhances the accuracy of bounding box regression and target detection. This method ensures an accurate representation of the aspect ratio, thus improving the model's performance. The formulation of Alpha-CIoU is as follows:

$$L_{Alpha-CIoU} = 1 - IoU^\alpha + \left(\frac{\left|C/\left(B \cup B^{gt}\right)\right|}{|C|}\right)^\alpha \qquad (7)$$

According to this formula, Alpha-CIoU ensures the accurate representation of the aspect ratio, effectively captures the variance between the width and height of the defect of automotive parts, and effectively improves the robustness of the model to improve the accuracy of the model compared with CIoU.

The variable α is selected to enhance the accuracy of bounding box regression through loss and adaptive gradient weighting for the target. In this study, α is set to 3 [44].

4. Experiments and Results

This section details extensive experiments and analyses to confirm the effectiveness of the proposed approach. During these evaluations, datasets from BYD car parts and the Northeastern University steel dataset are utilized to broaden the scope of the experiments. A detailed analysis of the results clearly indicates the superior performance of the proposed method.

4.1. Implementation Details

(1) Training strategy: The experimental environment utilizes PyTorch 1.91+CPU as the software framework, with Python 3.8 as the programming language. The model training hardware environment includes a GPU model NVIDIA GeForce RTX 3070 with 8 GB memory, and CUDA version 11.1 is utilized to accelerate model training.

(2) Evaluation: In this experiment, precision (P), recall (R), average precision (AP), mean average precision (mAP), and frames per second (FPS) are primarily selected as the evaluation indexes. The formulas for P, R, AP, and mAP are as follows:

$$P = \frac{T_P}{T_P + F_P} \tag{8}$$

$$R = \frac{T_P}{T_P + F_N} \tag{9}$$

$$AP = \int_0^1 P(r)dr \tag{10}$$

$$mAP = \frac{1}{c}\sum_{i=1}^{c} AP(i) \tag{11}$$

where T_P denotes the number of positive samples correctly predicted by the model, and F_P denotes the number of positive samples predicted by the model that were actually negative. F_N denotes the number of positive samples predicted by the model that are negative; r denotes the recall of the class, c denotes the number of all classes, and $AP(i)$ denotes the average precision of the ith cumulative iteration. FPS stands for frames per second, which refers to the number of images displayed per second in video or image processing. In practical applications, increasing FPS means being able to capture and process images faster, thereby improving productivity and inspection speed.

4.2. Datasets

The dataset utilized in this study consists of a custom collection focused on the interior button trim rings of BYD vehicles, which are made from PC+ABS material. This dataset comprises several defect types: friction, scratch, particle, black spot, and particle swarm, with 240 images for each category. The details of these defect types and features are illustrated in Figure 5. The dataset is divided into training, validation, and test sets with a ratio of 10:1:1. To assess whether the experimental improvements can be applied to other types of defect detection, we also utilize the NEU-DUT dataset from Northeastern University for comparative analysis.
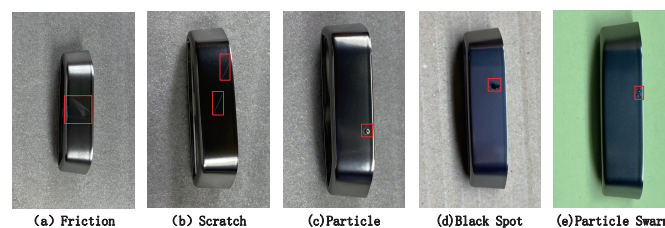

(a) Friction   (b) Scratch   (c)Particle   (d)Black Spot   (e)Particle Swarm

Figure 5. Samples of various types of defects in the BYD dataset.

Detecting defects in automotive components often involves dealing with noisy backgrounds. This challenge requires an algorithm with strong generalization capabilities and the ability to process diverse features and background information. In response, this study enhances the original YOLOv7's Mosaic-4 method by introducing a Mosaic-9 enhancement approach. This new method selects nine images at random from the dataset, applies various enhancement methods such as rotation and cropping, and combines them into a single image for input into the network.

The Mosaic-9 method, compared to the Mosaic-4 approach, compiles images that incorporate a richer array of feature and background information as well as targets of varying sizes, as demonstrated in Figure 6. This strategy increases the diversity of the data samples and enlarges the dataset, which significantly boosts the network's ability to generalize and minimizes the risk of model overfitting.
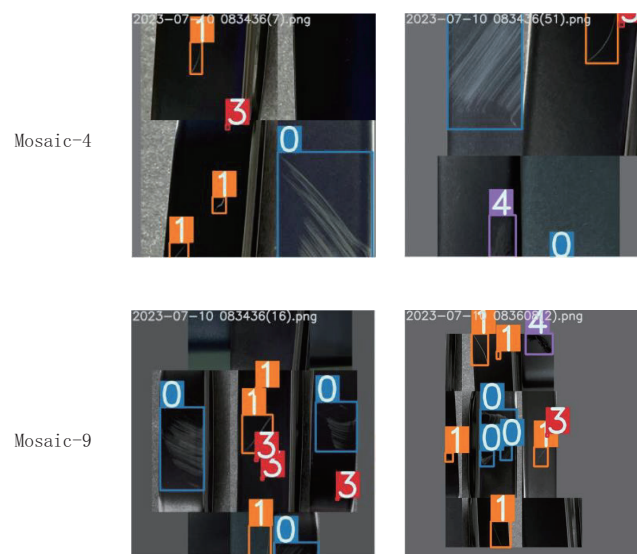
Figure 6. Comparison of different data enhancement methods.

To further verify the enhancement capabilities of our methodology for detecting different defects, we once again reference the NEU-DUT dataset from Northeastern University.

4.3. Ablation Study

4.3.1. Impact of Attention Mechanisms

This paper explores the effect of incorporating attention mechanisms on model accuracy. We integrate SE, SimAM, and ECA attention modules before the prediction layer in the model's multi-scale fusion process. The findings, presented in Table 1, exhibit that the model achieves $mAP50$ scores of 93.1%, 93.3%, and 93.5% after incorporating SE, SimAM, and ECA, respectively. These results indicate that leveraging attention mechanisms significantly boosts the defect detection performance for automotive parts.

Table 1. Performance comparison of different attention mechanisms based on the BYD dataset.

Method	AP_{SC}	AP_{BS}	AP_{Pa}	AP_{Ps}	AP_{Fr}	mAP^{50}	FPS	GFLPS	Params
SE	88.2	94.8	88.2	99.5	94.5	93.1	16.7	103.5	36.8
SimAM	90.1	92.3	89.0	98.4	96.5	93.3	38.0	103.2	36.5
ECA	86.9	95.2	90.6	99.6	95.1	93.5	38.6	103.3	36.5

ECA's ability to detect defects in automobile parts outperforms that of SE and SimAM, though it is marginally less effective at identifying Sc and Fr. Despite this, ECA features a faster FPS than both SE and SimAM while maintaining a similar count of parameters and floating-point operations. This indicates ECA's overall excellence in enhancing model detection capabilities without adding complexity.

4.3.2. Effect of Loss Function Hyperparameters

The paper further explores the effect of α in Equation (7) on the accuracy of defect detection in automobile parts. This is achieved through ablation studies with various α values. According to Table 2, the appropriate selection of α can fine-tune the accuracy of bounding box regression. Selecting an α that is either too high or too low can compromise detection accuracy. Adjusting α allows for a focused improvement to achieve a high IoU, leading to better regression accuracy. The findings suggest that setting α to 3 yields the best performance.

Table 2. Comparison of model performance for different α parameters.

α	AP_{Sc}	AP_{BS}	AP_{Pa}	AP_{Ps}	AP_{Fr}	mAP^{50}
1	85.0	94.0	84.2	99.6	95.0	91.6
2	83.5	94.5	85.8	99.6	95.6	91.8
3	88.6	93.5	88.6	99.5	95.5	93.1
4	84.3	94.2	83.8	99.6	95.2	91.6

4.3.3. Ablation Experiments with Different Modules

To verify the detection efficacy of the proposed algorithm and the effect of each enhancement method, the study conducts ablation tests using the YOLOv7 model on a dataset of BYD car parts. The presence of a "✓" in the table signifies the implementation of a particular enhancement strategy, where A stands for Mosaic-9, B stands for BiFPN, C stands for ECA attention, and D stands for Alpha-CIoU. Each experiment utilizes identical hyperparameters and training approaches, and the results are detailed in Table 3.

Table 3. Comparison of ablation experiment results for the MBEA-YOLOv7 model.

Test	Mosaic-9	BiFPN	ECA	Alpha-CIoU	AP_{Sc}	AP_{BS}	AP_{Pa}	AP_{Ps}	AP_{Fr}	mAP^{50}	R	P	FPS
1					84.7	85.2	82.3	99.5	92.1	88.7	89.8	86.1	19
2	✓				85.3	94.4	84.6	99.6	96.3	91.9	92.5	87.5	19
3		✓			87.3	94.9	83.7	99.6	94.3	92.0	93.0	89.0	30
4			✓		86.9	95.2	90.6	99.6	95.1	93.5	89.1	88.8	39
5				✓	88.6	93.5	88.6	99.5	95.5	93.1	90.1	87.8	30
6	✓	✓			90.2	94.8	88.0	99.6	95.1	93.6	91.0	90.3	29
7	✓	✓	✓		91.0	95.2	89.3	99.6	95.0	94.1	94.7	89.5	46
8	✓	✓	✓	✓	91.1	95.5	88.5	99.6	95.4	94.2	95.7	93.0	48

In the experiment, we reflected on the BYD dataset by adding each module. In Experiment 1, the baseline YOLOv7 model was employed. Experiment 2 enhanced the initial Mosaic-4 to Mosaic-9, leading to significant improvements. Specifically, the mAP50, recall, and precision increased by 3.2%, 2.7%, and 1.4%, respectively. This demonstrates that Mosaic-9 is capable of capturing a broader range of feature and background information as well as targets with varying scales, thus bolstering the model's robustness.

In Experiments 3 and 4, incorporating the BiFPN module, ECA attention module, and Alpha-IoU loss function resulted in significant enhancements. Specifically, the model's computational speed increased by 11.6/s, 19.9/s, and 10.9/s, respectively. Moreover, the mAP50 saw increases of 3.3%, 4.8%, and 4.4%, respectively. The accuracy for identifying various defects also demonstrated improvements. These enhancements mark a significant advancement over the original model. As illustrated in Figure 7, the YOLO-Former model demonstrates rapid convergence and requires shorter training periods. This is beneficial for fine-tuning and optimizing the algorithm. Observations from Figure 8a,b reveal that YOLO-Former performs well across most categories.

In Experiment 5, the adoption of Alpha-CIoU led to a 4.4% increase in mAP50 and slight improvements in detection accuracy and recall rates of 0.3% and 1.7%, respectively. In addition, the integration of three bounding box regression methods—DIoU [26], GIoU [27], and EIoU [28]—was introduced. The experiment consisted of 200 iteration rounds, with the mAP50 comparison represented in Figure 9.

The figure evidently demonstrates that the Alpha-CIoU loss function not only converges more swiftly but also achieves a higher mAP50 than the competing functions. This superior performance is attributed to the prevalent issue of mismatched anchors due to the anchor presetting mechanism, which results in a scarcity of high-quality anchors. In the regression loss calculation and backpropagation process, the predominance of low-quality frames has a more significant effect than the less frequent high-quality frames.
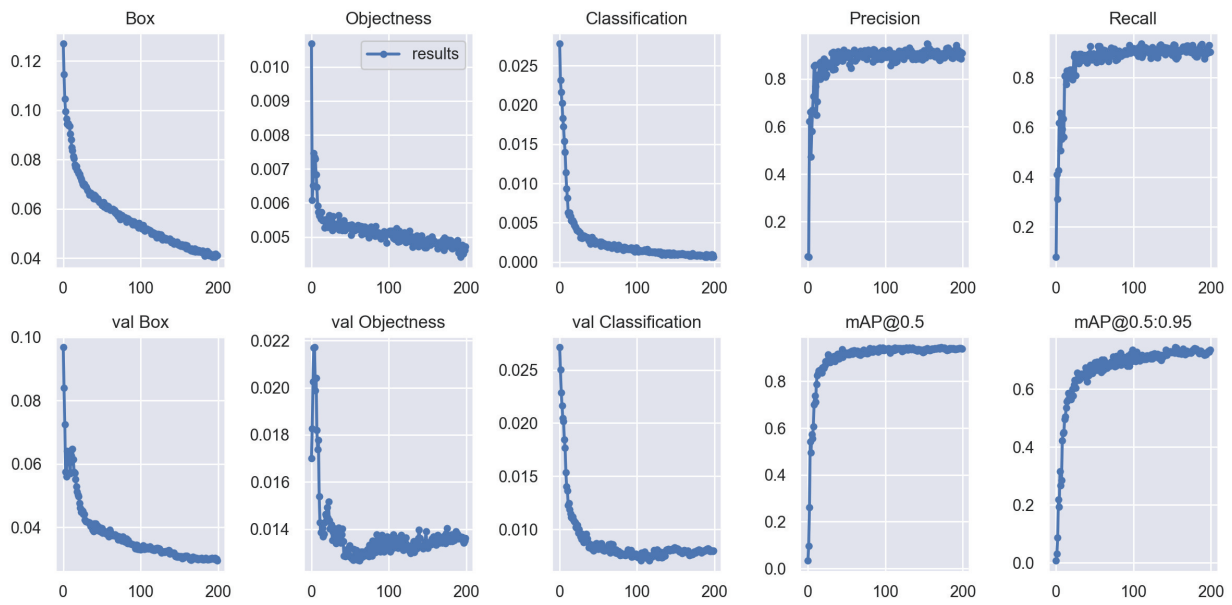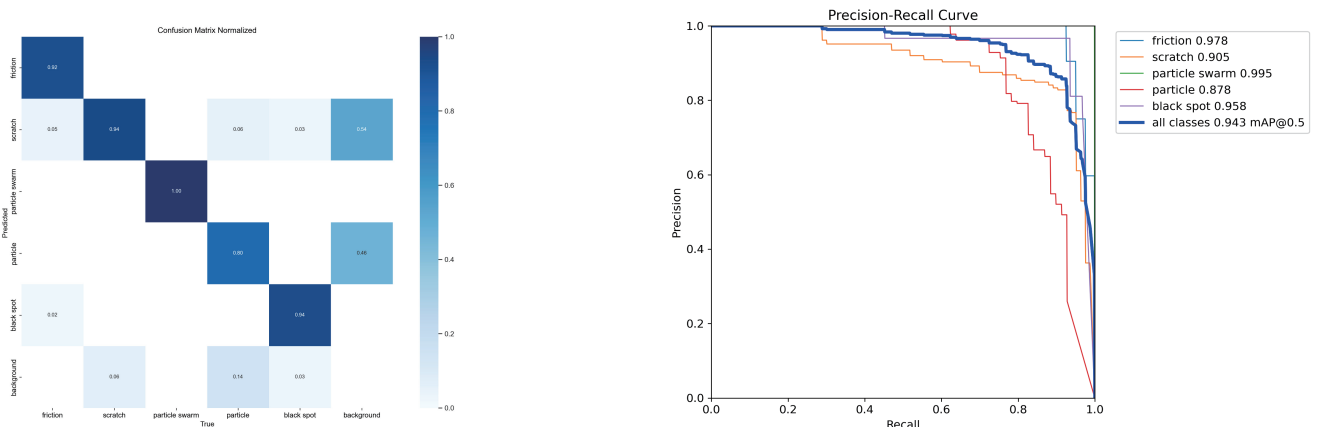
Figure 7. Loss curves of MBEA-YOLO on BYD dataset.

(**a**) Confusion matrix of MBEA-YOLO on BYD dataset.　　(**b**) Precision–recall curve of MBEA-YOLO on BYD dataset.

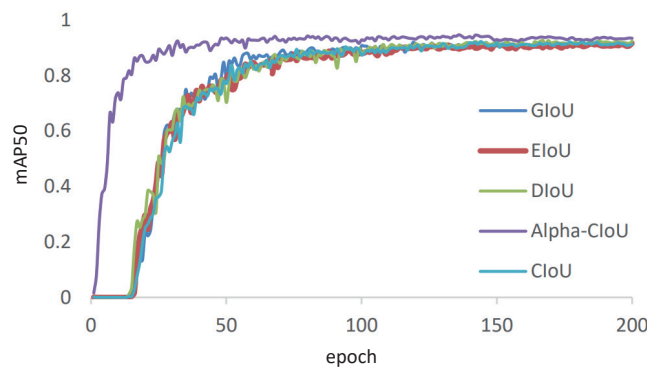**Figure 8.** (**a**) Confusion matrix and (**b**) precision–recall curve of MBEA-YOLO on BYD dataset.

Figure 9. Comparison of different loss functions.

To mitigate this issue, an exponential term ($0 < $ IoU) is added for weighting. When the IoU is larger, indicating a higher quality of the anchor, the weighting value of the exponential term is larger. Conversely, lower IoU scores, indicating poorer anchor quality, receive less weight. Accordingly, the enhanced loss function prioritizes high-quality anchors

and minimizes the effect of lower-quality ones. This adjustment effectively counters the challenge of unbalanced training samples in the original bounding box regression model.

4.3.4. Comparison of Detection Effects

To visualize the effect of the enhanced model on defect detection, this study analyzed the test set with both the original and the enhanced models. The comparison between YOLOv7 and the upgraded MBEA-YOLOv7 in some image tests is presented in Figure 10. This comparison aims to clearly demonstrate the enhancements the enhanced model introduces for detecting target defects.
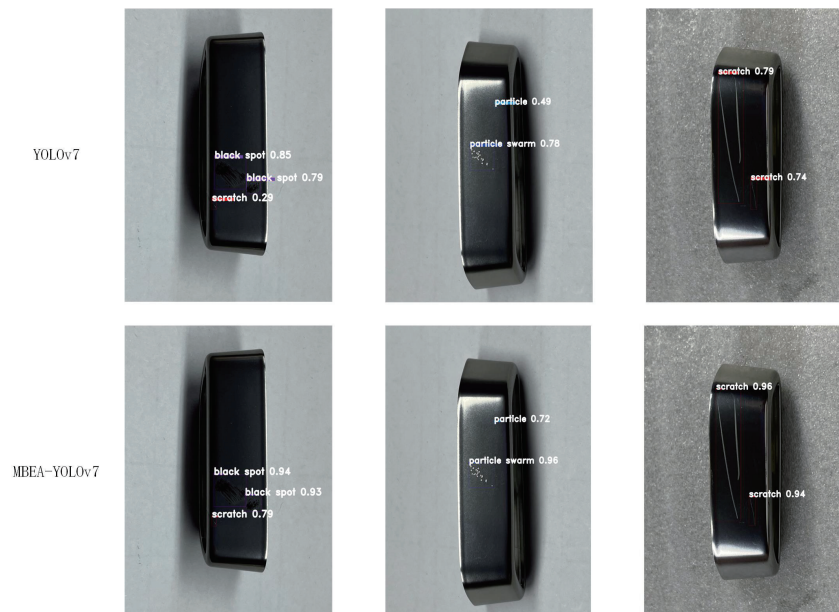

Figure 10. Comparison of detection performance on the BYD dataset.

4.3.5. Comparison and Generalization Experiments

To establish the superiority of the MBEA-YOLOv7 model introduced in this study over existing general target detection algorithms, we conducted comparative experiments on the NEU-DET dataset. These experiments pitted the newly proposed method against mainstream methods such as Faster-RCNN, SSD, YOLOv4, YOLOv5, YOLOv8, and YOLOv7.

The analysis, as illustrated in Table 4, reveals that the $mAP50$ of our proposed method surpasses that of Faster-RCNN, SSD, YOLOv4, YOLOv5, and YOLOv8 by 8.2, 12.9, 13.1, 12.8, and 3.8 percentage points, respectively. It also achieves the highest values in precision, recall, and detection speed. Despite a lower FPS compared to YOLOv8, it manages to detect targets in real time. In summary, the performance of the enhanced MBEA-YOLOv7 model exceeds that of other algorithms.

The exceptional performance of MBEA-YOLOv7 across various datasets confirms the algorithm's consistent detectability and strong ability to generalize across different datasets. The accuracy of MBEA-YOLOv7 reached 94.2% on the self-made BYD dataset and 68.8% on NEU-DUT. Compared with the original YOLOv7 model, the improvement is 5.5% and 7.5%, respectively, and the effect is remarkable.

Indeed, as indicated in the table, MBEA-YOLOv7 enhances the $mAP50$ by 7.5% in comparison to YOLOv7, showcasing enhanced accuracy and recall. The detection results depicted in Figure 11 further demonstrate that the algorithm we propose outperforms the original YOLOv7 model on various defect datasets. This highlights our algorithm's ability to maintain stable detection performance and exhibit a strong generalization capability across diverse datasets.

Table 4. Performance comparison of different algorithms on NEU-DUT.

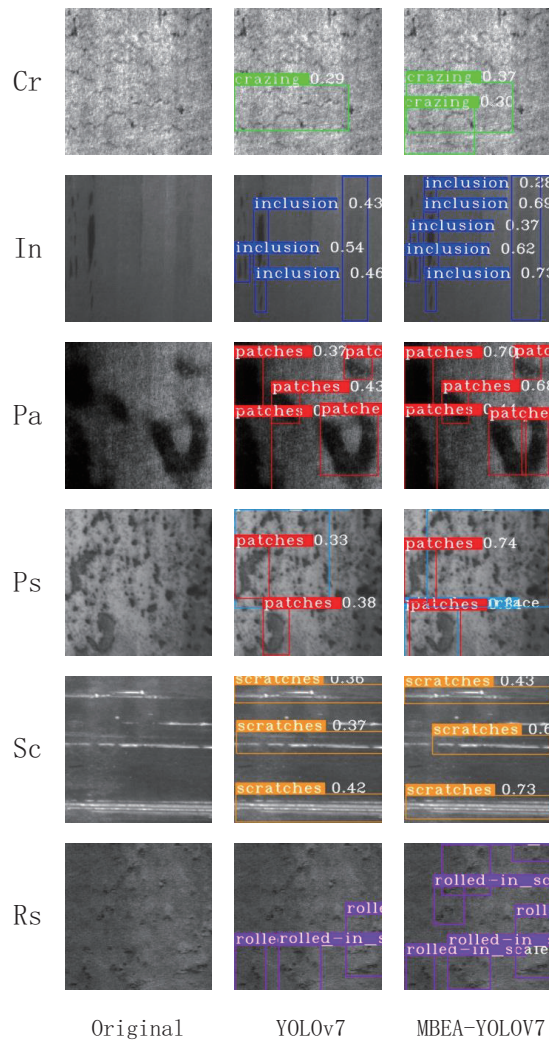| Method	Precision/%	Recall/%	mAP/%	FPS/s
Faster-RCNN	63.1	65.3	67.3	18
SSD	62.5	60.3	62.6	26
YOLOv4	61.5	61.6	62.4	31
YOLOv5	60.7	64.5	62.7	70
YOLOv7	68.3	65.8	68.0	48
YOLOv8	66.1	68.3	71.7	126
MBEA-YOLOv7	68.8	71.5	75.5	76

Figure 11. Comparison chart of detection effect.

5. Conclusions

In this paper, we propose a multi-class object detection method, the MBEA-YOLO network, for defect detection in automotive parts. This approach integrates feature extraction and fusion with attention mechanisms. For the self-built BYD data volume, compared with the original YOLOv7, accuracy increased by 3.3% and 4.8%, respectively, and the speed increased by 11.6/s and 19.9/s, respectively. These components enhance the accuracy for detecting small defects while maintaining the speed necessary for real-time, dynamic inspection tasks. Not only has our MBEA-YOLO network proven itself on the BYD dataset that it was trained on, but the Northeastern University steel dataset (NEU-DUT) has also been evaluated, and the effect is equally significant. The accuracy and speed also have

significant advantages compared to some popular deep learning methods. Compared to the original model, the accuracy and speed have improved by 7.5% and 28/s, respectively. Even compared to the latest YOLOv8, the accuracy has a 4.8% advantage. This confirms the outstanding efficacy of the method at identifying defects in automotive components. However, it can be seen that MBEA-YOLOv7 is 50/s slower than YOLOv8 for detection speed, and it can be found that YOLOv8 is more lightweight, which is also a goal for the future. The goal is to achieve higher accuracy and faster detection speed with a more lightweight model.

For future studies, we aim to compile extensive datasets from various locations, including a wide array of defect types in automotive parts. This comprehensive data collection is designed to broaden the adaptability of the methods we advocate. We also plan to delve into cutting-edge technologies to forge more powerful and efficient solutions leveraging diverse deep learning architectures. The objective seeks to narrow the divide between theoretical innovation and practical utility, and we aspire to make a significant contribution to the field. We envision our work benefiting traffic survey efforts under demanding conditions; this marks a promising direction for future research. We are eager to witness the progress in this arena.

Author Contributions: Conceptualization, H.H. and K.Z.; Methodology, H.H. and K.Z.; Software, H.H.; Validation, H.H.; Formal analysis, H.H.; Investigation, H.H.; Resources, H.H.; Data curation, H.H.; Writing—original draft, H.H.; Writing—review & editing, K.Z.; Supervision, K.Z.; Funding acquisition, K.Z. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data are contained within the article.

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

References

1. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [CrossRef]
2. Yan, H.; Cai, J.-F.; Zhao, Y.; Jiang, Z.; Zhang, Y.; Ren, H.; Zhang, Y.; Li, H.; Long, Y. A lightweight high-resolution algorithm based on deep learning for layer-wise defect detection in laser powder bed fusion. *Meas. Sci. Technol.* **2023**, *35*, 025604. [CrossRef]
3. Li, Z.; Zhang, Y.; Fu, X.; Wang, C. Metal surface defect detection based on improved yolov5. In Proceedings of the 2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS), Chengdu, China, 16–18 June 2023; pp. 1147–1150.
4. Kumar, A. Computer-vision-based fabric defect detection: A survey. *IEEE Trans. Ind. Electron.* **2008**, *55*, 348–363. [CrossRef]
5. Kim, K.J.; Lee, J.-W. Light-weight design and structure analysis of automotive wheel carrier by using finite element analysis. *Int. J. Precis. Eng. Manuf.* **2022**, *23*, 79–85. [CrossRef]
6. Xu, J.; Xi, N.; Zhang, C.; Shi, Q.; Gregory, J. Real-time 3d shape inspection system of automotive parts based on structured light pattern. *Opt. Laser Technol.* **2011**, *43*, 1–8. [CrossRef]
7. Ho, C.-C.; Hernandez, M.A.B.; Chen, Y.-F.; Lin, C.-J.; Chen, C.-S. Deep residual neural network-based defect detection on complex backgrounds. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 5005210. [CrossRef]
8. Yu, X.; Lyu, W.; Wang, C.; Guo, Q.; Zhou, D.; Xu, W. Progressive refined redistribution pyramid network for defect detection in complex scenarios. *Knowl.-Based Syst.* **2023**, *260*, 110176. [CrossRef]
9. Yang, L.; Fan, J.; Huo, B.; Li, E.; Liu, Y. A nondestructive automatic defect detection method with pixelwise segmentation. *Knowl.-Based Syst.* **2022**, *242*, 108338. [CrossRef]
10. Zou, X.; Zhao, J.; Li, Y.; Holmes, M. In-line detection of apple defects using three color cameras system. *Comput. Electron. Agric.* **2010**, *70*, 129–134.
11. 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, 24–27 June 2014; pp. 580–587.
12. 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, NE, USA, 26 June–1 July 2016; pp. 779–788.

13. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14; Springer: Cham, Switzerland, 2016; pp. 21–37.

14. Redmon, J.; Farhadi, A. Yolo9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.

15. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.

16. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.

17. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]

18. Cha, Y.-J.; Choi, W.; Suh, G.; Mahmoudkhani, S.; Büyüköztürk, O. Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 731–747. [CrossRef]

19. Tao, X.; Zhang, D.; Wang, Z.; Liu, X.; Zhang, H.; Xu, D. Detection of power line insulator defects using aerial images analyzed with convolutional neural networks. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *50*, 1486–1498. [CrossRef]

20. He, Y.; Song, K.; Meng, Q.; Yan, Y. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 1493–1504. [CrossRef]

21. Cheng, J.C.; Wang, M. Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques. *Autom. Constr.* **2018**, *95*, 155–171. [CrossRef]

22. Lei, H.; Wang, B.; Wu, H.; Wang, A. Defect detection for polymeric polarizer based on faster r-cnn. *J. Inf. Hiding Multim. Signal Process.* **2018**, *9*, 1414–1420.

23. Zhao, Z.; Zhen, Z.; Zhang, L.; Qi, Y.; Kong, Y.; Zhang, K. Insulator detection method in inspection image based on improved faster r-cnn. *Energies* **2019**, *12*, 1204. [CrossRef]

24. Neuhauser, F.M.; Bachmann, G.; Hora, P. Surface defect classification and detection on extruded aluminum profiles using convolutional neural networks. *Int. J. Mater. Form.* **2020**, *13*, 591–603. [CrossRef]

25. Sun, X.; Gu, J.; Huang, R.; Zou, R.; Palomares, B.G. Surface defects recognition of wheel hub based on improved faster r-cnn. *Electronics* **2019**, *8*, 481. [CrossRef]

26. Chen, J.; Liu, Z.; Wang, H.; Núñez, A.; Han, Z. Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network. *IEEE Trans. Instrum. Meas.* **2017**, *67*, 257–269. [CrossRef]

27. Li, Y.; Huang, H.; Xie, Q.; Yao, L.; Chen, Q. Research on a surface defect detection algorithm based on mobilenet-ssd. *Appl. Sci.* **2018**, *8*, 1678. [CrossRef]

28. Zhang, C.; Chang, C.-C.; Jamshidi, M. Concrete bridge surface damage detection using a single-stage detector. *Comput.-Aided Civ. Infrastruct. Eng.* **2020**, *35*, 389–409. [CrossRef]

29. Zheng, Q.; Wang, L.; Wang, F. Small object detection in traffic scene based on improved convolutional neural network. *Comput. Eng.* **2020**, *46*, 26–33.

30. Ju, M.; Luo, J.; Wang, Z.; Luo, H. Multi-scale target detection algorithm based on attention mechanism. *Acta Opt. Sin.* **2020**, *466*, 132–140.

31. Cui, Z.; Qin, Y.; Zhong, Y.; Cao, Z.; Yang, H. Target Detection in High-Resolution Sar Image via Iterating Outliers and Recursing Saliency Depth. *Remote Sens.* **2021**, *13*, 4315. [CrossRef]

32. Liu, J.; Jia, R.; Li, W.; Ma, F.; Abdullah, H.M.; Ma, H.; Mohamed, M.A. High precision detection algorithm based on improved retinanet for defect recognition of transmission lines. *Energy Rep.* **2020**, *6*, 2430–2440. [CrossRef]

33. Liu, J.; Liang, H.; Cui, X.; Zhong, M.; Li, C. SSD visual target detector based on feature integration and feature enhancement. *J. Comput. Eng. Appl.* **2022**, *58*, 150-159.. [CrossRef]

34. Li, Z.; Zhou, F. Fssd: Feature fusion single shot multibox detector. *arXiv* **2017**, arXiv:1712.00960.

35. Shi, W.; Bao, S.; Tan, D. Ffessd: An accurate and efficient single-shot detector for target detection. *Appl. Sci.* **2019**, *9*, 4276. [CrossRef]

36. Zhao, P.; Xie, L.; Peng, L. Deep small object detection algorithm integrating attention mechanism. *J. Front. Comput. Sci. Technol.* **2022**, *16*, 927-937.

37. Ren, J.; Ren, R.; Green, M.; Huang, X. Defect detection from X-ray images using a three-stage deep learning algorithm. In Proceedings of the 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), Edmonton, AB, Canada, 5–8 May 2019; pp. 1–4.

38. Du, W.; Shen, H.; Fu, J.; Zhang, G.; He, Q. Approaches for improvement of the X-ray image defect detection of automobile casting aluminum parts based on deep learning. *NDT E Int.* **2019**, *107*, 102144.

39. Tsai, D.-M.; Fan, S.-K.S.; Chou, Y.-H. Auto-annotated deep segmentation for surface defect detection. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [CrossRef]

40. Shin, S.; Jin, C.; Yu, J.; Rhee, S. Real-time detection of weld defects for automated welding process base on deep neural network. *Metals* **2020**, *10*, 389. [CrossRef]

41. Block, S.B.; da Silva, R.D.; Dorini, L.B.; Minetto, R. Inspection of imprint defects in stamped metal surfaces using deep learning and tracking. *IEEE Trans. Ind. Electron.* **2020**, *68*, 4498–4507. [CrossRef]

42. Chen, X.; Chen, J.; Han, X.; Zhao, C.; Zhang, D.; Zhu, K.; Su, Y. A light-weighted cnn model for wafer structural defect detection. *IEEE Access* **2020**, *8*, 24006–24018. [CrossRef]

43.   Huang, S.-C.; Le, T.-H.; Jaw, D.-W. Dsnet: Joint semantic learning for object detection in inclement weather conditions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 2623–2633. [CrossRef]

44.   He, J.; Erfani, S.; Ma, X.; Bailey, J.; Chi, Y.; Hua, X.-S. Alpha-iou: A family of power intersection over union losses for bounding box regression. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 20230–20242.