

Article

Enhanced YOLOv8 with BiFPN-SimAM for Precise Defect Detection in Miniature Capacitors

Ning Li ¹, Tianrun Ye ², Zhihua Zhou ^{1,2}, Chunming Gao ^{1,2,3,*} and Ping Zhang ^{1,2,3,*}

¹ School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; 13880777799@163.com (N.L.); zhouzhihua2022@163.com (Z.Z.)

² Yibin Park of University of Electronic Science and Technology of China, Yibin 644000, China; 15696420859@163.com

³ Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen 518000, China

* Correspondence: gaocm@uestc.edu.cn (C.G.); pingzh@uestc.edu.cn (P.Z.)

Abstract: In the domain of automatic visual inspection for miniature capacitor quality control, the task of accurately detecting defects presents a formidable challenge. This challenge stems primarily from the small size and limited sample availability of defective micro-capacitors, which leads to issues such as reduced detection accuracy and increased false-negative rates in existing inspection methods. To address these challenges, this paper proposes an innovative approach employing an enhanced ‘you only look once’ version 8 (YOLOv8) architecture specifically tailored for the intricate task of micro-capacitor defect inspection. The merging of the bidirectional feature pyramid network (BiFPN) architecture and the simplified attention module (SimAM), which greatly improves the model’s capacity to recognize fine features and feature representation, is at the heart of this methodology. Furthermore, the model’s capacity for generalization was significantly improved by the addition of the weighted intersection over union (WISE-IOU) loss function. A micro-capacitor surface defect (MCSD) dataset comprising 1358 images representing four distinct types of micro-capacitor defects was constructed. The experimental results showed that our approach achieved 95.8% effectiveness in the mean average precision (mAP) at a threshold of 0.5. This indicates a notable 9.5% enhancement over the original YOLOv8 architecture and underscores the effectiveness of our approach in the automatic visual inspection of miniature capacitors.



Citation: Li, N.; Ye, T.; Zhou, Z.; Gao, C.; Zhang, P. Enhanced YOLOv8 with BiFPN-SimAM for Precise Defect Detection in Miniature Capacitors. *Appl. Sci.* **2024**, *14*, 429. <https://doi.org/10.3390/app14010429>

Academic Editor: Antonio Fernández-Caballero

Received: 23 November 2023

Revised: 26 December 2023

Accepted: 27 December 2023

Published: 3 January 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/>).

Keywords: micro-capacitor quality control; defect detection; YOLOv8 architecture; bidirectional feature pyramid network (BiFPN); weighted intersection over union (WISE-IOU) loss function

1. Introduction

In the contemporary landscape of electronic product development, the significance of miniature capacitors is increasingly recognized, particularly in applications requiring high-precision charge storage and voltage regulation. These components are vital in the operation of high-frequency circuits and communication systems, playing a critical role in maintaining voltage stability and suppressing noise interference. A primary challenge in the utilization of these devices is the accurate detection and characterization of minuscule defects, which, despite their small size, can lead to substantial complications, such as system instability, data corruption, and increased security vulnerabilities [1,2].

Traditional methodologies for the quality assessment of micro-capacitors, including microscopic inspection and electrical performance tests, are progressively being outpaced by the demands of modern technology. These methods, albeit being reliable to a certain extent, are limited in their ability to detect defects at the micron scale. For instance, microscopic inspection is heavily dependent on the skill level of the operator and necessitates specialized expertise. Additionally, electrical performance tests may fail to identify minor

physical anomalies that do not immediately affect functionality but could potentially lead to long-term degradation [3,4].

The use of automated visual inspection technologies has become the norm to overcome these constraints. These systems take advantage of recent developments in pattern recognition and image processing, especially with the introduction of deep learning methods like convolutional neural networks (CNNs). Numerous CNN-based models, such as Fast R-CNN, Faster R-CNN, and R-CNN itself, have shown a great deal of effectiveness in detecting flaws in electronic components [5,6]. Nevertheless, these models are not very good at identifying minuscule flaws, which is made worse by the difficulty in labeling the data.

To surmount these challenges, this paper proposes an advanced approach employing the YOLOv8 model, known for its proficient single-stage target detection. YOLOv8 has a proven track record of processing complicated visual data and provides a speed–accuracy balance that is better than that of conventional two-stage models, such as the R-CNN family models [7,8]. To improve YOLOv8, we introduced the BiFPN (bidirectional feature pyramid network) and SimAM (simplified attention module). The goal of adding the SimAM was to improve the model’s capacity to identify important elements in intricate visual data—a situation that frequently arises in the identification of micro-capacitor defects. SimAM accomplishes this by more efficiently directing the model’s attention, which improves the accuracy of small-scale flaw identification [8]. The issues, caused by a lack of labeled data, are addressed by the incorporation of the BiFPN architecture. Through the establishment of a modified feature layer hierarchy, BiFPN enables more efficient feature integration and transfer across scales, improving the accuracy of the defect detection model [9–11]. In addition, the WISE-IOU (WIoU) loss function was introduced to improve the robustness and generalization of the model [12]. This is essential in real-world situations where sample variability and the effects of outliers are common. This enhancement was achieved via a refined feature layer hierarchy, which facilitates more effective feature transfer and integration across various scales, thus boosting the accuracy and dependability of defect detection [13,14]. Moreover, the incorporation of the WISE-IOU (WIoU) loss function significantly bolstered the model’s generalization capabilities and robustness [12]. These improvements culminated in a notable enhancement of the model’s performance, as evidenced by achieving a 95.8% mAP@0.5, representing a substantial 9.3 percentage point increase compared with the baseline YOLOv8 architecture. Additionally, the model maintained a real-time detection speed that aligns with industrial standards, underscoring the practicality and effectiveness of our approach in identifying defects in miniature capacitors.

Our work’s contributions are multifaceted:

- (1) Integrated algorithmic enhancement and performance efficiency: The deployment of YOLOv8 for detecting defects in micro-capacitors was notably advanced by integrating the SimAM attention mechanism with the BiFPN architecture. This combination significantly improved the model’s precision in identifying small defects amidst complex visual backgrounds. The SimAM mechanism sharpened feature discernment, while the BiFPN architecture enhanced multi-scale feature fusion, leading to increased detection accuracy. Furthermore, the adoption of the WISE-IOU loss function marked a crucial progression in boosting the model’s generalization and robustness. This adjustment is vital for ensuring real-time detection efficiency, effectively addressing challenges of sample variability and outlier impacts, and maintaining a balance between computational demand and operational effectiveness.
- (2) Dataset compilation and application relevance: The development of a high-quality dataset, derived from real industrial scenarios, underpins this paper. This dataset not only serves as a benchmark for model evaluation but also enriches the pool of data available for future research in this field. The way it is structured plays a crucial role in pushing forward the use of deep learning to identify defects in micro-capacitors, particularly in settings in which data are scarce.

We start with a review of existing literature in Section 2, summarizing key knowledge and relevant studies to lay the groundwork for our work. In Section 3, we explore the architecture and essential elements of our proposed method, giving a detailed view of its design and execution. Section 4 provides a comprehensive evaluation of our technique, including an ablation study to highlight the effectiveness of various components and a performance comparison. Finally, Section 5 wraps up our study, reflecting on what our findings mean and suggesting future research avenues in this area.

2. Related Work

Increasing complexity and the miniaturization trend in electronic products have brought significant challenges to quality control, especially in detecting minute defects in capacitors. Traditional methods, such as microscopic inspection and electrical testing, are becoming less effective in this evolving context [13]. This has led to the adoption of automated visual inspection methods using deep learning and convolutional neural networks (CNNs).

Recent advancements in this field include optimizing network structures to enhance real-time performance while maintaining detection accuracy. Techniques like depth-separable convolution and more efficient channel attention (MECA) have been utilized to reduce computational demands and improve feature extraction. Furthermore, to address the loss of detail in small targets due to pooling operations, methods using dilated convolution with varying rates (atrous spatial pyramid pool, or ASPF) have been developed. These methods extract more contextual information, which is crucial for detecting tiny defects [14,15].

In addition to network enhancements, new feature fusion methods have been proposed. These methods introduce shallower feature maps and employ dense multiscale weighting to fuse more detailed information, thereby improving detection accuracy. Optimization techniques such as the K-means++ algorithm for reconstructing prediction frames have also been integrated to accelerate model convergence, along with the combination of the Mish activation function and the SIOU loss function to further refine model performance [16–18].

Another critical aspect is addressing overfitting, especially in contexts with limited data. A coarse-grained regularization method for convolution kernels (CGRCKs) was introduced to maximize the difference between convolution kernels within the same layer. This method enhances the extraction of multi-faceted features and shows effectiveness compared with traditional L1 or L2 regularization, particularly in CNN applications in which dataset sizes are constrained [19–21].

The development of high-quality test datasets, often derived from real-world production settings, remains vital. These datasets serve as practical benchmarks for training and validating models, which are essential for assessing their real-world applicability, especially in industrial settings for detecting defects in miniature capacitors [22,23].

In summary, the field of miniature capacitor defect detection is rapidly evolving, with deep learning technologies at the forefront. Advances in network optimization, feature fusion techniques, and regularization methods have significantly improved detection efficiency and accuracy. These developments not only enhance quality control in the electronics manufacturing industry but also set the stage for future research exploring a balance between the processing speed, accuracy, model generalization, and effective management of large diverse datasets.

3. Materials and Methods

3.1. YOLOv8 Improved Model

In the dynamic landscape of target detection algorithms, the YOLO (you only look once, only vision) series stands as a paragon of speed and accuracy. Among its iterations, the latest, YOLOv8, has rapidly gained prominence within the series. The model is offered in various configurations, such as YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and

YOLOv8x [24], each tailored to specific performance requirements. In our work, YOLOv8n was selected as the foundational model. The structure of YOLOv8n is shown in Figure 1.

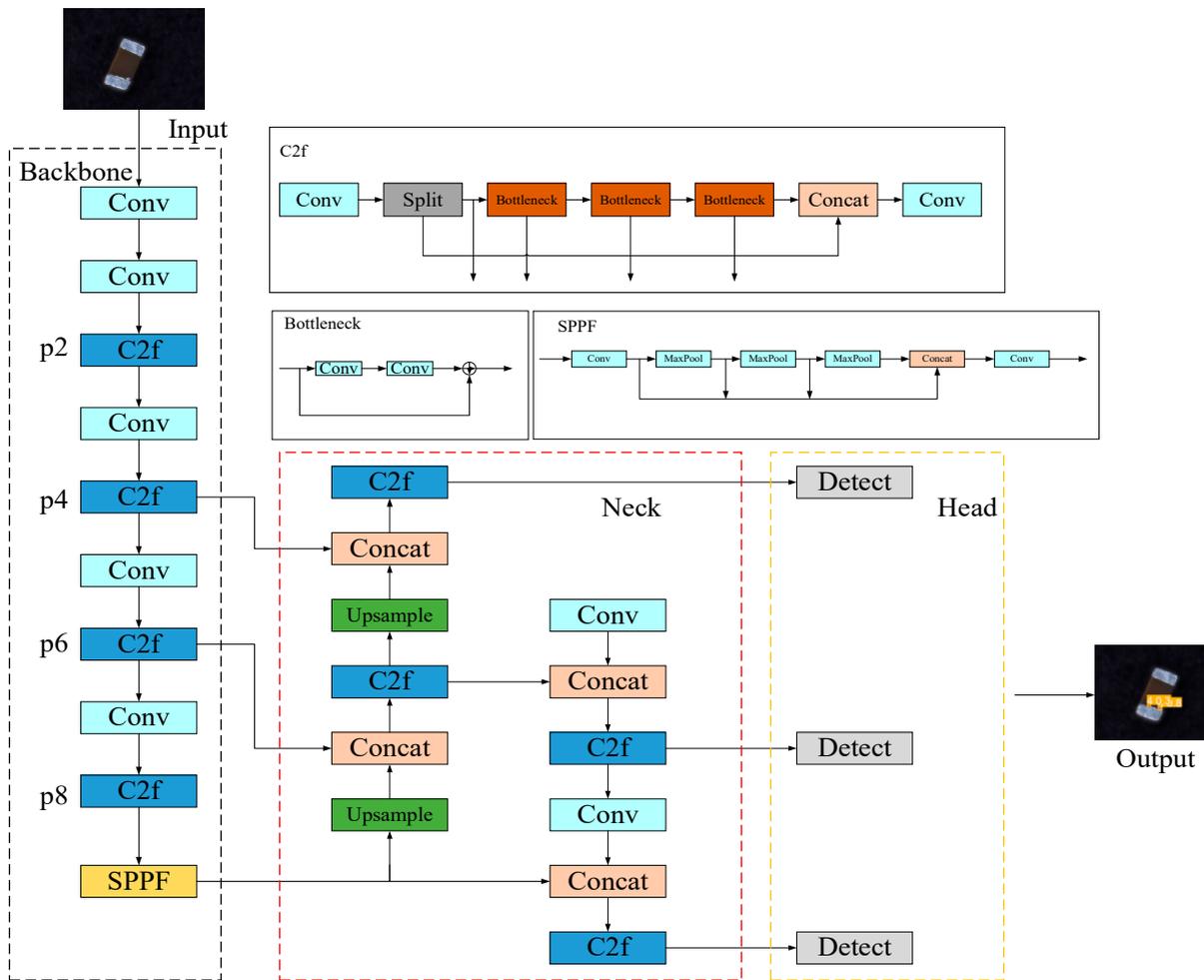


Figure 1. YOLOv8n network structure.

While the original YOLOv8n model boasts commendable overall performance, it exhibits certain limitations in identifying small dense defects. To surmount these challenges and enhance the model’s efficacy in pinpointing diminutive flaws in micro-capacitors, this paper introduces three significant enhancements to the YOLOv8 architecture. Figure 2 displays the enhanced network architecture. First, we integrated an attention mechanism known as the simplified attention module (SimAM) into the YOLOv8 framework. The purpose of this integration was to increase the model’s attention to smaller targets to improve its ability to identify and categorize critical minor flaws during micro-capacitor inspections. Second, the model’s multi-scale information fusion was improved by BiFPN combined with the p2 layer, which also increases the accuracy of small target detection. Lastly, the WISE-IOU [12,25] loss function replaced the traditional CIoU loss function. The purpose of this modification was to focus the model’s attention on higher-quality, representative examples. WISE-IOU adoption greatly strengthened the model’s capacity for generalization, which in turn improved the model’s robustness and performance in a variety of detection scenarios.

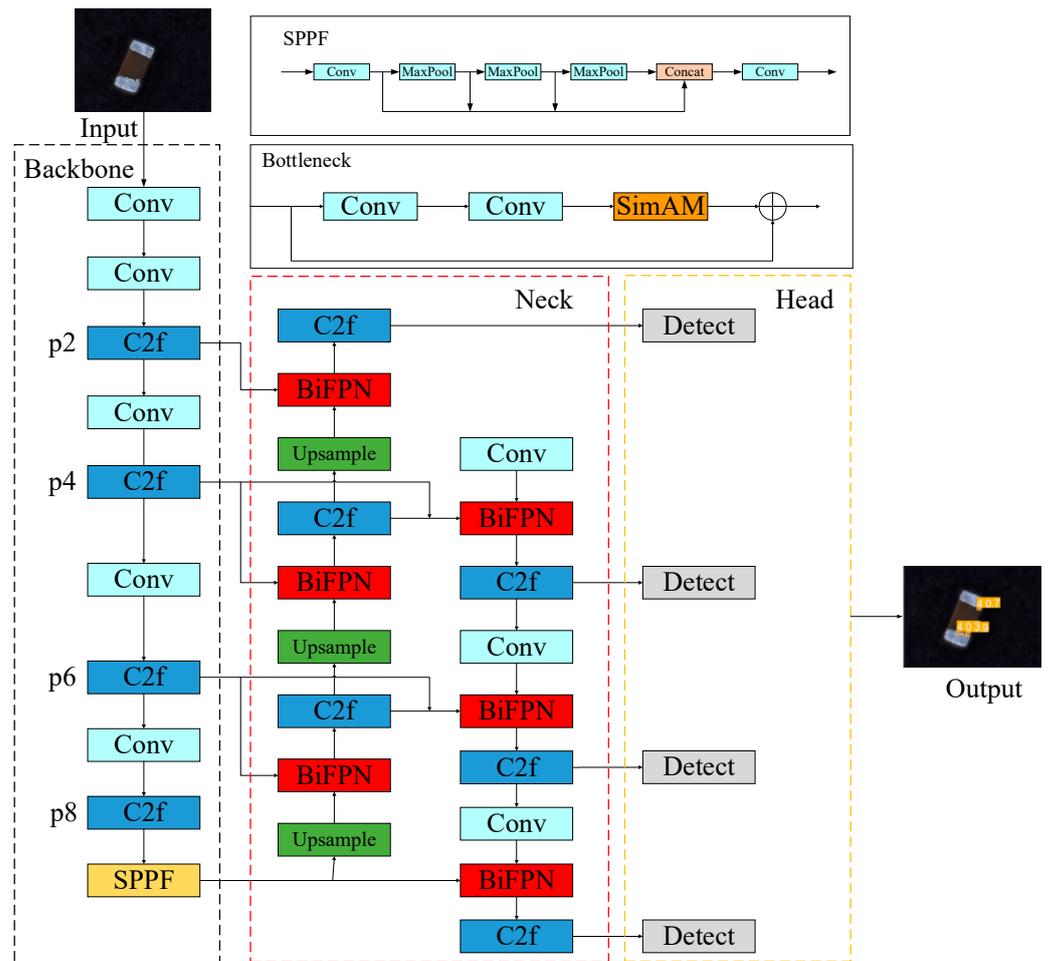


Figure 2. Improved YOLOv8n network structure.

3.2. SimAM Attention Mechanism

Drawing inspiration from human neuronal activities, this mechanism mimics the distinctive firing patterns of information-rich neurons, which have the capacity to suppress the activity of their less informative neighbors. In the context of neuroscience, these pivotal neurons are identified through the formulation of a specific energy function for each neuron, facilitating the pinpointing of neurons that carry crucial information. Translating this concept into the domain of machine learning, the SimAM attention mechanism adeptly applies this principle. It possesses the ability to dynamically learn and harness similarity information among targets, thereby accurately determining the similarity metric across various features. Subsequently, it assigns appropriate weights to these features on the basis of their informational value. This method effectively bolsters the model’s proficiency in detecting target defects within miniature capacitors. The underlying principle of SimAM is grounded in the recognition that not every feature contributes equally to the detection task. Certain features are more informative and play a pivotal role in the accuracy of defect detection. Similar to how information-rich neurons in human neuroscience display distinct firing patterns and modulate the activity of nearby neurons lacking in informative content, SimAM’s focus on these salient features guarantees that the model’s attention is focused on the most relevant aspects of the data [25].

$$e_t(w_t, b_t, y, x_i) = (y_t - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2, \quad (1)$$

The number of neurons is indicated by M in the preceding equation, and $\hat{t} = w_t t + b_t$ and $\hat{x}_i = w_t x_i + b_t$ represent the linear transformations of the target neuron and other neurons in the same channel in the input eigenmap [26]. To make things simpler, the scalars y_t and y_0 have binary labels, and to recombine the energy equations and obtain Equation (2), the regular term λ is added.

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - (w_t x_i + b_t))^2 + (-1 - (w_t t + b_t))^2 + \lambda w_t^2 \tag{2}$$

Solving the above equation yields the weights w_t and deviations b_t :

$$w_t = -\frac{2(t - \mu_t)}{(t - \mu_t)^2 + 2\sigma t^2 + 2\lambda} \tag{3}$$

$$b_t = -\frac{1}{2}(t + \mu_t)w_t \tag{4}$$

μ_t and σ represent the mean variance of every neuron in that channel except t , which eventually results in the formula for the least energy:

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \tag{5}$$

According to the above equation, a neuron’s significance increases with decreasing energy e_t^* since it allows for greater neuronal differentiation from nearby neurons. Equation (6) [27], when applied to deep neural networks, displays the final formula:

$$\tilde{X} = Sigmoid\left(\frac{1}{E}\right)X \tag{6}$$

The use of the sigmoid function is primarily to suppress the value of the energy in E that is too great, where X represents the input features. E is the sum of the minimum energy in all spatial and channel dimensions. Figure 3 displays the construction of the SimAM attention module. In this paper, the YOLOv8 bottleneck is integrated with the SimAM attention mechanism, which makes the model more focused on detecting targets. Furthermore, SimAM is very design-oriented, does not include too many extra parameters, and can ensure a high detection speed while improving the model’s detection accuracy [28–30].

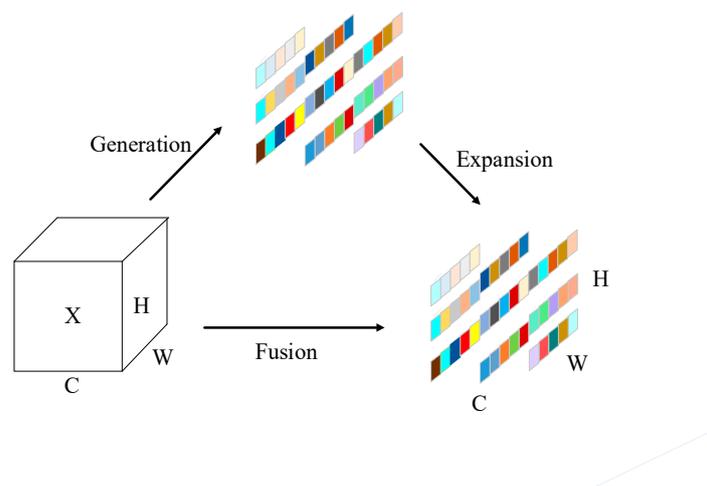


Figure 3. SimAM attention module structure.

3.3. Bidirectional Feature Pyramid Network

The neck network of YOLOv8 uses a synergistic combination of the path aggregation network (PAN) and feature pyramid network (FPN) architectures, as shown in Figure 4b. The FPN framework adeptly channels deep feature information to shallower layers, thereby enriching them with critical, high-level insights. Conversely, the PAN architecture facilitates the upward flow of precise positional data from the superficial layers to the deeper, feature-rich strata. This fusion, coined the PANet structure, masterfully amalgamates shallow and deep features, significantly bolstering the model's aptitude for discerning even the most nuanced characteristics [31].

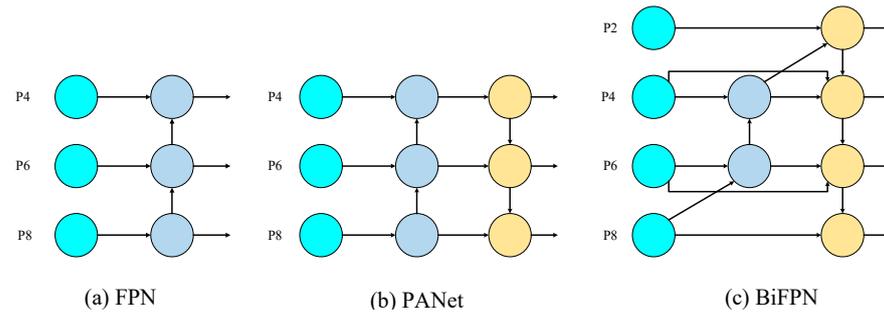


Figure 4. Neck feature network design: (a) FPN, (b) PANet, and (c) BiFPN.

However, our analysis revealed a notable shortfall in the PANet configuration. The pathway feeding into the PAN, previously processed by the FPN, inadvertently filters out some quintessential feature information originally harvested from the YOLOv8 backbone. To rectify this, we innovatively integrated a bidirectional feature pyramid network (BiFPN) into our model, as illustrated in Figure 4c. The BiFPN architecture innovates by introducing two additional lateral connection paths to the existing FPN+PAN framework. These novel pathways adeptly preserve and incorporate the raw features extracted directly from the backbone network into the detection feature map [32].

Moreover, we strategically incorporated the P2 layer into our neck network. This layer, characterized by its expansive feature map size and minimal convolution operations, is accompanied by an additional detection head. These augmentations serve a dual purpose: they not only intensify the fusion of positional and feature information within the model but also markedly elevate the precision in detecting minuscule targets. The culmination of these enhancements is vividly showcased in Figure 4.

3.4. WISE-IoU Loss Function

In the field of machine vision, the YOLOv8 model uses the complete intersection over union (CIoU) method for bounding box regression loss computation, which is particularly useful in the complex task of detecting faults in microscopic capacitors. A highly regarded metric for object detection accuracy, the CIoU measures how well anticipated and real bounding boxes match. This is accomplished by calculating the Euclidean distance and taking the detection frame's aspect ratio into account, providing a thorough similarity measurement [33]. This method is essential for improving object localization and size estimation accuracy in a variety of settings.

Nevertheless, the application of CIoU encounters notable challenges within datasets of tiny capacitor defects, which are characterized by their diminutive size and intricate defect patterns. These datasets frequently include samples of subpar quality due to the inherent complexity of such small-scale defects. In this context, CIoU's reliance on distance and aspect ratio metrics can disproportionately penalize these lower-quality examples. This bias may lead to a decline in the model's generalization capability, as it tends to overfit to these outlier samples, diminishing its efficacy in more typical scenarios.

The weighted intersection over union (WIoU) technique, which includes a dynamic and non-monotonic focusing mechanism, was developed as a solution to this problem. It

uses the dataset's "outliers" in a novel way to assess the anchor frame's quality [34]. This innovative strategy marks a notable stride forward in object detection, particularly in the challenging milieu of minuscule objects or complex backdrops. The WIoU methodology adeptly modulates its focus across different samples, mitigating the disproportionate impact of low-quality or extreme cases. By integrating the concept of "outliers," WIoU offers a refined and effective means of assessing prediction quality, which is especially beneficial in datasets characterized by high variability or inconsistent quality. The formula for WIoU is outlined below:

$$L_{WIoU} = R_{WIoU} L_{IoU} \quad (7)$$

$$R_{WIoU} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)}\right) \quad (8)$$

$$L_{IoU} = 1 - IoU \quad (9)$$

In the modified loss function L_{WIoU} used in YOLOv8, the R_{WIoU} term scales the IoU loss to focus on sample quality. The variables within R_{WIoU} are defined as follows:

x and y : The expected bounding box's center coordinates.

x_{gt} and y_{gt} : The coordinates of the ground truth bounding box's center.

w_g and h_g : The ground truth bounding box's height and width.

The term R_{WIoU} is calculated using the exponential function to emphasize the influence of anchor frames closer to the ground truth by reducing the impact of those further away. The IoU loss L_{IoU} is the complement of the IoU between the predicted and ground truth bounding boxes, which inherently focuses on the overlap quality. In Equation (9), the loss function is modified to give priority to anchor frames of average quality and reduce the impact of extreme samples, improving the generalization capacity and overall performance of the model. By making this modification, the model is guaranteed to focus more on normal, high-probability samples and less on outliers, resulting in a more stable and efficient gradient throughout training.

4. Experiments and Analysis of Results

4.1. Experimental Setup and Dataset

4.1.1. Dataset

In the field of machine vision, the micro-capacitor surface defect (MCSD) dataset was meticulously compiled for the inspection and analysis of minute defects in various types of capacitors. This dataset is unique in its focus on the intricacies of tiny capacitors, encompassing a wide range of defect types and capacitor sizes, suitable for rigorous AVI applications.

Our MCSD dataset contained 1358 high-resolution images of four unique types of miniature capacitors, each varying in size and defect characteristics. This dataset was rich, with 2450 meticulously annotated ground truth boxes, showcasing a wide variety of defects. For each capacitor type, the size, type, and specific defects are detailed in the table below.

Every image in the MCSD dataset was carefully marked with precise bounding boxes, identifying the exact location and extent of each defect. This dataset presents an in-depth view of defect distribution, considering both size and position, as detailed in Table 1. Sourced from real industrial settings, the MCSD dataset is highly relevant and applicable for practical use.

Table 1. Representative samples from the micro-capacitor surface defect (MCSD) dataset.

Capacitor Size	offsets		scratches		breakage		
10 mm							
5 mm	lead warpage	mushy spots		magnet indentation		chipping	
0.4 mm	spillage		scratches		deformation		soiling
2 mm	breakage	scratches		leakage		pits	spillage

In terms of experimentation, we split the dataset into training and validation sets at a 9:1 ratio. This split was designed to optimize model training and evaluate performance effectively. Such a division provided a thorough understanding of the model's ability to detect a broad spectrum of defects across various sizes of capacitors.

4.1.2. Experimental Setup

This experiment used an NVIDIA GeForce RTX 4090 GPU, Python 3.8 programming language, Pytorch 2.0.0, and CUDA 11.8. The image input size was 640×640 , and the Mosaic method was used to enhance the data. The number of training rounds was 400, and the initial learning rate was set to 0.01. The starting learning rate was set to 0.01 in all experiments in this article, while the batch size and number of training rounds were set to 32 and 300, respectively [35,36].

4.1.3. Evaluation Metrics

In assessing the performance of micro-capacitor defect detection, we considered several metrics:

Precision: This is the product of the number of successfully discovered defects, or true positive detections (TP), and the total number of false positives (FP), or occurrences of false positives that were mistakenly labeled as defects.

Recall: This metric calculates the number of true positives divided by the sum of true positives and false negatives FN , where false negatives represent defects that the model failed to detect.

Average precision (AP): AP was derived by summing the precision at each instance P_{ri} and dividing by the sum of all ranks (r), which collectively assess the model's precision at different levels of recall or detection confidence.

Mean average precision (mAP): By dividing the AP by the total number of classes (num_classes), one can calculate the mean of the AP values for all classes (mAP), which provides a unified performance metric that illustrates the accuracy of the model for all defect types.

These measures are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (11)$$

$$\text{AP} = \frac{\sum P_{ri}}{\sum r} \quad (12)$$

$$\text{mAP} = \frac{\text{AP}}{\text{num_classes}} \quad (13)$$

4.2. Comparative Experiment of Attention Module and Loss Function

To ascertain the efficacy of various attention mechanisms and loss functions in augmenting the YOLOv8 algorithm for our dataset, this study meticulously crafted a suite of comparative experiments. These experiments meticulously integrated a variety of attention mechanisms and loss functions into the YOLOv8 framework to rigorously assess their capability to tackle practical challenges.

Table 2 presents a comparative analysis of four distinct attention mechanisms: SE, CBAM, CA, and Simam. Furthermore, Table 3 focuses on six divergent loss functions: CIoU, DIoU, GIoU, SIoU, EIoU, and WIoU [22]. Both tables evaluate their respective methods against three common metrics: parameters (M), frame rate (Fps), and mean average precision (mAP@0.5).

Table 2. Comparison of results with different loss functions.

Model	IoU	Parameters/M	Fps	mAP@0.5 (%)
YOLOv8n	CIoU	3.01	96	86.3
	DIoU	3.01	101	85.8
	GIoU	3.01	98	87.1
	SIoU	3.01	103	87.0
	EIoU	3.01	96	86.8
	WIoU	3.01	99	88.6

Table 3. Comparison of results with different attention modules.

Model	Attention Mechanisms	Parameters/M	Fps	mAP@0.5 (%)
YOLOv8n	SE	3.01	109	87.4
	CBAM	3.08	104	87.8
	CA	3.04	101	89.1
	Simam	3.01	94	90.3

After evaluating the performance of various attention mechanisms, the Simam attention mechanism emerged as the superior choice, achieving an optimal mean average precision (mAP@0.5) score of 0.903. However, it processed frames at a slightly reduced rate of 94 frames per second, indicating a trade-off between high accuracy and the system's real-time response. When it comes to selecting a loss function, the WIoU loss function stood out with a leading mAP@0.5 score of 0.886 while facilitating real-time processing capabilities at

99 frames per second. This balance suggests that WIoU is well-suited for real-time object detection systems, striking an effective balance between enhancing detection accuracy and maintaining processing speed.

Taking into account both precision and computational efficiency, we chose to integrate the Simam attention mechanism and WIoU loss function into YOLOv8, a decision supported by our comprehensive ablation study. This combination not only enhanced the detection accuracy in our dataset but also met the stringent real-time performance requirements of industrial applications [37]. The ablation study further corroborated the efficacy of both the Simam attention mechanism and WIoU loss function in our specific dataset. Looking ahead, our future endeavors will focus on deploying these methodologies across broader and more heterogeneous datasets, aiming to confirm their scalability and effectiveness in a variety of real-world scenarios.

4.3. Ablation Experiment

Table 4 outlines the ablation studies we conducted, which were designed to thoroughly assess the improvements added to the YOLOv8 model. Specifically, these enhancements included the integration of the SimAM attention mechanism, the BiFPN architecture, and the Wise-IoU (WIoU) loss function. The results of these experiments offer clear, quantifiable evidence of the performance boosts each modification brings to the table.

Table 4. Comparison of results from ablation experiments.

Model	Parameters/M	FPS/f.s-1	mAP@0.5
YOLOv8n	3.01	96	86.3
YOLOv8n + Simam	3.01	94	90.3
YOLOv8n + BiFPN	3.37	83	90.0
YOLOv8n + WIoU	3.01	99	88.6
YOLOv8n + Simam + WIoU	3.01	93	93.0
YOLOv8n + Simam + BiFPN + WIoU	3.37	78	95.8

The YOLOv8n framework's application of the SimAM attention mechanism produced a 2.3% increase in mean average precision (mAP) over the baseline YOLOv8 model, indicating the mechanism's efficacy in enhancing the detection of minute and complex information. A comparable 2.3% gain in mAP over the baseline was obtained by integrating the BiFPN, which improved the feature fusion capabilities of the model. This emphasizes how the BiFPN helps enhance feature representation, which helps provide more precise object localization and recognition. With a 3.3% gain in mAP over the initial model, the switch from the conventional CIoU loss function to the WIoU loss function showed a noteworthy improvement. This modification shows how well the WIoU loss function optimizes the model for improved generalization [38]. The SimAM attention mechanism and the WIoU loss function resulted in a 2.7% increase in mAP over the baseline in the model. This combination suggests a synergistic impact that improves the overall performance of the model.

Most impressively, the integration of SimAM, BiFPN, and WIoU into the YOLOv8n framework boosted the mAP by 9.5% over the baseline model. The combination of these three enhancements substantially elevated performance, demonstrating their collective impact in creating a robust model capable of detecting minute and low-contrast defects with high precision.

4.4. Comparison with Other Algorithms

According to the empirical findings displayed in Table 5, the suggested small-scale capacitor identification model has significant benefits in comparison with the current cutting-edge target detection algorithms, such as SSD, Faster RCNN, and YOLO versions 5n, 7-tiny, and 8. The model introduced in this paper, referred to as "Ours", achieved a noteworthy enhancement in mean average precision (mAP) at a threshold of 0.5—recording

a significant 95.8% as opposed to the original YOLOv8's 86.3%. This improvement marks a 9.5% increase in mAP, which is substantial considering the operational parameters of the model. Our model not only excelled in precision but also demonstrated an impressive frame rate of 78 FPS, which, while slightly lower than that of YOLOv7-tiny, is considerably higher than that of the more computationally demanding models like Faster RCNN. This balance of speed and accuracy indicates that our model is not only more precise but also operationally efficient. The marginal increase in the number of parameters from YOLOv8 to our model (from 3.01 M to 3.37 M) is a modest trade-off for the significant gains in detection accuracy. In comparison with other models' recall rates, our model demonstrated an approximate 2.5% improvement over the highest-performing model in the category, Gold-YOLO, which has a recall rate of 87.9%. This enhancement signifies that our model is more proficient in correctly identifying a greater number of true positives, i.e., defects. This capability is particularly crucial in the context of micro-capacitor defect detection, as it ensures the comprehensive identification of all potential defects. The slight increment in model complexity did not impede the processing speed, as evidenced by the fps metric, which is crucial for real-time detection tasks. Other updated models' performance on our MCSD dataset demonstrated a minor reduction from their findings on their respective originally selected public datasets, aside from the foundational models in the YOLO series. This discrepancy in performance might be explained by these models' restricted capacity for generalization, especially in the narrow field of micro-capacitor defect detection. These restrictions draw attention to the difficulty of modifying current models to fit extremely precise and subtle jobs, highlighting the necessity for customized methods in machine vision applications meant for complex micro-capacitor defect detection.

Table 5. Comparison of results from different algorithms.

Model	Parameters/M	FPS/f.s-1	Precision/P	Recall/R	mAP@0.5
SSD [39]	26.29	63	72.8	66.5	70.1
Faster RCNN [40]	134.08	16	80.1	72.4	76.4
YOLOv5n [41]	7.05	58	86.7	86.8	88.8
TPH-YOLOv5 [42]	46.3	61	92.7	82.3	88.4
YOLOv7-tiny [43]	6.05	120	73.5	81.2	81.8
YOLOv8n [44]	3.01	96	90.3	80.7	86.3
Gold-YOLO [45]	5.61	95	83	87.9	83.5
Ours	3.37	78	97.9	91.2	95.8

These discoveries hold manifold implications. Primarily, the substantial increase in mAP signifies a significant augmentation in the model's capacity to identify minute capacitive anomalies—a pivotal facet in ensuring quality control within the realm of electronics manufacturing. Second, the preservation of computational efficiency, i.e., the steadfastness in computation and fps, underscores the model's potential for real-time applications. Sustaining this equilibrium between high detection accuracy and operational efficiency stands as a linchpin for practical implementation in industrial settings.

Moreover, our model evinces robustness and efficacy in grappling with the intricacies entailed in the detection of minute capacitor defects compared with established models, such as SSD, Faster RCNN, YOLOv5n, and YOLOv7-tiny. This facet bears paramount significance in the sphere of electronic component fabrication, wherein the discernment of the tiniest imperfections assumes a pivotal role in assuring the reliability and performance of the entire system.

Furthermore, the detailed comparison presented in Table 6 elucidates the superior performance of our model over a range of defect types. Specifically, our model demonstrated exceptional precision in identifying 'B-leakage' and 'B-pit' defects, with an AP of 97.3% and 99.9% respectively, eclipsing the other YOLO variants by a significant margin. These findings are not just statistical victories; they translate into tangible benefits in the manufac-

turing process, in which early and accurate defect detection is crucial to maintaining the integrity of the production line.

Table 6. Comparative analysis of defect detection performance across YOLO models (AP %).

Model	4 B-Leakage	6 B-Pit	9 C-Mushy Spot	12 C-Chipping	14 D-Defect	16 D-Soiling
YOLOv5n	82.4	95.0	84.3	89.9	75.0	82.4
YOLOv7-tiny	66.7	62.5	71.3	82.1	60.5	63.9
YOLOv8n	92.5	72.1	86.1	88.9	42.0	71.9
Ours	97.3	99.9	91.0	91.3	75.5	95.6

In instances of ‘C-mushy spot’ and ‘C-chipping’ defects, our model again surpassed the YOLOv5n and YOLOv7-tiny, with APs of 91.0% and 91.3%, respectively. This high level of accuracy in defect detection ensures that even the most subtle irregularities are captured, which is vital for the longevity and reliability of electronic components. For ‘D-defect’ and ‘D-soiling’, the improved model maintained its lead with an AP of 75.5% and 95.6%. These scores are not only reflective of the model’s ability to detect a wide variety of defect types but also its adaptability to different defect characteristics, which is essential for a comprehensive quality control system.

4.5. Visualization of the Results

Figure 5 shows the visualization of defect detection on the MCSD dataset for a qualitative assessment. Each row in the figure indicates a different type of defect. The outputs from YOLOv7-tiny, YOLOv5n, and YOLOv8 are displayed in the third, fourth, and fifth columns, respectively, and the first column displays the findings from our model and three comparable algorithms.

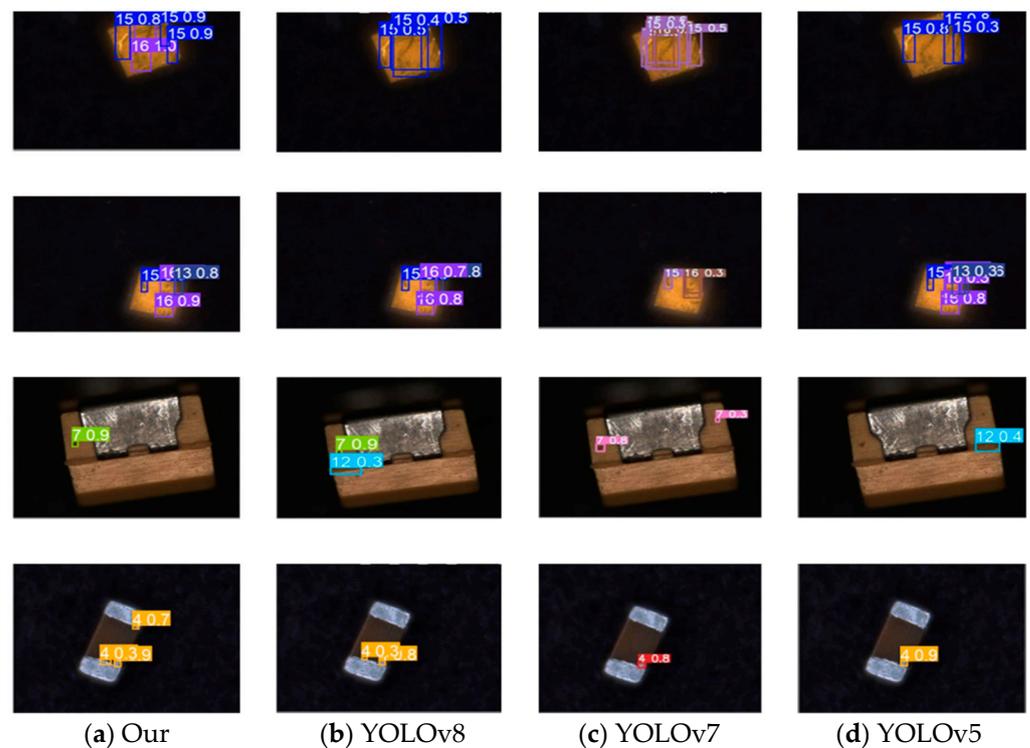


Figure 5. Comparison of results of different models on the MCSD dataset.

Our model excelled notably in pinpointing small and subtle defects, as can be seen in the first column, particularly in challenging scenarios, such as edge anomalies or defects with contours that are not clearly defined against their backgrounds. This proficiency

is evident in the rows showcasing these complex defect categories, in which our model consistently outperformed the others in detection clarity and accuracy.

The second, third, and fourth columns validate the comparative performance of the other algorithms. While YOLOv5n, YOLOv7-tiny, and YOLOv8 exhibited varying degrees of success, they each demonstrate certain limitations in detecting less conspicuous defects—a critical aspect in which our model demonstrates its robustness.

5. Conclusions

In the domain of automatic visual inspection (AVI) for miniature capacitor quality control, the accurate detection and characterization of small-sized defects remains a formidable challenge. These challenges stem from the small size and limited sample availability of defective micro-capacitors, leading to issues such as reduced detection accuracy and increased false-negative rates in existing inspection methods. In response to these issues, this research endeavored to introduce an enhanced YOLOv8 model algorithm, specifically designed to elevate the accuracy of detecting minuscule capacitive defects in industrial production settings. Our novel algorithm significantly augmented the model's capability to identify intricate, small-sized targets by leveraging the SimAM attention module and integrating the BiFPN (bidirectional feature pyramid network) structure. The BiFPN, with its sophisticated approach to feature fusion and layer connectivity, significantly bolstered the model's efficiency in discerning fine details, a key aspect in the realm of tiny defect detection.

Furthermore, we substituted the Wise-IoU loss function for the traditional CIoU loss function, a calculated step that successfully lessened the negative influence of anomalous samples during model training. This enhanced the model's generalization capabilities, leading to a fortified detection performance overall. Our experimental results, validated using the micro-capacitor surface defect (MCSD) dataset comprising 1358 images representing four distinct types of micro-capacitor defects, speak volumes about the effectiveness of our approach, especially evident in the significant uptick in the mean average precision (mAP) metric. Impressively, we achieved this without piling on computational costs or sacrificing frame rate. This underscores our model's robustness and its competitive edge in pinpointing minute capacitive defects in industrial processes.

We aim to broaden the scope of the MCSD dataset in future work to include a wider variety of faults. We are also dedicated to continuously improving and fine-tuning our system. We are optimistic that with persistent enhancements, our model will prove invaluable across a broad spectrum of practical applications, substantially elevating accuracy and efficiency in industrial quality control.

Although our proposed enhanced YOLOv8 model shows significant progress in the detection of defects in miniature capacitors, there are still some limitations. One of the main issues is the reliance on the MCSD dataset. Although this dataset contains a wide range of defect types, it may not be a comprehensive representation of the kinds of defects that can occur in various industrial environments, which may limit the applicability and effectiveness of the model when confronted with environments with unknown defect types or deviations from the parameters of the dataset. Furthermore, although replacing the Wise-IoU loss function with the CIoU loss function yielded positive results in our tests, the generalization of this approach to all possible industrial application scenarios needs to be further explored.

Author Contributions: Conceptualization, N.L. and Z.Z.; Methodology, N.L. and Z.Z.; Software, N.L.; Validation, N.L. and T.Y.; Formal analysis, N.L.; Investigation, N.L. and T.Y.; Resources, N.L.; Data curation, N.L.; Writing—original draft preparation, T.Y.; Writing—review and editing, T.Y.; Visualization, Z.Z.; Supervision, C.G. and P.Z.; Project administration, C.G. and P.Z.; Funding acquisition, C.G. and P.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 62075031.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data provided in this study are available upon request from the corresponding author. The data are not available to the public due to the confidentiality stage of the related R&D products.

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

References

1. Liao, Y.; Najafi-Haghi, Z.P.; Wunderlich, H.-J.; Yang, B. Efficient and Robust Resistive Open Defect Detection Based on Un-supervised Deep Learning. In Proceedings of the 2022 IEEE International Test Conference (ITC), Anaheim, CA, USA, 23–30 September 2022; pp. 185–193.
2. Zhou, Y.; Yuan, M.; Zhang, J.; Ding, G.; Qin, S. Review of vision-based defect detection research and its perspectives for printed circuit board. *J. Manuf. Syst.* **2023**, *70*, 557–578. [\[CrossRef\]](#)
3. Liu, B.; Wang, H.; Wang, Y.; Zhou, C.; Cai, L. Lane Line Type Recognition Based on Improved YOLOv5. *Appl. Sci.* **2023**, *13*, 10537. [\[CrossRef\]](#)
4. Nathan, L.P.A.; Hemamalini, R.R.; Jeremiah, R.J.R.; Partheeban, P. Review of condition monitoring methods for capacitors used in power converters. *Microelectron. Reliab.* **2023**, *145*, 115003. [\[CrossRef\]](#)
5. Chai, J.; Zeng, H.; Li, A.; Ngai, E.W. Deep Learning in Computer Vision: A Critical Review of Emerging Techniques and Application Scenarios. *Mach. Learn. Appl.* **2021**, *6*, 100134. [\[CrossRef\]](#)
6. Soori, M.; Arezoo, B.; Dastres, R. Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cogn. Robot.* **2023**, *3*, 54–70. [\[CrossRef\]](#)
7. Srivastava, S.; Divekar, A.V.; Anilkumar, C.; Naik, I.; Kulkarni, V.; Pattabiraman, V. Comparative analysis of deep learning image detection algorithms. *J. Big Data* **2021**, *8*, 66. [\[CrossRef\]](#)
8. Zhang, Q.; Zhang, H.; Lu, X. Adaptive Feature Fusion for Small Object Detection. *Appl. Sci.* **2022**, *12*, 11854. [\[CrossRef\]](#)
9. Mao, R.; Wang, Z.; Li, F.; Zhou, J.; Chen, Y.; Hu, X. GSEYOLOX-s: An Improved Lightweight Network for Identifying the Severity of Wheat Fusarium Head Blight. *Agronomy* **2023**, *13*, 242. [\[CrossRef\]](#)
10. Yu, X.; Yu, Q.; Mu, Q.; Hu, Z.; Xie, J. MCAW-YOLO: An Efficient Detection Model for Ceramic Tile Surface Defects. *Appl. Sci.* **2023**, *13*, 12057. [\[CrossRef\]](#)
11. Zhang, Y.-F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **2022**, *506*, 146–157. [\[CrossRef\]](#)
12. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. *arXiv* **2023**, arXiv:2301.10051.
13. Tian, B.; Chen, H. Remote Sensing Image Target Detection Method Based on Refined Feature Extraction. *Appl. Sci.* **2023**, *13*, 8694. [\[CrossRef\]](#)
14. Prunella, M.; Scardigno, R.M.; Buongiorno, D.; Brunetti, A.; Longo, N.; Carli, R.; Dotoli, M.; Bevilacqua, V. Deep Learning for Automatic Vision-Based Recognition of Industrial Surface Defects: A Survey. *IEEE Access* **2023**, *11*, 43370–43423. [\[CrossRef\]](#)
15. Bhatt, P.M.; Malhan, R.K.; Rajendran, P.; Shah, B.C.; Thakar, S.; Yoon, Y.J.; Gupta, S.K. Image-Based Surface Defect Detection Using Deep Learning: A Review. *J. Comput. Inf. Sci. Eng.* **2021**, *21*, 040801. [\[CrossRef\]](#)
16. Yin, Y.; Li, H.; Fu, W. Faster-YOLO: An accurate and faster object detection method. *Digit. Signal Process.* **2020**, *102*, 102756. [\[CrossRef\]](#)
17. Thakuria, A.; Erkinbaev, C. Improving the network architecture of YOLOv7 to achieve real-time grading of canola based on kernel health. *Smart Agric. Technol.* **2023**, *5*, 100300. [\[CrossRef\]](#)
18. Li, G.; Zhao, S.; Zhou, M.; Li, M.; Shao, R.; Zhang, Z.; Han, D. YOLO-RFF: An Industrial Defect Detection Method Based on Expanded Field of Feeling and Feature Fusion. *Electronics* **2022**, *11*, 4211. [\[CrossRef\]](#)
19. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
20. Liu, T.; Bao, J.; Wang, J.; Zhang, Y. A Coarse-Grained Regularization Method of Convolutional Kernel for Molten Pool Defect Identification. *J. Comput. Inf. Sci. Eng.* **2020**, *20*, 021005. [\[CrossRef\]](#)
21. Zhang, J.; Zhou, H.; Niu, Y.; Lv, J.; Chen, J.; Cheng, Y. CNN and multi-feature extraction based denoising of CT images. *Biomed. Signal Process. Control* **2021**, *67*, 102545. [\[CrossRef\]](#)
22. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [\[CrossRef\]](#)
23. Wang, D.; Wang, X.; Wang, L.; Li, M.; Da, Q.; Liu, X.; Gao, X.; Shen, J.; He, J.; Shen, T.; et al. MedFMC: A Real-World Dataset and Benchmark for Foundation Model Adaptation in Medical Image Classification. *arXiv* **2023**, arXiv:2306.09579.
24. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
25. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. Simam: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In Proceedings of the 38th International Conference on Machine Learning, PMLR, Virtual, 18–24 July 2021; pp. 11863–11874.
26. Qian, J.; Lin, J.; Bai, D.; Xu, R.; Lin, H. Omni-Dimensional Dynamic Convolution Meets Bottleneck Transformer: A Novel Improved High Accuracy Forest Fire Smoke Detection Model. *Forests* **2023**, *14*, 838. [\[CrossRef\]](#)

27. Tian, Z.; Yang, F.; Qin, D. An Improved New YOLOv7 Algorithm for Detecting Building Air Conditioner External Units from Street View Images. *Sensors* **2023**, *23*, 9118. [[CrossRef](#)] [[PubMed](#)]
28. Bai, W.; Zhao, J.; Dai, C.; Zhang, H.; Zhao, L.; Ji, Z.; Ganchev, I. Two Novel Models for Traffic Sign Detection Based on YOLOv5s. *Axioms* **2023**, *12*, 160. [[CrossRef](#)]
29. Li, J.; Tian, Y.; Chen, J.; Wang, H. Rock Crack Recognition Technology Based on Deep Learning. *Sensors* **2023**, *23*, 5421. [[CrossRef](#)]
30. Zhang, Y.; Ni, Q. A Novel Weld-Seam Defect Detection Algorithm Based on the S-YOLO Model. *Axioms* **2023**, *12*, 697. [[CrossRef](#)]
31. Qu, Z.; Gao, L.-Y.; Wang, S.-Y.; Yin, H.-N.; Yi, T.-M. An improved YOLOv5 method for large objects detection with multi-scale feature cross-layer fusion network. *Image Vis. Comput.* **2022**, *125*, 104518. [[CrossRef](#)]
32. Chiley, V.; Thangarasa, V.; Gupta, A.; Samar, A.; Hestness, J.; De Coste, D. RevBiFPN: The Fully Reversible Bidirectional Feature Pyramid Network. Proceedings of Machine Learning and Systems. *arXiv* **2023**, arXiv:2206.14098.
33. Zheng, Z.; Wang, P.; Ren, D.; Liu, W.; Ye, R.; Hu, Q.; Zuo, W. Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation. *IEEE Trans. Cybern.* **2021**, *52*, 8574–8586. [[CrossRef](#)]
34. Cho, Y.-J. Weighted Intersection over Union (wIoU): A New Evaluation Metric for Image Segmentation. *arXiv* **2023**, arXiv:2107.09858.
35. Li, L.; Yang, L.; Zeng, Y. Improving Sentiment Classification of Restaurant Reviews with Attention-Based Bi-GRU Neural Network. *Symmetry* **2021**, *13*, 1517. [[CrossRef](#)]
36. Luo, M.; Liu, X.; Huang, W. Gaze Estimation Based on Neural Network. In Proceedings of the 2019 IEEE 2nd International Conference on Electronic Information and Communication Technology (ICEICT), Harbin, China, 20–22 January 2019; pp. 590–594.
37. Xiao, Z.; Wan, F.; Lei, G.; Xiong, Y.; Xu, L.; Ye, Z.; Liu, W.; Zhou, W.; Xu, C. FL-YOLOv7: A Lightweight Small Object Detection Algorithm in Forest Fire Detection. *Forests* **2023**, *14*, 1812. [[CrossRef](#)]
38. Lei, F.; Tang, F.; Li, S. Underwater Target Detection Algorithm Based on Improved YOLOv5. *J. Mar. Sci. Eng.* **2022**, *10*, 310. [[CrossRef](#)]
39. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision—ECCV 2016; Lecture Notes in Computer Science*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer: Cham, Switzerland, 2016; Volume 9905, pp. 21–37, ISBN 978-3-319-46447-3.
40. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems 28 (NIPS 2015), Montreal, QC, Canada, 7–12 December 2015; Volume 28, pp. 91–99.
41. Wang, J.; Chen, Y.; Dong, Z.; Gao, M. Improved YOLOv5 network for real-time multi-scale traffic sign detection. *Neural Comput. Appl.* **2022**, *35*, 7853–7865. [[CrossRef](#)]
42. Wang, C.; He, W.; Nie, Y.; Guo, J.; Liu, C.; Han, K.; Wang, Y. Gold-YOLO: Efficient Object Detector via Gather-and-Distribute Mechanism. *arXiv* **2023**, arXiv:2309.11331v5.
43. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
44. Reis, D.; Kupec, J.; Hong, J.; Daoudi, A. Real-Time Flying Object Detection with YOLOv8. *arXiv* **2023**, arXiv:2305.09972.
45. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios. *arXiv* **2021**, arXiv:2108.11539.

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.