Article

MFAD-RTDETR: A Multi-Frequency Aggregate Diffusion Feature Flow Composite Model for Printed Circuit Board Defect Detection

Zhihua Xie [1,2,*] and Xiaowei Zou [1,2]

1   Key Lab of Optic-Electronic and Communication, Jiangxi Science and Technology Normal University, Nanchang 330038, China; 1020101062@jxstnu.edu.cn
2   Nanchang Key Laboratory of Failure Perception & Green Energy Materials Intelligent Manufacturing, Nanchang 330038, China
*   Correspondence: xie_zhihua68@aliyun.com

Abstract: To address the challenges of excessive model parameters and low detection accuracy in printed circuit board (PCB) defect detection, this paper proposes a novel PCB defect detection model based on the improved RTDETR (Real-Time Detection, Embedding and Tracking) method, named MFAD-RTDETR. Specifically, the proposed model introduces the designed Detail Feature Retainer (DFR) into the original RTDETR backbone to capture and retain local details. Subsequently, based on the Mamba architecture, the Visual State Space (VSS) module is integrated to enhance global attention while reducing the original quadratic complexity to a linear level. Furthermore, by exploiting the deformable attention mechanism, which dynamically adjusts reference points, the model achieves precise localization of target defects and improves the accuracy of the transformer in complex visual tasks. Meanwhile, a receptive field synthesis mechanism is incorporated to enrich multi-scale semantic information and reduce parameter complexity. In addition, the scheme proposes a novel Multi-frequency Aggregation and Diffusion feature composite paradigm (MFAD-feature composite paradigm), which consists of the Aggregation Diffusion Fusion (ADF) module and the Refiner Feature Composition (RFC) module. It aims to strengthen features with fine-grained awareness while preserving a certain level of global attention. Finally, the Wise IoU (WIoU) dynamic nonmonotonic focusing mechanism is used to reduce competition among high-quality anchor boxes and mitigate the effects of the harmful gradients from low-quality examples, thereby concentrating on anchor boxes of average quality to promote the overall performance of the detector. Extensive experiments are conducted on the PCB defect dataset released by Peking University to validate the effectiveness of the proposed model. The experimental results show that our approach achieves the 97.0% and 51.0% performance in mean Average Precision (mAP)@0.5 and mAP@0.5:0.95, respectively, which significantly outperforms the original RTDETR. Moreover, the model reduces the number of parameters by approximately 18.2% compared to the original RTDETR.

Keywords: feature composite paradigm; fine-grained awareness; receptive field synthesis mechanism; RTDETR; defect detection

1. Introduction

With the more sophisticated electronic circuits and the increasing market demand for high-quality computing products, the print circuit board (PCB), as an essential electronic device, plays a crucial role in the stability of the entire electronic industry's application. In the process of PCB manufacturing, circuit defects are one of the key factors leading to performance degradation. The common defects of PCB mainly include six categories: missing holes, mouse bites, open circuits, short circuits, spurious signals, and stray coppers. Automatic PCB defect detection refers to the use of computer vision technology to detect these defects, which is crucial to ensure the quality of PCB products. However, the

defects present on PCB are typically small and subtle, such as tiny soldering contact issues, incomplete or missing solder balls, and minute flaws. These issues affect the connectivity and reliability of the circuit and are difficult to detect by traditional inspection methods or manual detection approaches. Therefore, it is necessary to explore rapid and accurate defect detection technologies in PCB production.

Automatic PCB defect detection can be primarily categorized into two groups based on features: traditional, manual methods and deep learning methods. As industrial defects often manifest as regions of abrupt pixel changes in images, it enables traditional manual algorithms like edge detection operators [1] to locate defect areas. Traditional edge detection operators include Prewitt, Sobel, and Canny. From the perspective of frequency domain analysis, abrupt defects typically appear as significant high-frequency parts in the spectrum. Transform-based methods such as Fourier transform [2], Gabor transform [3], and wavelet transform [4] usually convert the image into the frequency domain for defect detection. There are some studies, such as the work proposed by Chetverikov et al., that detect abrupt defects on textile surfaces based on texture orientation. Hou et al. implemented the Gabor wavelet transform to extract frequency domain information in coordination with support vector machines [5] and random forests [6] for automatic defect classification. Wu et al. proposed an algorithm for diagnosing printed circuit solder joint defects based on Back Propagation (BP) neural networks [7] and genetic algorithms [8]. Before classifying features, they introduced genetic algorithms to select and remove redundant features to avoid the overfitting problem of BP neural networks. However, these methods struggle to handle complex backgrounds or low signal-to-noise ratio defects and have high requirements for imaging conditions. When detecting small targets in complex backgrounds, these traditional machine learning approaches will yield a high false detection rate.

In recent years, deep learning networks play a vital role in improving the performance of PCB defect detection, due to their powerful learning capability and strong feature representation ability on relatively large datasets. Generally, the prevalent Convolutional Neural Network (CNN) architectures exhibit high accuracy and good generalization capability on defect representation. Concretely, Park et al. addressed the issue of CNN performance degradation caused by a small training image dataset, named MarsNet [9], which enhances the resolution of feature maps through improvements in the dilated residual network (DRN) [10]. Wu utilized the Mask RCNN [11] for effective classification and localization of solder joint defects, achieving promising results. Ross B. Girshick et al. introduced the Fast R-CNN [12] with a multitask learning strategy, sharing feature layers during training to promote model generalization and detection speed. Ding et al. presented TDD-Net [13], which builds upon Faster RCNN [14] by integrating Res-Net and feature pyramid networks (FPN) for multi-scale anchor-based detection of small defects on PCBs, demonstrating robustness and generalization of performance. Hu et al. updated Faster RCNN with GARPN [15] and ShuffleNetV2 [16] units to improve detection speed and mAP. Li et al. extended the FPN model through semantic fusion across high and low-level layers and introduced a focal loss function [17] to further boost detection performance. Although transformer models [18], proposed by Vaswani et al., show limited performance in end-to-end object detection due to high computational costs and challenges in achieving high accuracy for small object detection, the parallel execution of the transformer model with ResNet to optimize detection of small targets has broad applicability, but may introduce huge computational burdens unsuitable for high-efficiency and real-time industrial applications. YOLO architecture [19] achieves fast end-to-end object detection with a simple design, yet its detection accuracy and robustness may decrease in complex backgrounds or dense target scenarios. Notably, RTDETR [20], proposed by Zhao et al., performs real-time end-to-end detection based on transformers, effectively handling multi-scale features through internal scale interactions and cross-scale fusion, showing superior performance compared to equivalently scaled YOLO detectors. Based on the good performance of RTDETR on diverse object detection, the main motivation of this work is to explore an efficient yet lightweight deep model based on improved REDETR.

As can be gleaned from the above review, the deep learning approaches for PCB defect detection exhibit higher accuracy and better robustness compared to traditional machine learning methods. However, detecting small defect targets often involves significant shape variations, irregularities, and diverse manifestations under different conditions, thereby increasing the complexity of detection algorithms. To improve the precision of detecting small defects while minimizing model complexity and boosting operational efficiency, we propose a Multi-Frequency Aggregate Diffusion Feature Flow Composite Paradigm for PCB defect detection, abbreviated as MFAD-RTDETR. The main contributions are summarized as follows:

(1) In terms of fine-grained feature representation on complex PCB defects, a novel multi-frequency aggregation and diffusion feature composite paradigm (MFAD-feature composite paradigm) is designed to enhance fine-grained features while preserving global attention, which includes the Aggregation Diffusion Fusion (ADF) module and the Refiner Feature Composition (RFC) module.

(2) With regard to detail preservation for small defects, a Detail Feature Retainer (DFR) is developed to better capture and retain local feature details through adaptive point movement and gating mechanisms.

(3) Regarding efficient feature fusion on PCB defects with various scales, a receptive field synthesis mechanism is introduced to achieve effective fusion between different scale features, thereby obtaining rich multi-scale information and reducing parameter complexity.

Our code for this model is available on: https://github.com/ZouXiaowei-zxw/MFAD (accessed on 1 March 2024).

2. Methodology

In this section, the detailed framework of RTDETR is depicted firstly to present the main stages of PCB defect detection. This is followed by an elaboration of the overall framework of MFAD-RTDETR and seven specific modules updated in the proposed model.

2.1. RTDETR Network

The original RTDETR model is an efficient and powerful single-stage object detection framework [20]. It introduces a real-time end-to-end object detector that outperforms early models in terms of real-time performance, accuracy, and stability. It eliminates the need for post-processing, maintains stable inference speed without latency, and proposes an IoU-aware query selection algorithm, significantly enhancing model performance and providing an effective approach for initializing target queries. The core architecture of the RTDETR comprises three components: the backbone network, the hybrid encoder, and the transformer decoder equipped with an auxiliary prediction head. The detailed framework of RTDETR is shown in Figure 1.

(1) As shown in Figure 1, the backbone network is responsible for feature extraction and primarily comprises ConvBN modules and Basic Block modules. The ConvBN module, which includes a convolutional layer and a batch normalization layer, extends the network's receptive field. The Basic Block module, based on the ResNet architecture, consists of two convolutional layers and a residual connection, which effectively address the vanishing gradient problem and enhance the model's representational power and performance.

(2) The hybrid encoder serves as the network's feature fusion component, incorporating an attention-based intra-scale feature interaction (AIFI) module [20], which uses a single transformer layer to extract rich high-level semantic information and effectively capture relationships between conceptual entities within the image. The Cross-Scale Feature Fusion (CCFM) module employs a feature pyramid network (FPN) for feature fusion, with the fusion module consisting of two 1×1 convolutions and several RepBlocks, thereby significantly leveraging the integration of features across different scales. Finally, based on IoU-aware queries, a specific number of image features are selected from the encoder output sequence as initial target queries for the decoder.

(3) The decoder is equipped with an auxiliary prediction head, which iteratively optimizes the target queries to generate bounding boxes and confidence scores.
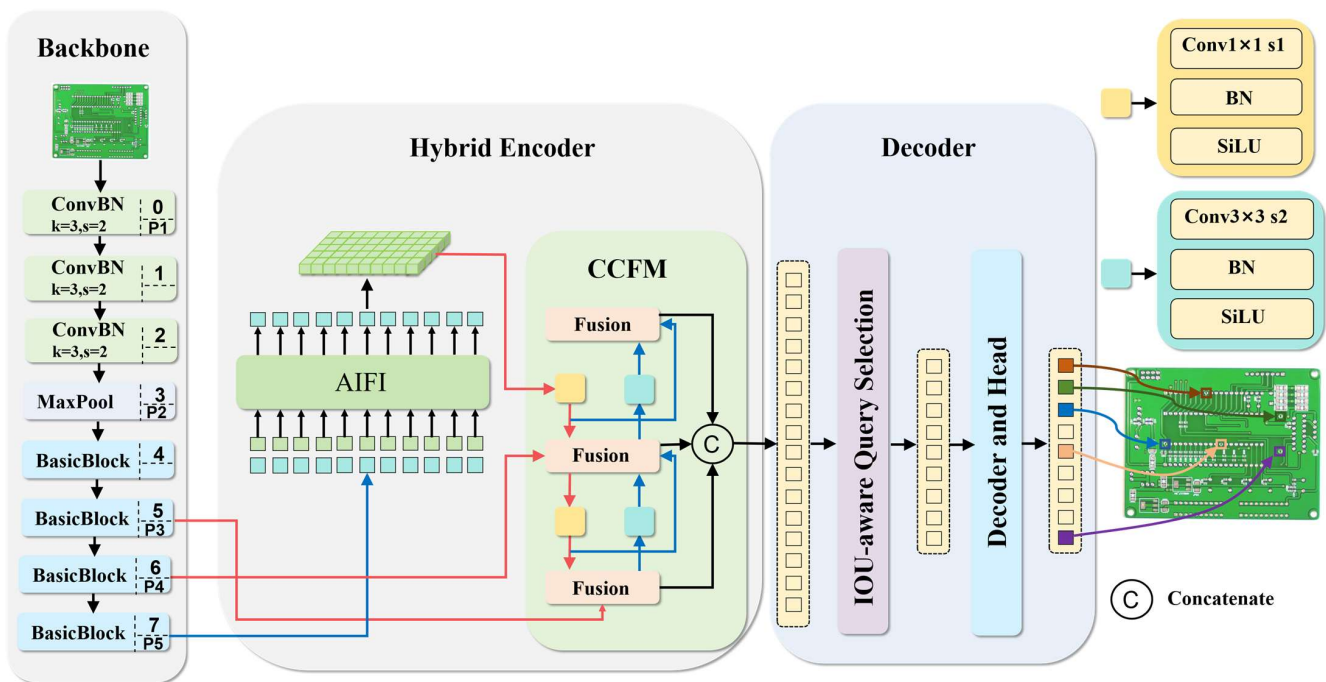

Figure 1. The framework of RTDETR.

2.2. The Proposed MFAD-RTDETR Model

This study aims to construct a lightweight MFAD-RTDETR model, which is shown in Figure 2. The main structure of MFAD-RTDETR, similar to that of RTDETR, also contains three main components: the backbone network, the hybrid encoder, and the decoder. As traditional feature extraction methods are insufficient in handling complex tasks of detecting small defects on PCBs, the proposed MFAD-RTDETR not only effectively improves attention to details but also minimizes network parameters. Briefly, compared with RTDETR, the main improvements of MFAD-RTDETR are the following seven aspects.

(1) The scheme first proposes a DFR module that employs an adaptive point movement mechanism [21] and a gating mechanism [22] to better capture and retain local feature details.

(2) Subsequently, the introduction of the VSS [23], centered on the Mamba architecture and the Cross-Scan Module (CSM), not only strengthens global attention but also reduces the original quadratic complexity to a linear level.

(3) The multi-frequency fusion module mainly comprises the FDASI and SSFF modules, which facilitate more effective feature extraction and multi-frequency fusion.

(4) Moreover, a deformable attention (DAttention) mechanism [24] is introduced to dynamically adjust the offsets of reference points, thereby achieving precise localization of defect regions. It significantly enhances the efficiency and accuracy of the transformer in handling complex visual tasks.

(5) Meanwhile, a receptive field synthesis mechanism has been introduced to employ the Dilated Reparam Block (DRB) [25]. This approach uses small convolutional kernels during the training phase, which can equivalently transform into nondilated layers with sparse larger kernels during the testing phase, thereby reducing the number of network parameters.

(6) Furthermore, a novel MFAD-feature composite paradigm is proposed. It consists of two parts: the first part is the ADF module, which aggregates different frequency feature flows and injects them into the improved Frequency and Dimension-Aware Selective

Integration (FDASI) Module. The FDASI module selectively highlights defect features in specific areas of the image. Then, the diffusion mechanism propagates these local details throughout the entire network, enabling the model to comprehensively capture essential features and structures. The second part is the RFC module, which employs the Scale Sequence Feature Fusion (SSFF) [26] strategy for initial shallow-to-deep fusion, enhancing the global attention while preserving detailed features. Additionally, a detail detection head P1 is introduced to improve the recognition of minute features.
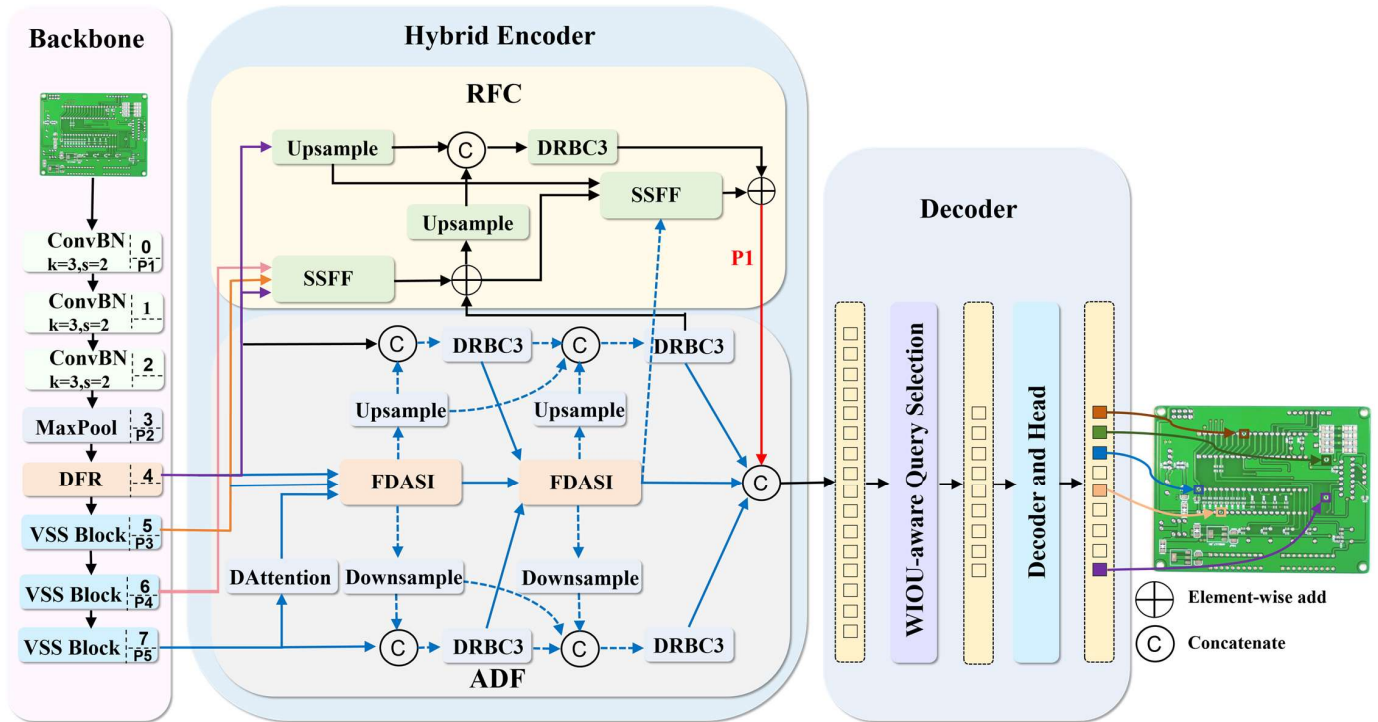

Figure 2. The framework of MFAD-RTDETR. With the blue dashed arrows, the diffusion mechanism effectively propagates these high-frequency details throughout the network.

(7) Finally, with the help of the dynamic nonmonotonic focusing mechanism of WIoU [27], which assesses anchor box quality based on "outlier," this approach minimizes competition among high-quality anchor boxes and mitigates the impact of adverse gradients originating from low-quality examples. It efficiently prioritizes anchor boxes of moderate quality, thus improving the comprehensive performance of the detector. As a result, this efficient and lightweight model remarkably boosts the detection accuracy of small defects and network complexity.

2.2.1. The DFR Module

When dealing with the diversity and complexity of data, conventional convolution operators may fail to capture detailed features effectively [28]. To cope with this problem, we innovatively combine SMPConv and CGLU to develop a DFR module. As shown in Figure 3a,b, this module mainly comprises the Self-Moving Point Convolution (SMPConv) module and the Convolutional GLU module. Concretely, the Self-Moving Point mechanism involves defining the SMP operator as the weighted sum of neighboring points for a query point $x \in R$, influenced by their distance to point X. Points beyond a certain distance (depending on the radius) do not affect the query point, enabling the model to capture complex patterns and variations in the data more accurately than the fixed-point representations. Each convolution layer has independently learnable parameters and shares position parameters to create arbitrarily large kernels while reducing the number of learnable parameters, thus achieving precise detail features. And the gating mechanism in CGLU consists of

two linear projection stages, one of which is activated by an activation function and then multiplied elementwise. By introducing depth-wise separable convolutions (DWConv), each token gains a gated channel attention based on nearest neighbor features, allowing more flexible selection and retention of local feature details. In addition, the DFR module is designed to accurately locate and extract fine features, enhancing the capability to extract fine-grained features.
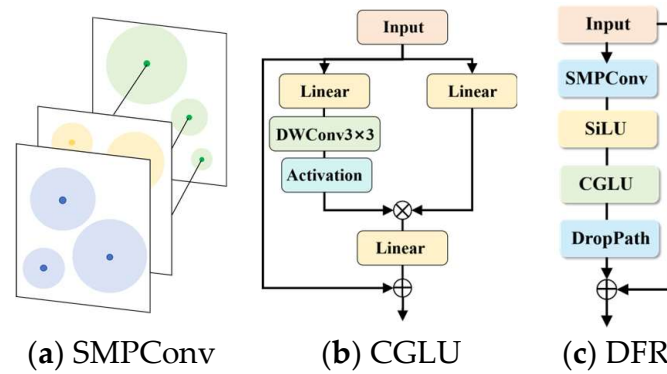

(**a**) SMPConv   (**b**) CGLU   (**c**) DFR

Figure 3. The structure diagrams of SMPConv, CGLU, and DFR modules.

2.2.2. VSS Block

Traditional transformer-based attention mechanisms suffer from high computational complexity and the limitations of fixed global attention. To address this, a VSS block based on the Mamba architecture is introduced, which effectively alleviates these issues. As shown in Figure 4, the Mamba (S6) architecture draws inspiration from the concept of a state machine.
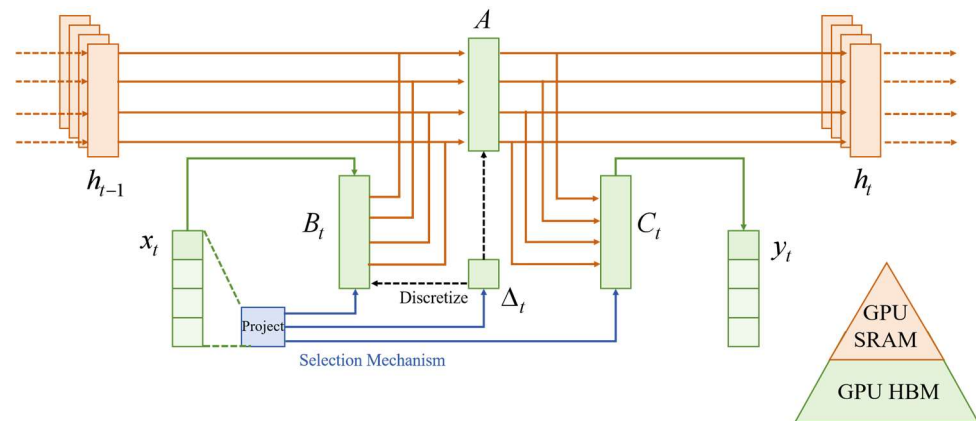

Figure 4. The Mamba (S6) architecture diagram.

The input x_t is first mapped through a selective mechanism to obtain B_t, $\triangle$ and C_t. Next, $\triangle$ and zero-order hold technology are used to discretize A and B_t, as shown in Figure 5. In the S6 block, the discretized B_t is multiplied by the input x_t, and the discretized A is multiplied by the previous state h_{t-1}. These products are then summed to obtain the new state h_t. Finally, the new state is multiplied by C_t to yield the output x_t. Furthermore, the Mamba architecture offers acceleration advantages through parallel scan operations and hardware-aware algorithms. The 2D-Selective-Scan (SS2D) module builds on the S6 foundation by adding a Cross-Scan Module (CSM), which decomposes images into small patches and scans them in a row and column order: from top-left to bottom-right, bottom-right to top-left, top-right to bottom-left, and bottom-left to top-right, generating scan sequences in four directions. These sequences are reshaped and merged into a new sequence, not only enhancing global attention and compressing hidden states but also

reducing complexity to a linear level. As shown in Figure 5, the VSS block leverages the SS2D module to effectively capture input features through long-range dependencies. It excels in global feature integration and significantly improves computational efficiency through its design, achieving a balance between accuracy and efficiency in the network.
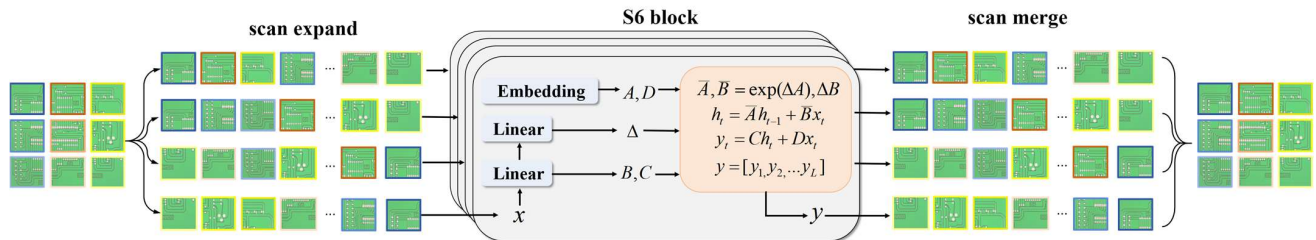

Figure 5. The structure diagram of the SS2D module.

2.2.3. Multi-Frequency Fusion Module

In the original RTDETR, the classic CCFM only fused features at different scales, which easily led to the loss of important details during down-sampling and caused information redundancy due to repeated stackings. To deal with this situation, the Dimension-Aware Selective Integration Module (DASI) [29] aims to represent detailed features. However, it preprocesses different scales independently, thus breaking the correlation between different scales. To address this issue, depicted in Figure 6, we alter course to approach it from the perspective of multi-frequency fusion. Natural images encompass a diverse frequency spectrum, with high and low frequencies playing different roles in capturing image features: the former focuses on local details, while the latter captures the global structures. To effectively integrate these features, this work designed the FDASI module. In the preprocessing stage of FDASI, the Frequency-Aware Fusion (FAF) module was developed. This module addresses three different frequency features by dividing them into three groups and inputting them pairwise into the Select Feature Fusion (SFF) module, which includes the Coordinate Attention (CA) module. After a single parallel multi-frequency operation, the features are multiplied elementwise with relatively high-frequency features to obtain fused high-frequency features and then combined with the original low-frequency features. This process results in fused features that contain rich multi-frequency information, providing more comprehensive input features for the subsequent modules. These features are subsequently fed into the DASI module, as described in Equations (1)–(4). In the equations, h_i, I_i, and u_i represent high-dimensional features, low-dimensional features, and the current layer's features, respectively. u_i' represents the selective aggregation result of each partition, F'_u denotes the output after channel dimension merging, $B(\cdot)$ and $\delta(\cdot)$ represent batch normalization (BN) and the rectified linear unit (RELU), and $\hat{F}_u$ is the final output. If the output of the activation function $\alpha > 0.5$, the model gives higher priority to fine-grained features. Otherwise, it emphasizes contextual features. The approach preserves rich multi-frequency information while paying attention to detail features.

$$\alpha = sigmod(u_i) \tag{1}$$

$$u_i' = \alpha I_i + (1 - \alpha)h_i \tag{2}$$

$$F'_u = [u_1', u_2', u_3', u_4'] \tag{3}$$

$$\hat{F}_u = \delta(B(Conv(F\prime u))) \tag{4}$$

To effectively integrate multi-frequency information from both deep and shallow layers, we introduced the Scale Sequence Feature Fusion (SSFF) module. As illustrated in Figure 7, this method can combine the high-dimensional information from deep feature maps with the low-dimensional information from shallow feature maps. Therefore, this

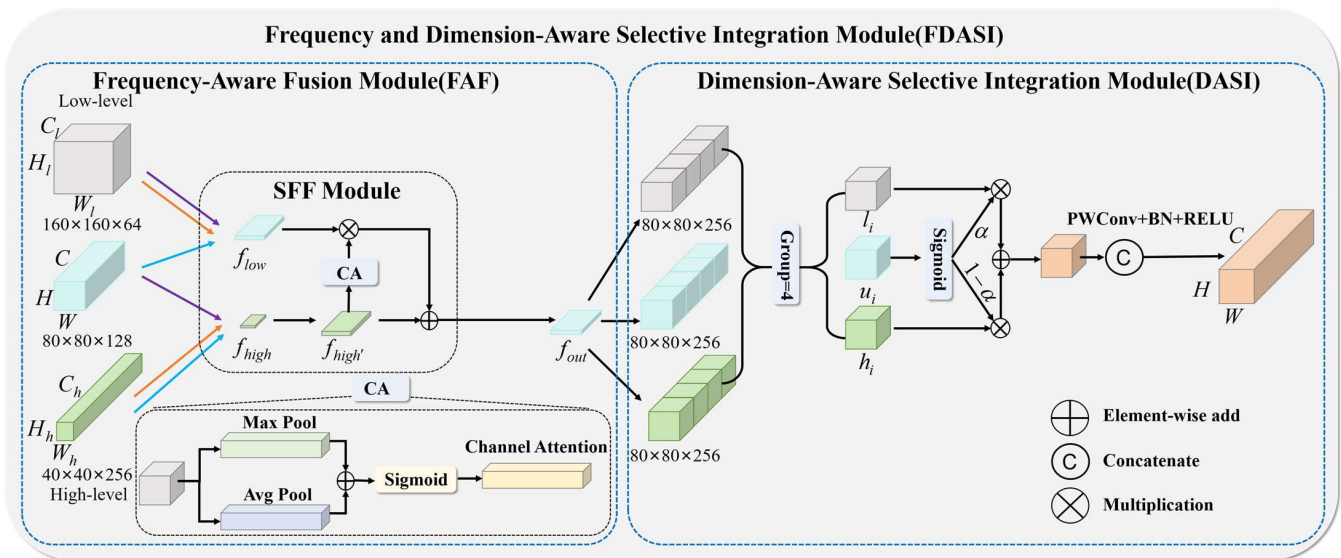integration achieves comprehensive capture and representation of defect features across different scales.

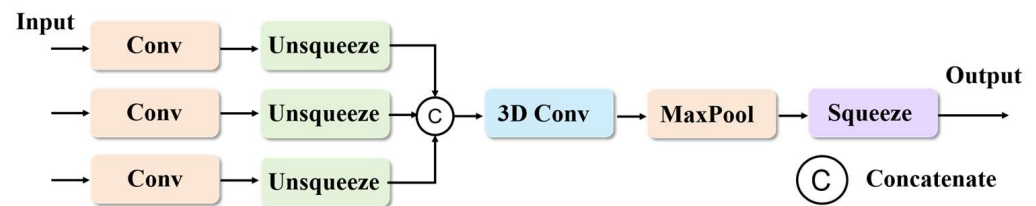**Figure 6.** The FDASI module structure diagram.

Figure 7. The SSFF module structure diagram.

2.2.4. DAttention Module

The transformer usually suffers from high computational complexity and low efficiency. As shown in Figure 8, the deformable attention module is introduced to address these issues. Specifically, DAttention improves the model's focus by generating multiple reference points and dynamic offsets, allowing it to concentrate on a small number of key areas. This significantly enhances the efficiency and accuracy of the visual transformer when handling complex visual tasks. Integrating DAttention into the internal scale feature interaction module enhances the network's ability to capture global dependencies and intricate local details within images. As shown in Figure 8a, a set of uniformly distributed reference points is first generated on the feature map, with their offsets learned through the offset network from query points q. Then, based on these deformable points, the corresponding $\widetilde{k}, \widetilde{v}$, and relative position bias are projected from the sampling features. This method dynamically optimizes the performance of output features in multi-head attention. Figure 8b shows the detailed structure of the offset network used to generate offsets for the deformable points. It achieves precise localization of target defect areas through dynamic adjustment of reference points' offsets.
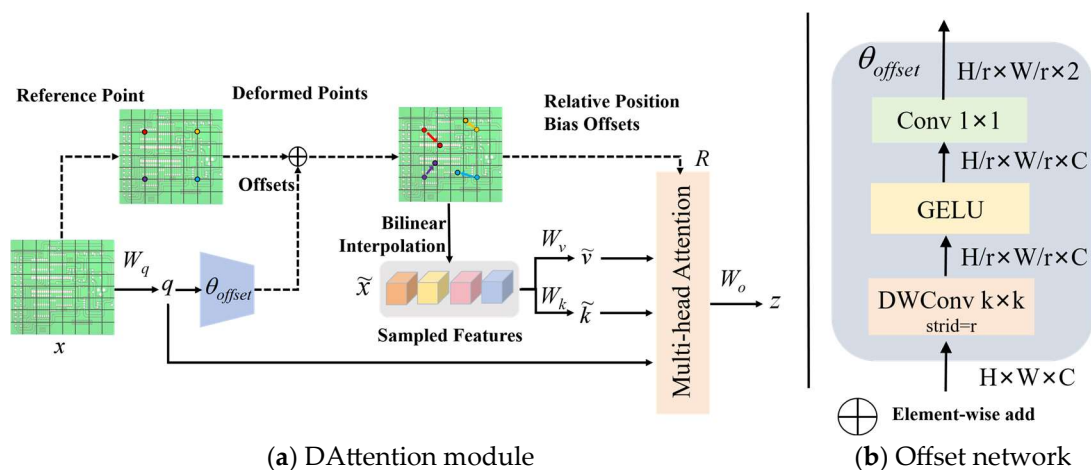
(**a**) DAttention module   (**b**) Offset network

Figure 8. The structure diagrams of DAttention module and Offset network.

2.2.5. DRBC3 Module

Traditional convolutional layers suffer from excessive parameter quantities, low computational efficiency, and fixed receptive fields.

As shown in Figure 9, the Dilated Reparam Block (DRB) is introduced to address these issues, which updates traditional convolutional layers with dilated convolutions. The size of the original convolution kernel and the size of the dilated convolution kernel are utilized within these dilated layers, effectively acting as large kernel convolutions with fewer parameters. Through reparameterization, multiple small kernel convolutional layers with different dilation rates and batch normalization layers are consolidated into an equivalent large-kernel convolutional layer. This operator not only optimizes the learnable parameters and computational efficiency, but also significantly boosts the network's ability to capture spatial information, providing a broader receptive field without increasing the network depth.
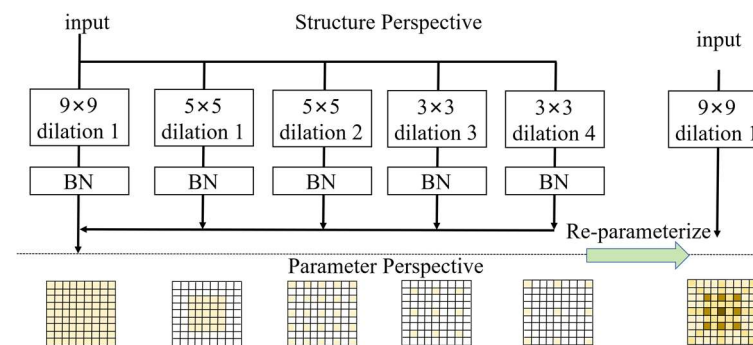

Figure 9. The DRBC3 module structure diagram.

2.2.6. MFAD-Feature Composite Paradigm

Traditional FPN methods in feature extraction typically process features of different scales serially or hierarchically, which can lead to information loss or redundancy. Particularly when dealing with multi-layer scales, this stepwise processing can not only cause these issues but also increase computational complexity. To effectively address this problem, we propose a multi-frequency aggregation diffusion feature composite paradigm based on PCB defect detection from the perspective of parallel frequency processing. As shown in Figure 1, this paradigm consists of two parts. The first part is the ADF, which targets feature flows of different frequencies from layers 4, 5, and 10. Along the blue solid lines, different frequency features generated by the backbone network are aggregated and injected into FDASI, selectively highlighting defect features in critical areas of the image. Then, along

the blue dashed lines, the diffusion mechanism effectively propagates these high-frequency details throughout the network. After diffusion, information of different frequencies within a certain range can complement each other, enhancing the capability of spatial relationship modeling. The second FDASI further aggregates the initially aggregated and diffused features with the features from the backbone network, allowing the model to comprehensively capture crucial high-frequency information in the data. After the second diffusion, these features are fused with the first diffusion features to consolidate and enhance fine-grained features. The second part is the RFC. It inputs initially extracted high-frequency information from layers 4, 5, and 6 of the backbone networks and uses Scale Sequence Feature Fusion (SSFF) for the initial composite of deep and shallow layers, preliminarily locating local details. This fusion, combined with the enhanced high-frequency features from ADF, helps retain and strengthen detail features in the early stages, preventing the loss of important details during subsequent processing. Next, it introduces features from the detail preserving layer 4 of the backbone network. To balance the focus on local details and global structure, mid-frequency semantic information from ADF is also extracted. These three feature streams are then sent into SSFF for deep compositing of deep and shallow layers, enhancing the capability of multi-frequency information compositing while focusing on high-frequency features. Additionally, the model introduces a multi-frequency information detail detection head P1. Designed for fine-grained PCB defect detection, this head performs multi-frequency domain analysis by integrating high-frequency and mid-frequency features, thereby scaling up the precision and reliability of detail detection. Experimental results confirm that this approach effectively detects minute detail defects.

2.2.7. The Loss Function WIoU

The traditional Generalized Intersection over Union (GIoU) has limitations. When two bounding boxes are in an inclusion relationship, GIoU degrades to IoU, failing to distinguish their relative positions. Moreover, when the two boxes intersect, convergence is slow in the horizontal and vertical directions. In addition, GIoU loss may overly focus on the smallest enclosing rectangle, resulting in a small overlap area between the predicted and ground truth bounding boxes. To address these issues, we adopt a dynamic nonmonotonic focusing mechanism. This mechanism evaluates the quality of anchor boxes using an "outlier" and introduces a gradient gain allocation strategy. This strategy reduces the competition among high-quality anchor boxes while effectively mitigating the negative gradient impact from low-quality examples. Furthermore, it can dynamically optimize the loss in order to weight small objects, thereby enhancing the model's detection performance. Particularly when handling complex visual tasks, it allows for greater focus on anchor boxes of ordinary quality, thereby significantly improving the overall performance of the detector.

3. Experiment and Result Analysis

3.1. Experimental Environment and Data Preprocessing

The configuration of our experiments includes an NVIDIA GeForce RTX 2060 SUPER GPU, and an Intel (R) Core (TM) i5-12400F CPU. The development language is Python 3.10.9, and the deep learning platform is the torch-2.0.1+cu117 based on the ubuntu22.04 system. The dataset used in the experiments was released by the Intelligent Robot Open Laboratory of Peking University. There are 693 pieces of original data in the dataset, which were divided into six categories: rat bite, leaky hole, short circuit, open circuit, burr, and pseudo-copper. Some classic defects are shown in Figure 10. The quantity of images in the PCB dataset and the number of each defect type are listed in Table 1. We augment the dataset to 1386 images using the rotation, scale, and translation operators. The dataset was then randomly divided into 970 samples for the training set, 138 samples for the validation set, and 278 samples for the test set, at a ratio of 7:1:2. The batch size was set to two, with a training duration of 300 epochs. The initial learning rate was set at 1×10^{-4}, and the final learning rate was set at 1.
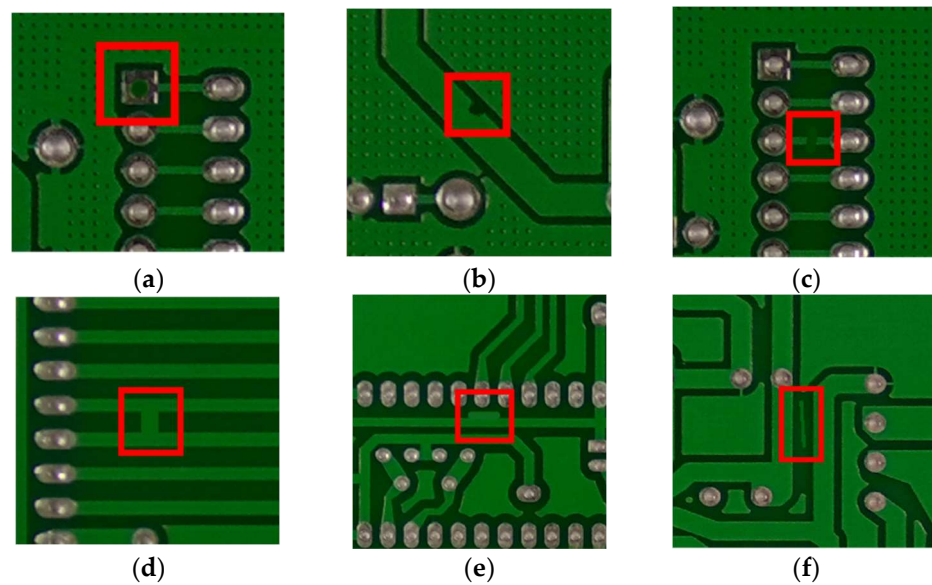
Figure 10. The classic defect images (**a**) missing hole; (**b**) mouse bite; (**c**) open circuit; (**d**) short circuit; (**e**) spur; and (**f**) spurious copper. The red boxes mark the corresponding defects.

Table 1. The quantity of images in the PCB dataset and the count of each defect type.

Defect Type	Missing Hole	Mouse Bite	Open Circuit	Short Circuit	Spur	Spurious Copper
Number of images	230	230	232	232	230	232
Number of defects	199	215	166	190	204	193

3.2. Evaluation Metrics

The prevalent precision (*P*), recall (*R*), mean Average Precision (*mAP*) [30], *F1* score [31] and GFLOPs [32] were set as the detection performance evaluation criteria. Specifically, the F1 score, which aims to consolidate precision and recall into one indicator, is a harmonic mean of *P* and *R*. *mAP* computes the average AP across all classes, providing a comprehensive evaluation of the model's overall performance in multiclass classification tasks. Moreover, the index of GFLOPs is used to measure the computational complexity.

3.3. Experimental Result Analysis

3.3.1. Comparison Experimental Analysis

Keeping the other structures unchanged, we solely compare the detection performance of different attention modules. The DAttention module is compared with other advanced attention modules, and the comparison results are shown in Table 2. Despite the slightly higher GFLOPs introduced by incorporating DAttention, its detection metrics outperform those of other attention modules. Notably, with an mAP50 of 97.0% and an mAP50-95 of 51.0%, the effectiveness of the attention mechanism in DAttention is clearly demonstrated.

The comparison results of AP values of various defect detection methods are report-ed in Table 3. It can be seen that MFAD-RTDETR achieves better results compared with other algorithms. In particular, the value of AP for detecting missing hole defects in the test set reached 99.5%. Although the performance of YOLOv3, YOLOv5, and YOLOv6 models is comparable to the proposed model in detecting missing hole defects, the proposed model significantly outperforms these advanced detection models in other defect categories. These findings further verify the effectiveness of the design scheme.

Table 2. Performance comparison of different attention modules. Bold fonts indicate the best results.

Method	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)	Parameters	GFLOPs
AIFI	95.1%	92.4%	94.6%	49.1%	19879464	57.0
HiLoAttention [33]	95.6%	92.9%	95.6%	50.5%	17281921	178.3
BiLevelRoutingAttention [34]	93.5%	87.2%	91.6%	46.7%	16437147	161.9
DAttention	**96.5%**	**94.5%**	**97.0%**	**51.0%**	**16268912**	**176.5**

Table 3. Comparison of AP values. Bold fonts indicate the best results.

Model	Spur	Open Circuit	Mouse Bite	Spurious Copper	Missing Hole	Short
RTDETR [20]	89.6%	93.4%	92.7%	97.0%	98.5%	96.3%
Faster R-CNN [14]	82.9%	94.4%	89.2%	88.2%	97.5%	94.8%
YOLOv3 [35]	91.7%	92.0%	92.9%	98.3%	**99.5%**	96.6%
YOLOv5 [30]	89.4%	94.3%	88.2%	97.0%	**99.5%**	97.2%
YOLOv6 [32]	84.2%	90.4%	83.3%	95.8%	**99.5%**	96.2%
YOLOv8 [36]	87.3%	94.7%	83.2%	94.6%	99.2%	95.1%
MFAD-RTDETR	**92.9%**	**94.8%**	**98.4%**	**99.0%**	**99.5%**	**97.4%**

The metrics of the proposed algorithm are compared with other object detection algorithms, and the results are listed in Table 4.

Table 4. Comparison of various advanced object detection algorithms. Bold fonts indicate the best results.

Model	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)	F1-Score
RTDETR	95.1%	92.4%	94.6%	49.1%	0.937
Faster RCNN	-	-	91.2%	43.8%	-
YOLOv3	95.4%	94.3%	95.8%	50.3%	0.948
YOLOv6	92.6%	88.9%	91.6%	46.8%	0.907
YOLOv8	93.9%	89.3%	93.2%	48.1%	0.915
GOLD-YOLO [37]	-	-	93.4%	48.6%	-
MFAD-RTDETR	**96.5%**	**94.5%**	**97.0%**	**51.0%**	**0.955**

As presented in Table 4, the MFAD-RTDETR uses 1386 samples to achieve 97.0% mAP50 values and 51% mAP50-95 values, respectively. In addition, the proposed model surpasses the second metrics in both mAP50 and mAP50-95 by nearly 1.2% and 0.7%, respectively. It also outperforms the recent GOLD-YOLO model. Furthermore, the results show that the F1 score achieved 0.955, indicating an ideal tradeoff between precision and recall.

3.3.2. Ablation Experiments

To verify the contributions of different modules, this work conducts extensive ablation experiments. The ablation results are listed in Table 5. Compared with the original RTDETR, integration of the DFR module improves the mAP50, mAP50-95, and F1-score metrics by 0.3%, 0.8%, and 1.0%, respectively. The intuitive explanation is that the proposed detail preserver can effectively perform the preliminary extraction of details. After adopting DAttention and the VSS Block, the model also shows improvements in mAP50, mAP95, and F1 score metrics. This indicates that this attention mechanism and the Mamba-based approach can further promote experimental results. Furthermore, after integrating the ADF and RFC modules, the model yields improvements of 1.1%, 1.6%, and 0.4% in mAP50, mAP50-95, and F1-score, respectively. It can be inferred that the proposed MFAD-feature composite paradigm significantly enhances model performance. And the introduced WIoU further increases the mAP50 and mAP50-95 values, indicating that it can better focus on anchor boxes of ordinary quality, thereby facilitating more focused learning of

important features. At the same time, the model has approximately 18.2% fewer parameters compared to the original RTDETR. In other words, these modules proposed in this study have significant contributions for the detection of tiny industrial defects on PCB.

Table 5. Results on Ablation experiments. $\sqrt{}$ means the corresponding module is embedded in the model. Bold fonts indicate the best results.

RT-DETR	DFR	DAttention	VSS	ADF	RFC	WIoU	mAP50	mAP50-95	Params (M)	F1-Score
√							94.6%	49.1%	19879464	0.937
√	√						94.9%	49.5%	19740400	0.947
√		√					95.3%	49.3%	19882920	0.951
√			√				95.8%	49.9%	19266088	0.944
√				√			96.1%	50.1%	15506748	0.955
√					√		96.2%	49.6%	15433251	0.951
√						√	95.6%	49.4%	19879464	0.942
√	√	√					95.5%	49.8%	19854892	0.944
√	√	√	√				95.7%	49.9%	19129200	0.952
√	√	√	√	√			96.6%	50.3%	18672780	0.954
√	√	√	√	√	√		96.8%	50.7%	15747420	0.956
√	√	√	√	√	√	√	**97.0%**	**51.0%**	**16268912**	**0.955**

3.3.3. Visual Analysis of Experimental Results

To better prove the robust performance of the MFAD-RTDETR model, we visualized the changes in the loss function during the whole training and validation process of MFAD-RTDETR in Figure 11. From Figure 11, it can be seen the giou_loss curves of training and validation both ultimately converge smoothly to less than 0.5, while the cls_loss curves converge between 0.4 and 0.5. Additionally, the l1_loss curves converge to within 0.1. Furthermore, mAP50 and mAP50-95 eventually stabilize between 0.8 and 1.0, and above 0.5, respectively. In the end, after 300 epochs in the training process, all curves converge smoothly and flat, indicating that the model training results are relatively ideal. To further verify the effectiveness of the proposed model on each defect category, the normalized confusion matrixes of the RT-DETR and MFAD-RTDETR models at their respective optimal performances are shown in Figure 12. The comparative visualization demonstrates that the MFAD-RTDETR model achieves ideal performance, which surpasses the original RT-DETR model across defect categories except for the missing_hole defect. As the detection accuracy for the missing_hole defect reaches 100%, the MFAD-RTDETR model proposed in this paper exhibits excellent performance in detecting small defect targets, thereby confirming its effectiveness and reliability in real-world industry scenarios. Therefore, this advancement indicates significant potential for the application of the MFAD-RTDETR model in detecting small object defects in the industrial field.

Moreover, we adopt the PR curve and F1-condifence curve to prove the robustness of our MFAD-RTDETR model for PCB defect detection. The PR curve and F1-condifence curve of the MFAD-RTDETR model are shown in Figure 13. Based on Figure 13, the PR curve for the test set is very close to the upper right corner, which indicates that the classifier's performance is exceptional. Additionally, the F1-confidence curve reveals that the proposed model achieves a good balance between precision and recall, demonstrating strong stability and reliability.
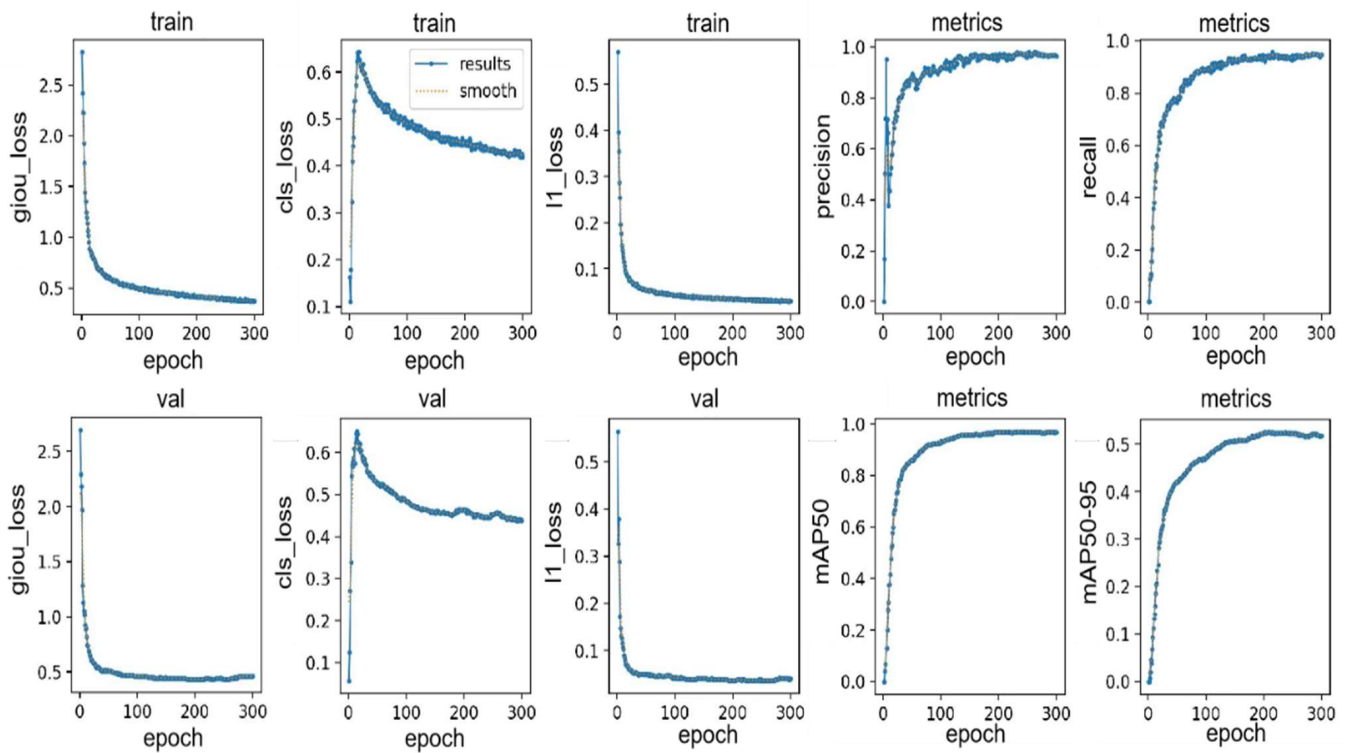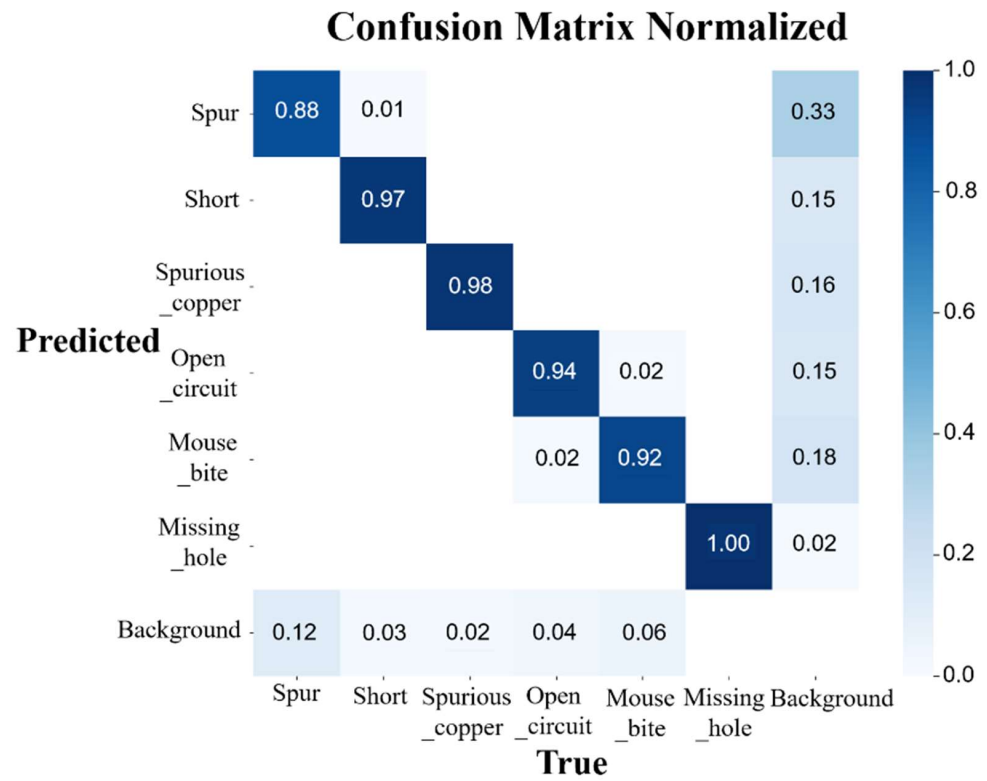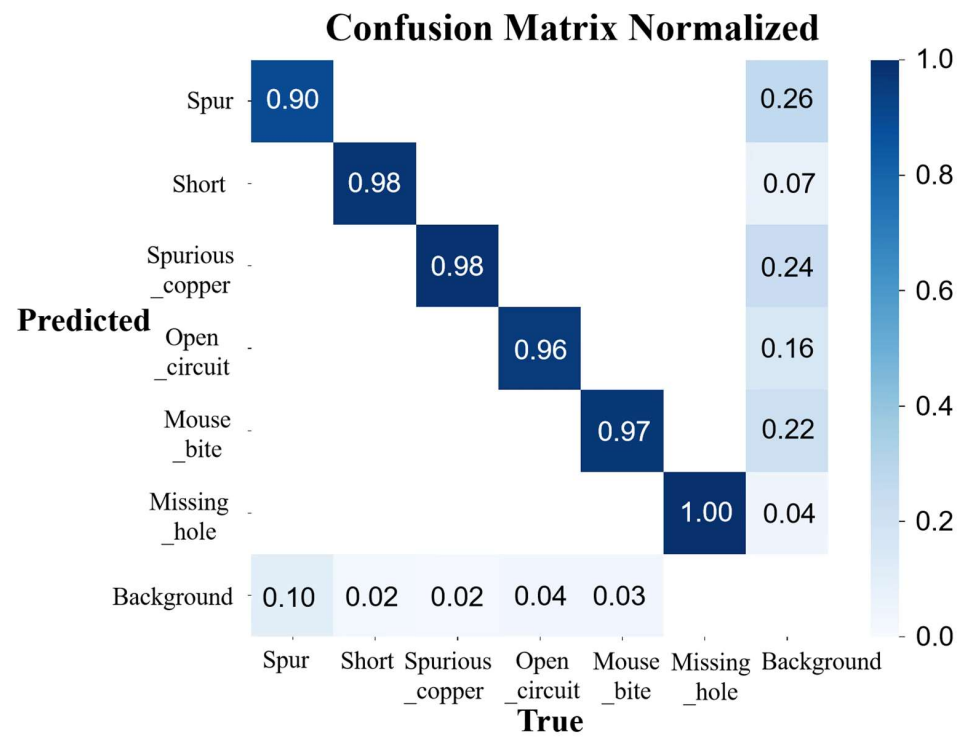
Figure 11. The loss, precision, recall, mAP50, and mAP50-95 curves of the MFAD-RTDETR model.

In addition, the visualization of heat maps corresponding to before and after the addition of the P1 detection head is shown in Figure 14. Obviously, the addition of the P1 detection head significantly enhances the detection performance and the detection effect on the test set is ideal. As shown in Figure 14a, the network without the P1 detection head performed poorly in detecting defects, failing to identify all defect points. Even when defects were detected, the heat map showed noticeable inaccuracies in localization, unable to precisely mark the exact locations of the defects. In contrast, after introducing the P1 detection head, a marked improvement was observed. In Figure 14b, the network with the P1 detection head successfully detected all missing holes on the PCB, with the heat map accurately pinpointing the location of each defect, demonstrating high localization precision. This improvement indicates that the P1 detection head plays a crucial role in enhancing the network's overall detection performance and further confirms its effectiveness and reliability in practical applications. In summary, these results clearly demonstrate the positive impact of adding the P1 detection head, particularly in handling complex and fine-grained detection tasks, where its advantages become even more pronounced.

To better validate the effectiveness of detection on six types of defects, the partial detection results of MFAD-RTDETR are shown in Figure 15. The detection confidence for all defects, except for mouse bite and open circuit, is above 80%, while the confidence for mouse bite and open circuit ranges between 70% and 80%.
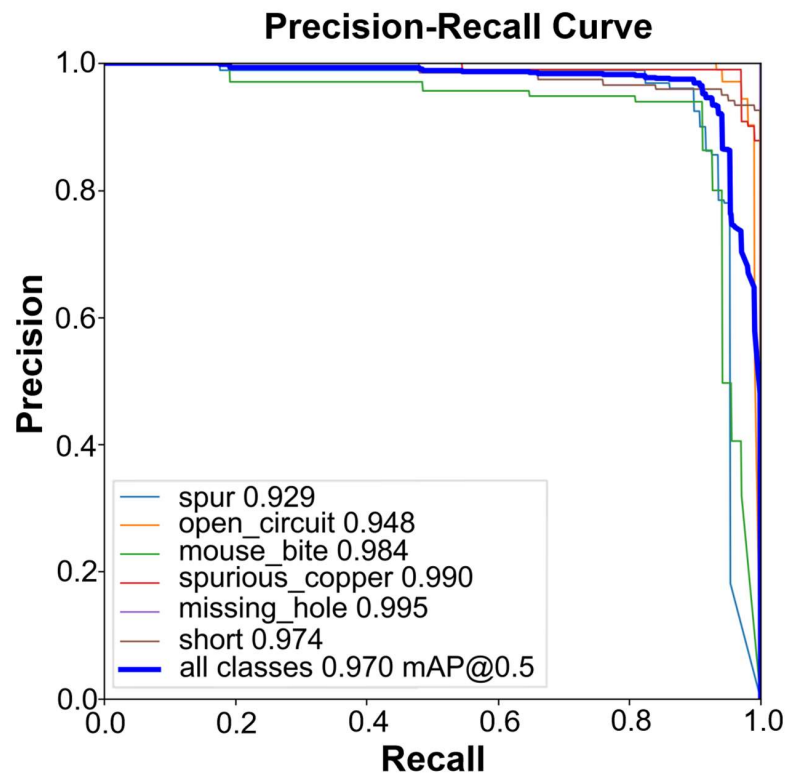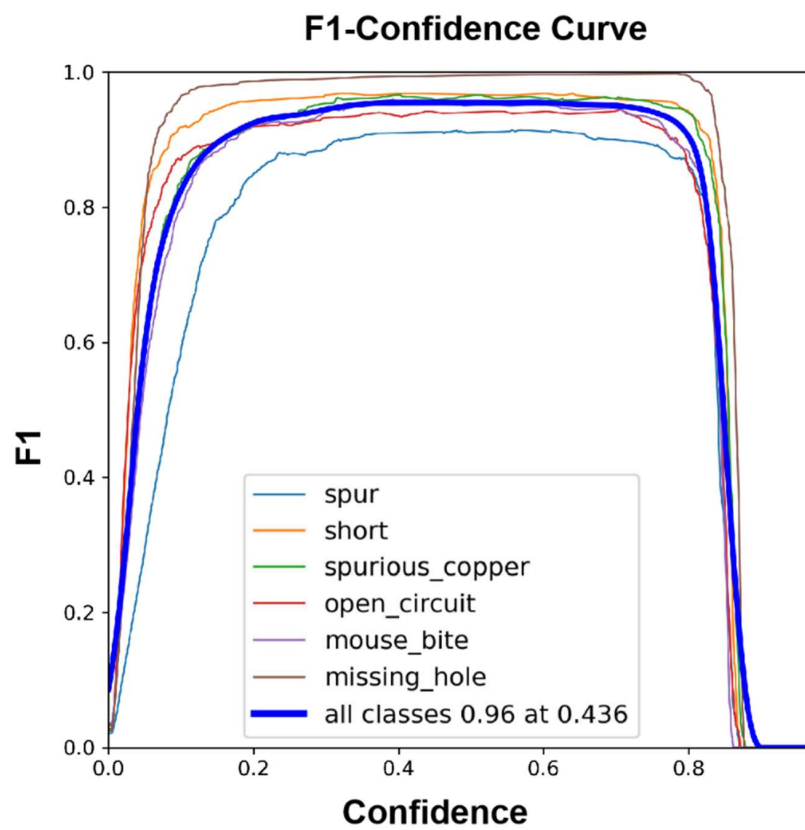
(a)

(b)

Figure 12. Confusion matrix comparison plot of RTDETR and MFAD-RTDETR. (**a**) represents the confusion matrix for RTDETR, (**b**) represents the confusion matrix for MFAD-RTDETR.

Precision-Recall Curve

spur 0.929
open_circuit 0.948
mouse_bite 0.984
spurious_copper 0.990
missing_hole 0.995
short 0.974
all classes 0.970 mAP@0.5

(**a**) PR curves

F1-Confidence Curve

spur
short
spurious_copper
open_circuit
mouse_bite
missing_hole
all classes 0.96 at 0.436

(**b**) F1-codfidence curves

Figure 13. The PR curves and F1-confidence curves of the MFAD-RTDETR model.

(**a**) Before adding the P1 detection head

(**b**) After adding the P1 detection head

Figure 14. Heat Maps Corresponding to Before and After Adding the P1 Detection Head.

Figure 15. Samples of test results for (**a**) missing hole; (**b**) mouse bite; (**c**) open circuit; (**d**) short circuit; (**e**) spur; and (**f**) spurious copper.

4. Conclusions

In this paper, we propose an improved PCB defect detection scheme based on the RT-DETR model, which intends to address the issues of low accuracy in PCB defect detection with current general object detection algorithms. The method introduces a DFR in the original RTDETR backbone to capture and preserve local feature details. Following processing through the VSS based on the Mamba architecture, it enhances global attention and reduces complexity. The deformable attention mechanism dynamically adjusts reference points to achieve the precise localization of target defects. Additionally, a receptive field synthesis mechanism enriches multi-scale semantic information while reducing model complexity. Furthermore, the model proposes the MFAD-feature composite paradigm, comprising the ADF and RFC modules, aimed at facilitating fine-grained feature perception while maintaining global attention. Finally, utilizing the WIoU dynamic nonmonotonic focusing mechanism, the model focuses on average-quality anchor boxes to improve detection performance. Experimental results demonstrate that the proposed algorithm achieves the 97.0% and 51.0% detection performance in mAP@0.5 and mAP@0.5:0.95, which exhibits significantly improved detection accuracy for small PCB defects compared to other defect detection networks; while there is still room for robust real-time PCB defect detection, we will incorporate more effective attention mechanisms and lightweight strategies to explore efficient yet simple PCB defect detection solutions in the future.

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

Data Availability Statement: The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request. All other public datasets used are available and cited in the references.

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

References

1. Dharampal, V.M. Methods of image edge detection: A review. *J. Electr. Electron. Syst.* **2015**, *4*, 150. [CrossRef]
2. Bracewell, R.N. The fourier transform. *Sci. Am.* **1989**, *260*, 86–95. [CrossRef] [PubMed]
3. Yao, J.; Krolak, P.; Steele, C. The generalized Gabor transform. *IEEE Trans. Image Process.* **1995**, *4*, 978–988. [PubMed]
4. Zhang, D.; Zhang, D. Wavelet transform. In *Fundamentals of Image Data Mining: Analysis, Features, Classification and Retrieval*; Springer: Cham, Switzerland, 2019; pp. 35–44.
5. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28. [CrossRef]
6. Brophy, J.; Lowd, D. Machine unlearning for random forests. In Proceedings of the International Conference on Machine Learning, Bellevue, WA, USA, 28 June–2 July 2011; pp. 1092–1104.
7. Li, J.; Cheng, J.; Shi, J.; Huang, F. Brief introduction of back propagation (BP) neural network algorithm and its improvement. In *Advances in Computer Science and Information Engineering*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 2, pp. 553–558.
8. Reeves, C.R. Genetic algorithms. In *Handbook of Metaheuristics*; Springer: Cham, Switzerland, 2010; pp. 109–139.
9. Park, J.Y.; Hwang, Y.; Lee, D.; Kim, J.-H. MarsNet: Multi-label classification network for images of various sizes. *IEEE Access* **2020**, *8*, 21832–21846. [CrossRef]
10. Yu, F.; Koltun, V.; Funkhouser, T. Dilated residual networks. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 472–480.
11. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, New Orleans, LA, USA, 5–9 March 2017; pp. 2961–2969.
12. Girshick, R. Fast R-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.

13. Ding, R.; Dai, L.; Li, G.; Liu, H. TDD-net: A tiny defect detection network for printed circuit boards. *CAAI Trans. Intell. Technol.* **2019**, *4*, 110–116. [CrossRef]
14. Liu, B.; Zhao, W.; Sun, Q. Study of object detection based on Faster R-CNN. In Proceedings of the 2017 Chinese Automation Congress (CAC), Jinnan, China, 20–22 October 2017; pp. 6233–6236.
15. Hu, B.; Wang, J. Detection of PCB surface defects with improved faster-RCNN and feature pyramid network. *IEEE Access* **2020**, *8*, 108335–108345. [CrossRef]
16. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. Shufflenet v2: Practical guidelines for efficient CNN architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.
17. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21002–21012.
18. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 87–110. [CrossRef]
19. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo algorithm developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [CrossRef]
20. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. Detrs beat yolos on real-time object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–21 June 2024; pp. 16965–16974.
21. Kim, S.; Park, E. Smpconv: Self-moving point representations for continuous convolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 10289–10299.
22. Shi, D. TransNeXt: Robust Foveal Visual Perception for Vision Transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–21 June 2024; pp. 17773–17783.
23. Ruan, J.; Xiang, S. Vm-unet: Vision mamba unet for medical image segmentation. *arXiv* **2024**, arXiv:2402.02491.
24. Xia, Z.; Pan, X.; Song, S.; Li, L.E.; Huang, G. Vision transformer with deformable attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 4794–4803.
25. Ding, X.; Zhang, Y.; Ge, Y.; Zhao, S.; Song, L.; Yue, X.; Shan, Y. UniRepLKNet: A Universal Perception Large-Kernel ConvNet for Audio Video Point Cloud Time-Series and Image Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 5513–5524.
26. Kang, M.; Ting, C.M.; Ting, F.F.; Phan, R.C.-W. ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation. *Image Vis. Comput.* **2024**, *147*, 105057. [CrossRef]
27. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism. *arXiv* **2023**, arXiv:2301.10051.
28. Gong, L.H.; Pei, J.J.; Zhang, T.F.; Zhou, N.-R. Quantum convolutional neural network based on variational quantum circuits. *Opt. Commun.* **2024**, *550*, 129993. [CrossRef]
29. Xu, S.; Zheng, S.C.; Xu, W.; Xu, R.; Wang, C.; Zhang, J.; Teng, X.; Li, A.; Guo, L. HCF-Net: Hierarchical Context Fusion Network for Infrared Small Object Detection. *arXiv* **2024**, arXiv:2403.10778.
30. Tang, J.; Liu, S.; Zhao, D.; Tang, L.; Zou, W.; Zheng, B. PCB-YOLO: An improved detection algorithm of PCB surface defects based on YOLOv5. *Sustainability* **2023**, *15*, 5963. [CrossRef]
31. Cyril, G.; Gaussier, E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In Proceedings of the European Conference on Information Retrieval, Santiago de Compostela, Spain, 21–23 March 2005; pp. 345–359.
32. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A single-stage object detection framework for industrial application. *arXiv* **2022**, arXiv:2209.02976.
33. Pan, Z.; Cai, J.; Zhuang, B. Fast vision transformers with hilo attention. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 14541–14554.
34. Zhu, L.; Wang, X.; Ke, Z.; Zhang, W.; Lau, R. Biformer: Vision transformer with bi-level routing attention. In Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 10323–10333.
35. Redmon, J. Yolov3: An incremental improvemen. *arXiv* **2018**, arXiv:1804.02767.
36. Ling, Q.; Isa, N.A.M.; Asaari, M.S.M. Precise detection for dense PCB components based on modified YOLOv8. *IEEE Access* **2023**, *11*, 116545–116560. [CrossRef]
37. Wang, C.; He, W.; Nie, Y.; Guo, J.; Liu, C.; Han, K.; Wang, Y. Gold-YOLO: Efficient object detector via gather-and-distribute mechanism. *Adv. Neural Inf. Process. Syst.* **2024**, *36*, 51094–51112.