forests

MDPI

Article

FDD-YOLO: A Novel Detection Model for Detecting Surface Defects in Wood

Bo Wang [1,2], Rijun Wang [3,4,*], Yesheng Chen [3], Chunhui Yang [3], Xianglong Teng [3] and Peng Sun [5]

[1] School of Artificial Intelligence and Smart Manufacturing, Hechi University, Yizhou 546300, China; 05041@hcnu.edu.cn
[2] Key Laboratory of AI and Information Processing, Hechi University, Yizhou 546300, China
[3] School of Teachers, College for Vocational and Technical Education, Guangxi Normal University, Guilin 541004, China; 17707734324@163.com (Y.C.); 19830029462@163.com (C.Y.)
[4] Guangxi University Engineering Research Center of Agricultural and Forestry Intelligent Equipment Technology, Guilin 546300, China
[5] School of Information and Communication Engineering, North University of China, Taiyuan 030051, China; sunpeng@nuc.edu.cn
* Correspondence: wangrijun1982@126.com

Abstract: Wood surface defect detection is a critical step in wood processing and manufacturing. To address the performance degradation caused by small targets and multi-scale features in wood surface defect detection, a novel deep learning model is proposed in this study, FDD-YOLO, specifically designed for this task. In the feature extraction stage, the C2f module and the funnel attention (FA) mechanisms are integrated into the design of the C2f-FA module to enhance the model's ability to extract features of wood surface defects of various sizes. Additionally, the Dual Spatial Pyramid Pooling-Fast (DSPPF) module is developed, and the Context Self-attention Module (CSAM) is introduced to address the limitations of traditional max pooling methods, which often overlook global contextual information when extracting local features, thereby improving the detection of small-scale wood defects. In the feature fusion stage, a Dual Cross-scale Weighted Feature-fusion (DCWF) module is proposed to fuse shallow, deep, and cross-scale features through a weighted summation approach, effectively addressing the challenge of scale variation in wood surface defects. Experimental results demonstrate that the proposed FDD-YOLO model significantly improves detection performance, increasing the mAP of the baseline model YOLOv8 from 78% to 82.3%, a substantial enhancement of 4.3 percentage points. Furthermore, FDD-YOLO outperforms other mainstream defect detection models in terms of detection accuracy. The proposed model demonstrates significant potential for industrial applications by improving detection accuracy, enhancing production efficiency, and reducing material waste, thereby advancing quality control in wood processing and manufacturing.

Keywords: detection model; funnel attention mechanism; dual spatial pyramid pooling-fast; dual cross-scale weighted feature fusion; wood surface defect

check for updates

Received: 9 January 2025
Revised: 4 February 2025
Accepted: 8 February 2025
Published: 10 February 2025

Citation: Wang, B.; Wang, R.; Chen, Y.; Yang, C.; Teng, X.; Sun, P. FDD-YOLO: A Novel Detection Model for Detecting Surface Defects in Wood. *Forests* **2025**, *16*, 308. https://doi.org/10.3390/f16020308

Copyright: © 2025 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/).

1. Introduction

The detection of surface defects in wood plays a critical role in the automation of wood processing [1]. This technology enables manufacturers to effectively identify various defects on the wood surface, such as cracks, wormholes, and knots. By promptly repairing or removing defective wood, the overall quality and reliability of wood-based products can be significantly improved, thereby preventing potential safety hazards caused by surface

defects [2]. Consequently, an in-depth investigation into wood surface defect detection technologies is not only necessary but also of paramount importance. The development of wood surface defect detection technologies has made significant progress. Traditional manual inspection methods, although still in use, are time-consuming, prone to human error, and lack the consistency required for large-scale industrial applications [3]. To overcome these limitations, automated systems leveraging advanced technologies such as sensor-based detection, machine vision, and deep learning have been increasingly adopted [4–6].

Sensor-based defect detection methods utilize various physical properties such as sound, vibration, light, and heat to identify irregularities in wood structure, making them invaluable for applications where visual inspection may fall short. Ultrasonic testing is a nondestructive testing (NDT) technique that utilizes the propagation characteristics of high-frequency sound waves to detect surface and internal defects in materials. This method is widely applied in wood inspection, particularly for identifying internal, invisible defects such as cracks, knots, and decay [7–9]. Laser-based wood defect detection technology [10–12] enables the generation of precise geometric and texture information of wood surfaces, making it particularly effective for detecting subtle cracks, depressions, knots, and other surface defects. By analyzing the interaction between laser beams and the wood surface, detailed surface shapes and features can be extracted for defect identification. Infrared and thermal imaging-based wood defect detection methods are non-contact approaches that identify defects by detecting anomalies in surface or internal temperature distribution [13–15]. These techniques leverage variations in infrared light or thermal radiation signals to effectively detect surface cracks, internal voids, decay, moisture changes, and other defects associated with thermal conductivity properties. Hybrid sensor systems integrate multiple sensing technologies to enhance the accuracy and reliability of wood defect detection [16]. By fusing data from different types of sensors, this approach aims to overcome the limitations of individual sensing techniques, combining their strengths to enable comprehensive defect detection in more complex and diverse application scenarios. The multimodal data acquired from various sensors provide multidimensional information about the wood, including surface textures, internal structures, and thermal properties. This enables more accurate defect identification [17].

Machine vision-based defect detection techniques primarily rely on manually designed feature extraction algorithms and classifiers [18]. Common algorithms include edge detection, region segmentation (e.g., Otsu thresholding, region growing), and texture analysis (e.g., Gray-Level Co-occurrence Matrix, GLCM) [19]. These methods are straightforward to implement and cost-effective. However, they are highly sensitive to complex textures and variations in lighting, which limits their detection accuracy. In recent years, the integration of machine vision with deep learning has led to significant advancements in wood surface defect detection. Convolutional Neural Networks (CNNs), such as Residual Network (ResNet) [20], Very Deep Convolutional Networks (VGG) [21], and Inception [22], have demonstrated strong capabilities in automatically extracting hierarchical features. This makes them highly effective in addressing challenges posed by complex textures and irregular defects in wood surface detection. By incorporating object detection algorithms such as You Only Look Once (YOLO) [23–26], Faster Region-based Convolutional Neural Network (Faster R-CNN) [27], Mask R-CNN [28], Single-Shot multibox Detector (SSD) [29], and RetinaNet [30], these methods not only classify wood surface defects but also localize them by generating bounding boxes. This enables the simultaneous identification and precise localization of multiple types of defects on wood surfaces, significantly enhancing detection performance.

Wood surface defect detection plays a critical role in the wood processing industry, particularly in sectors such as furniture manufacturing and construction materials, where undetected defects can lead to significant product quality issues and material waste. For instance, in the production of high-end wooden flooring, even minor surface defects can result in substantial financial losses due to rejected batches [31]. Despite the importance of this task, it faces significant challenges, including defect diversity, multi-scale characteristics, and the difficulty of detecting small defects, which are particularly problematic in complex industrial environments. Although deep learning techniques [32–34] have made remarkable progress in wood surface defect detection in recent years, the accuracy and generalization capability for detecting multiple defect types still need improvement. Existing research has shown limited success in addressing the challenges of small-object detection and reducing false positives and false negatives, especially in complex industrial environments where robustness remains suboptimal.

To address these issues, this paper proposes a novel wood surface defect detection algorithm based on YOLOv8 as the baseline model. By leveraging the multi-scale characteristics and efficient detection capabilities of YOLOv8, the proposed method introduces specific optimization strategies to significantly enhance small-object detection performance while effectively reducing false positives and missed detections. The main contributions of this paper are as follows:

1. A Novel Detection Model: This paper presents a novel detection model for wood surface defect detection named FDD-YOLO. The model is designed to address the challenges of small-object detection and multi-scale defect detection, thereby improving detection performance.
2. DSPPF+CSAM Module for Feature Extraction: To enhance the ability to detect small wood surface defects, a dual-branch feature extraction strategy is proposed, incorporating the DSPPF+CSAM module. This addresses the limitations of the SPPF module, which uses max pooling to extract local features of multi-scale wood surface defects but overlooks the extraction of global contextual features.
3. DCWF Module for Feature Fusion: By adopting a weighted sum strategy, the module effectively combines deep and shallow features as well as cross-scale features, addressing the scale variation of wood surface defects and improving the model's multi-scale feature fusion capability.

2. Materials and Methods

2.1. Dataset Description

The dataset utilized in this study was sourced from the literature [35] and was collected from a real-world sawmill environment, using a custom-designed image acquisition device. The dataset includes images representing 10 typical wood surface defects. After filtering, seven prevalent defect types with a sufficient sample size were retained for training and evaluation of the proposed model. A total of 3600 images were selected. The defect types included are: Live Knot, Dead_Knot, Knot_with_Crack, Knot_Missing, Crack, Marrow, and Resin. Representative examples of each defect type are shown in Figure 1. The distribution of these defects in the dataset after filtering is summarized in Table 1.

Wood surface defects are characterized by a wide variety of multi-scale features and small target sizes, which pose significant challenges for detection, leading to potential issues such as false positives and missed detections. To address these challenges, this study uses YOLOv8 as the baseline network model and aims to develop a wood surface defect detection algorithm. The goal is to enhance detection accuracy while minimizing false positives and missed detections.

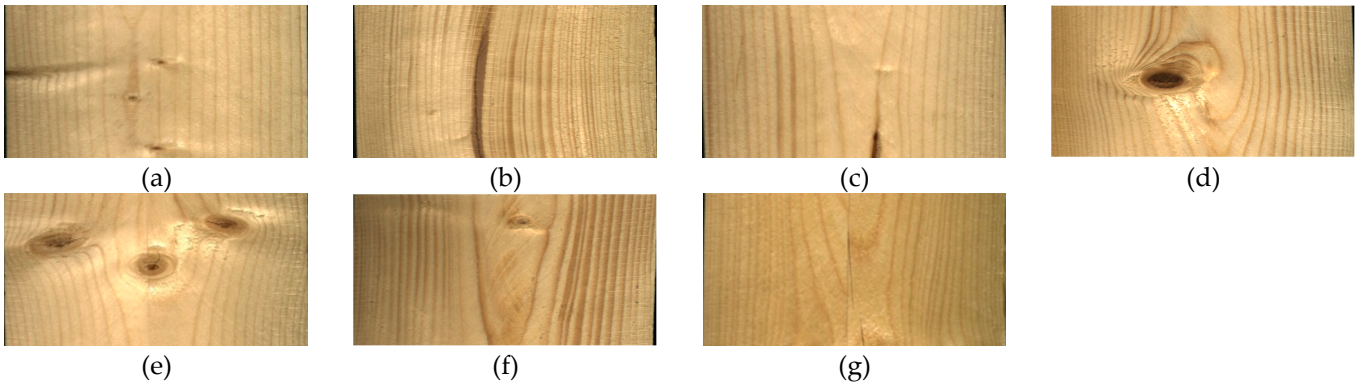


Figure 1. Examples of wood defects [36]: (**a**) Live_Knot; (**b**) Marrow; (**c**) Resin; (**d**) Dead_Knot; (**e**) Knot_with_Crack; (**f**) Knot_Missiong; (**g**) Crack.

Table 1. Distribution of wood surface defects in the dataset.

Defect Type	Number of Occurrences	Number of Images with the Defect	Images in the Dataset (%)
Live_Knot	4070	2256	62.7
Marrow	206	191	5.3
Resin	650	523	14.5
Dead_Knot	2934	1875	52.1
Knot_with_crack	542	398	11.1
Knot_missing	121	110	3.1
Crack	517	371	10.3
Without any defects	7	—	0.2

2.2. FDD-YOLO Model for Wood Surface Defect Detection

The proposed FDD-YOLO model is developed based on YOLOv8 [36] as the baseline network and retains its fundamental architecture, comprising three components: backbone, neck, and prediction. In the backbone, the C2f module is integrated with the funnel attention (FA) mechanism to construct a novel C2f-FA module. This enhancement improves the feature extraction capability of the model for multi-scale wood surface defects. Furthermore, to address the limitations of the Spatial Pyramid Pooling-Fast (SPPF) module, which focuses on extracting local features via max pooling while overlooking the global contextual features of wood defects, a module combining Dual Spatial Pyramid Pooling-Fast (DSPPF) and Context Self-attention Module (CSAM) is designed. This module aggregates both the local salient features and the global contextual features of wood defects, thereby improving the model's ability to detect small-scale surface defects. In the neck, to tackle the challenges posed by varying defect scales, the model incorporates the Dual Cross-scale Weighted Feature-fusion (DCWF) module, which employs a weighted summation approach for feature fusion. By integrating deep and shallow features along with cross-scale features, the DCWF module effectively addresses scale variation issues in defect detection. The architecture of the proposed FDD-YOLO model is illustrated in Figure 2.

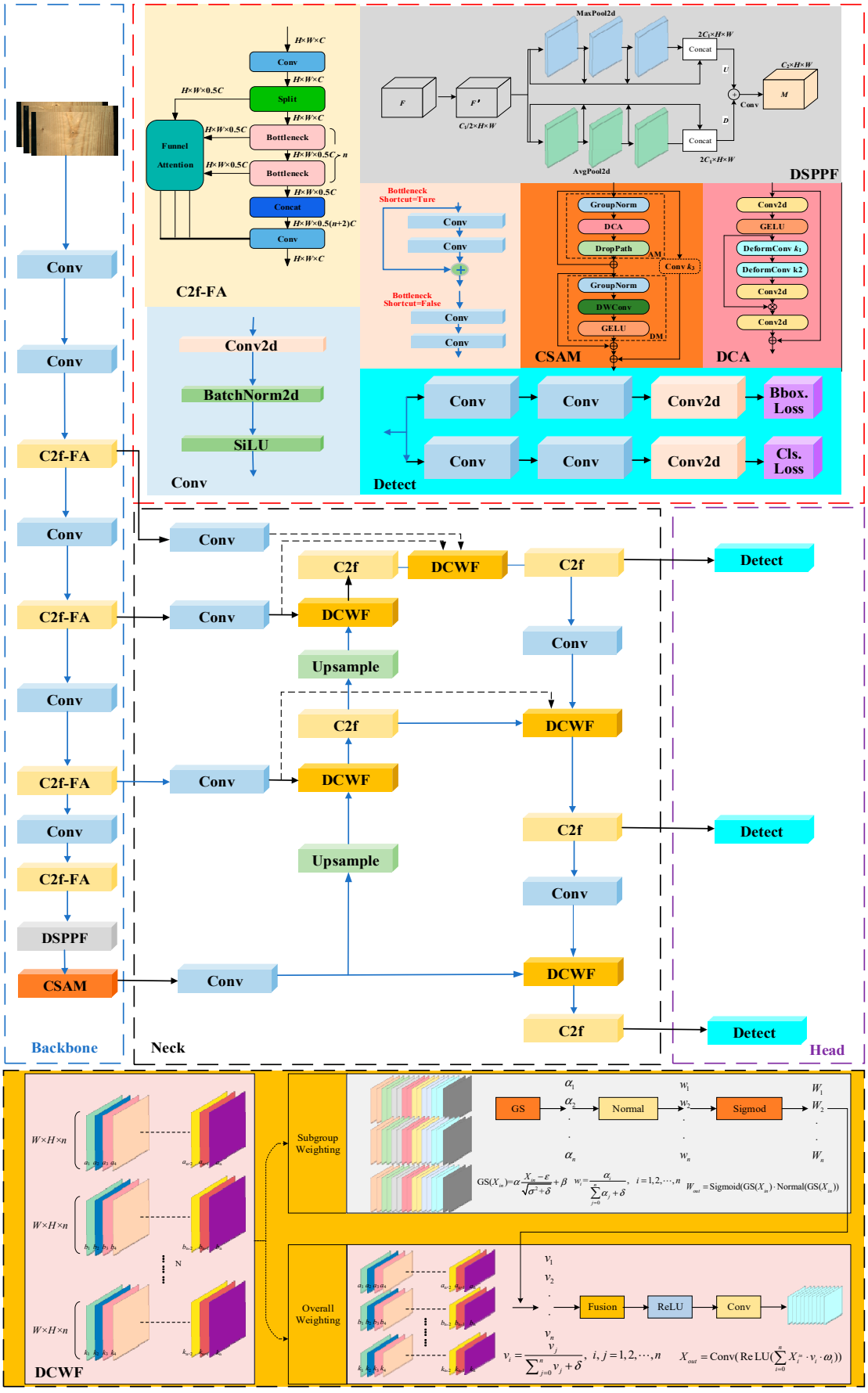


Figure 2. Architecture of the proposed FDD-YOLO model.

2.2.1. C2f-FA Module

The C3 module, which serves as the primary feature extraction component in YOLOv5 [37], was replaced in YOLOv8 by the C2f structure to enhance gradient flow and improve residual feature learning. As a critical module for feature extraction, C2f significantly enhances the model's capability to extract features effectively. To further strengthen the feature extraction ability of the model for multi-scale wood surface defects, this study introduces a lightweight FA mechanism [38]. The FA mechanism is integrated with the C2f module to form a novel C2f-FA module, which enhances the model's capacity to capture features of wood surface defects of varying sizes.

The structure of the designed FA module is illustrated in Figure 3. Within the FA module, the channel mean module is employed to compute the average pixel value across all channels. This module preserves global information while capturing inter-channel correlations, thereby accelerating the convergence speed of the attention mechanism. This ensures that the model converges efficiently. The calculation formula for the Channel mean is as follows:

$$F_{avg}(x_{i,j,n}) = \frac{\sum_{n=1}^{c} x_{i,j,n}}{c} \quad (1)$$

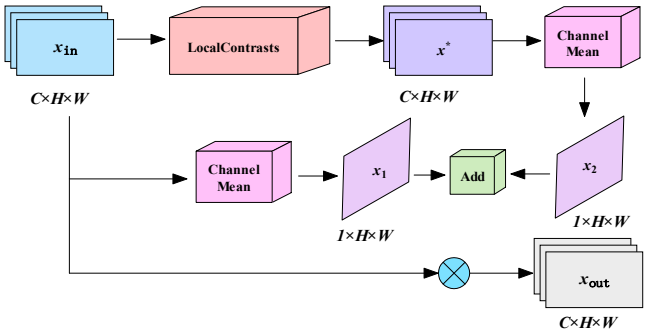


Figure 3. Structure of the FA module.

Here, (x, j) represents the pixel position within the wood image, with a size not exceeding (w, h). c denotes the total number of channels, and n represents the specific channel index.

The Local Contrast module is designed to extract high-frequency information within a pixel and its surrounding regions. Specifically, the operation compares the input pixel with the average subpixel value extracted via convolution within a parameter-defined neighborhood w, using a max operation to evaluate the intensity relationships between the pixel and its surrounding regions. This approach facilitates the capture of local extrema in small defect features. Compared to traditional self-attention mechanisms, this method optimizes the computation scope from a global scale to the vicinity of each pixel, significantly reducing the computational cost of the module. The process is as follows, where the $Avg(x)$ function calculates the average subpixel value within the convolutional kernel defined by the parameter pool:

$$f(x_{c,i,j}) = max(x_{c,i,j}, Avg(x_{c,i,j})) \quad (2)$$

The overall computation process of the FA module can be described as follows: First, the channel-wise average of the input pixel x_{in} is calculated to represent the weight value that retains global information, denoted as x_1. Next, the local extrema x^* of x_{in} are computed through the Local Contrast module, followed by calculating the channel-wise average of the extrema, which represents the weight value considering local extremum information, denoted as x_2. Subsequently, x_1 (global information) and x_2 (local information) are aggregated

using an add operation, enabling the integration of multi-scale feature information. This aggregated weight is then multiplied element-wise with x_{in}. By leveraging the multi-scale weight information generated through the add operation, the model enhances its focus on high-frequency information across defects of varying scales, ultimately producing the output feature map.

The proposed FA-C2f module enhances the original C2f architecture by incorporating a funnel attention (FA) mechanism to compute local weights based on 2D spatial context, thereby improving inter-channel correlations and multi-scale defect perception. As illustrated in Figure 4, the FA-C2f module operates as follows:

1. Feature transformation: A 1 × 1 convolutional layer first adjusts the channel dimensions of the input features.
2. Channel splitting: The transformed features are split into two equal partitions along the channel axis using the Split operation.
   (1) Primary Path: One partition passes through multiple bottleneck blocks to expand the receptive field while preserving hierarchical feature representations.
   (2) Residual Path: The other partition retains the original information via a residual connection, ensuring gradient flow richness with minimal parameter overhead.
3. Feature fusion: The outputs from both paths are concatenated and further refined.
4. Funnel attention integration: To address the weakened inter-channel dependencies caused by splitting, the primary path features are fed into the FA mechanism. This module dynamically recalibrates channel-wise attention weights, prioritizing defect-related regions across varying scales.

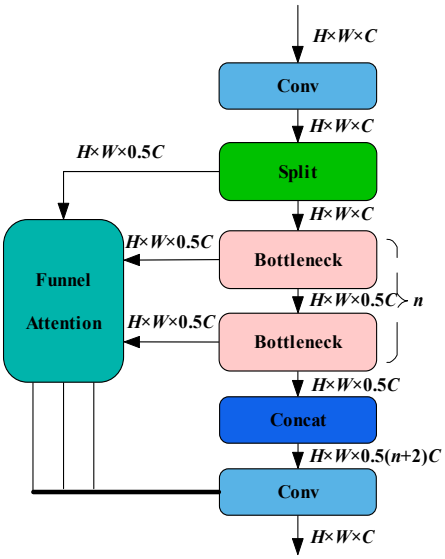


Figure 4. Structure of the C2f-FA module.

This design not only reduces computational complexity but also strengthens the model's ability to distinguish subtle defects in high-texture wood surfaces.

2.2.2. Dual Spatial Pyramid Pooling-Fast (DSPPF) Module

To address the issue of small target features in wood surface defects, it is crucial to not only focus on the local features of the defects but also consider the broader global contextual information. The SPPF module [39] in YOLOv8 uses max-pooling to extract local features at different scales but overlooks the extraction of global context features. To better aggregate local salient features and global contextual features, enhancing the feature generalization ability, the DSPPF module is designed.

The structure of the proposed DSPPF module is illustrated in Figure 5. Unlike the traditional SPPF, DSPPF introduces an additional average pooling branch alongside the max pooling operation. This dual-branch design offers two key advantages:

1. Local feature extraction: The max pooling branch captures local salient features of wood surface defects, which are critical for identifying fine-grained details.
2. Global context aggregation: The average pooling branch extracts global contextual information, enabling the model to better understand the overall defect distribution and surrounding texture patterns.

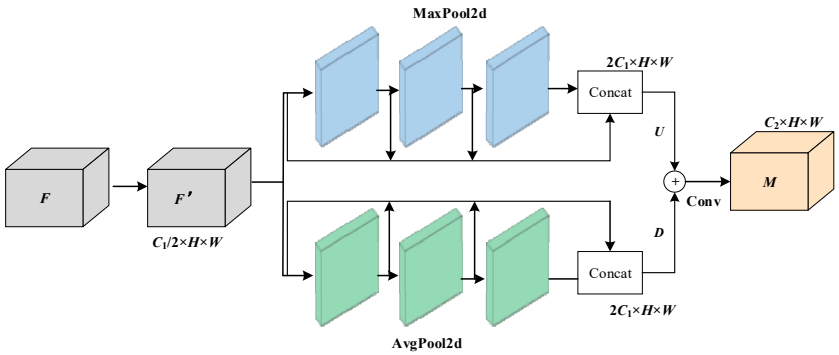


Figure 5. Structure of DSPPF module.

By combining local and global features, DSPPF enhances the backbone network's ability to detect small targets, which often require both detailed local information and broader contextual cues for accurate identification. Furthermore, average pooling retains more background details and contextual information, while max pooling extracts local salient features. The combination of these two pooling methods generates richer features, which is highly beneficial for detecting small target wood surface defects. The working The principle of DSPPF is as follows:

Step 1: A 1×1 standard convolution is applied to the input feature map F for dimensionality reduction, resulting in a new feature map F', which reduces computational complexity.

Step 2: Pyramid max-pooling and average pooling are separately applied to F', yielding feature maps U and D, respectively.

Step 3: Element-wise addition is used to fuse these features, producing the feature map M.

The specific formula is as follows:

$$U = Cat\left(P_m\left(F'\right), P_m\left(P_m\left(F'\right)\right), P_m\left(P_m\left(P_m\left(F'\right)\right)\right), F'\right) \tag{3}$$

$$D = Cat\left(P_a\left(F'\right), P_a\left(P_a\left(F'\right)\right), P_a\left(P_a\left(P_a\left(F'\right)\right)\right), F'\right) \tag{4}$$

$$M = U \oplus D \tag{5}$$

where, $P_m(\cdot)$ and $P_a(\cdot)$ represent max pooling and average pooling, respectively, and $\oplus$ denotes element-wise addition with broadcasting.

2.2.3. Context Self-Attention Module

After the backbone network extracts features; the input feature map provides richer, fine-grained information for the overall model. However, this only results in limited contextual information, where the proportion of relevant detected information is small, and the proportion of background and undetected object information is large. This is particularly challenging for small target detection of wood surface defects, which are easily

confused with the background. To capture global long-range dependencies and achieve comprehensive and distinctive feature representations that differentiate small targets from the background, the Context Self-attention Module (CSAM) is proposed, whose overall structure is shown in Figure 6.

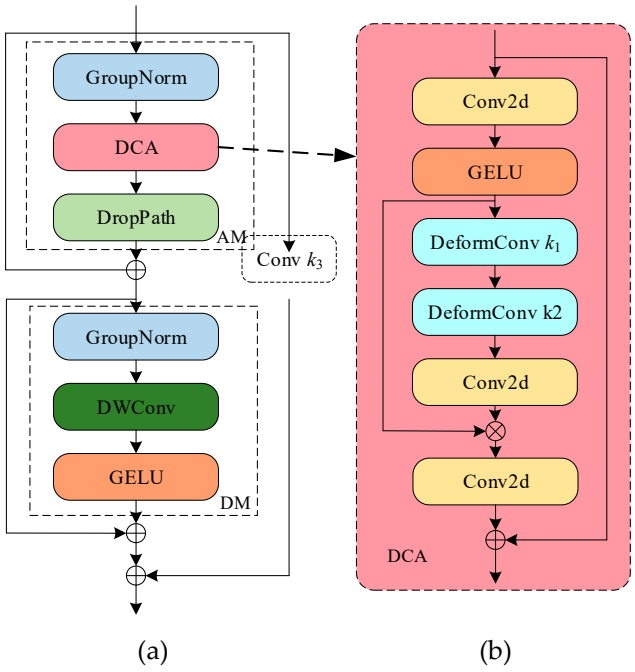


Figure 6. Overall structure of the CSAM module: (**a**) structure of CSAM; (**b**) structure of DCA.

The CSAM is primarily composed of two parts: the Attention Module (AM) based on deformable convolution [40] and the Depthwise Separable Convolution Module (DM) [41]. In AM, the input feature map undergoes group normalization. Next, it enters the Deformable Convolution Attention (DCA) module, where the feature map is initially processed using a 1×1 convolution followed by Gaussian Error Linear Unit (GELU) activation [42]. Two consecutive deformable convolutions of size 3×3 and 5×5 are then applied. The principle behind this is that, based on the target features in the input image, the standard sampling network is warped to make the convolution kernels more closely fit small targets, allowing for more effective feature extraction. Attention is dynamically applied within the receptive field to adapt to varying target sizes. Afterward, residual connections are added, and element-wise multiplication is performed, adjusting the channel number and adding the result to the input feature map. Finally, a DropPath regularization operation is introduced to enhance the generalization ability and robustness of the features.

In DM, the depthwise separable convolution is employed to enhance feature representation. Compared with traditional convolution, the depthwise separable convolution is a lightweight operation that reduces computational cost. Residual connections are also incorporated both outside the AM and DM and within the overall Contextual Attention Module (CAM), enriching the connections between pixel information and improving the model's ability to resist interference and detect defects in complex backgrounds. The entire process can be described as follows:

$$X' = \mathrm{DCA}\left(\mathrm{GN}\left(X_{\mathrm{input}}\right)\right) + X_{\mathrm{input}} \tag{6}$$

$$X_{\mathrm{output}} = \mathrm{GELU}\left(\mathrm{DWConv}\left(\mathrm{GN}\left(X'\right)\right)\right) + X' + X_{\mathrm{input}} \tag{7}$$

where X_{input} is the input to the CAM; X_{output} is the output of the CAM; X' is the output of the AM and also the input to the DM; GN represents group normalization; DWConv denotes the depthwise separable convolution with a 3 × 3 convolution kernel.

2.2.4. Dual Cross-Scale Weighted Feature-Fusion (DCWF) Module

In the feature fusion stage of YOLOv8, the multi-scale feature fusion network based on Feature Pyramid Network (FPN) compensates for the lack of contextual features and addresses the issue of scale variation. However, during the feature fusion process, equal treatment of different inputs may introduce unnecessary noise. Inspired by the approach in [43], we propose replacing the traditional FPN's direct summation with a weighted sum during feature fusion. Thus, we design the DCWF module (as shown in Figure 7) to replace the Concat operation in the original YOLOv8 model and modify the cross-scale fusion method (illustrated by the dashed line in Figure 2). This approach integrates both shallow and deep features as well as cross-scale features to solve this problem.

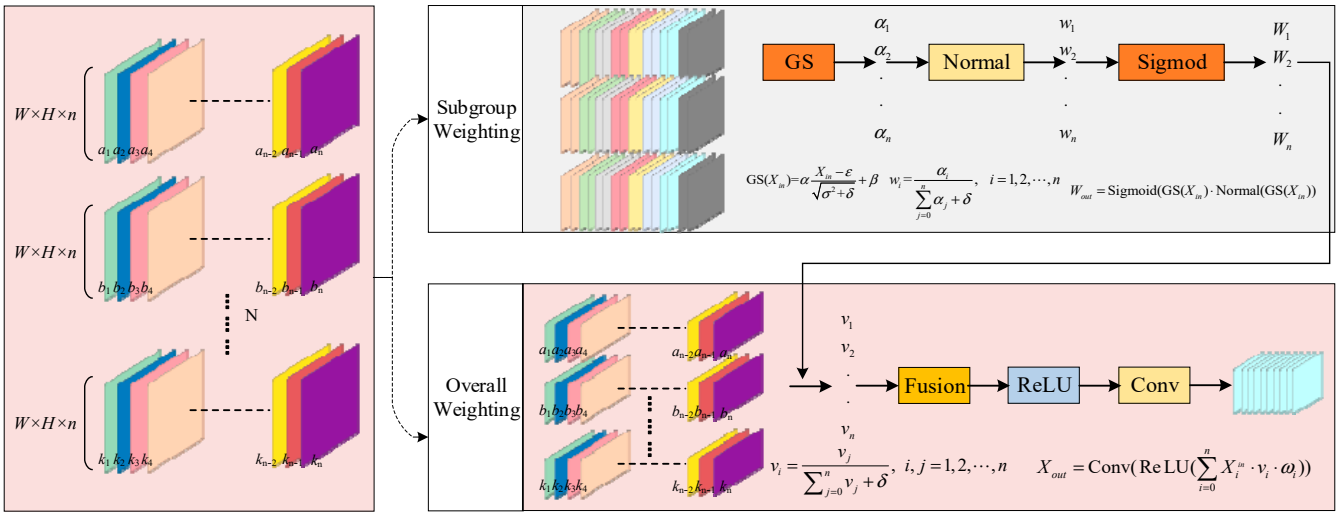


Figure 7. Structure of the DCWF module.

To fully utilize the channel feature information of different input feature maps and minimize the impact of redundant information on feature fusion, the DCWF module first applies a grouping and weighting operation to each input feature before performing an overall weighting operation on all feature inputs to refine the weights. The specific operations are as follows:

(1) Grouped weighting: Initially, the channels of each input feature map are separated into n groups (in this study, n is set to 8, with each group represented by a different color). The weights for each group of feature maps are obtained through GS and Normal operations. Finally, the sigmoid activation function is applied to map the weight values into the range of (0, 1).

$$W_{out} = \text{Sigmoid}(\text{GS}(X_{in}) \cdot \text{Normal}(\text{GS}(X_{in})) \tag{8}$$

$$\text{GS}(X_{in}) = \alpha \frac{X_{in} - \varepsilon}{\sqrt{\sigma^2 + \delta}} + \beta \tag{9}$$

Here, GS refers to Group Standard normalization applied to the input for evaluating the information content of different feature maps. The symbols σ and ϵ represent the standard deviation and mean of X, respectively, while α and β are trainable parameters. Additionally, δ is a small constant added for balancing the division.

The normalized feature maps are multiplied by learnable parameters for normalization, resulting in the weight values for each feature map group. The learnable parameter α in the GS layer is used to measure the spatial pixel variance of each channel. Richer spatial information reflects greater variation in spatial pixels, leading to a larger value of α. Finally, the weights are multiplied by the normalized input feature maps, and the Sigmoid activation function is applied to map the weights to the range of (0, 1), yielding the final group weight value:

$$w_i = \frac{\alpha_i}{\sum_{j=0}^{n} \alpha_j + \delta}, \quad i = 1, 2, \cdots, n \tag{10}$$

where i represents the number of input feature map groups, and δ is a small constant added for numerical stability.

(2) Global weighting: After multiplying the different input feature maps from the feature pyramid network by their corresponding group weights, they are further multiplied by a normalized global learnable weight. The weighted feature maps are then fused to generate a single feature map. The feature values are subsequently mapped to an appropriate range through the Rectified Linear Unit (ReLU) activation function. Finally, a 1×1 convolution is applied to obtain the output feature map, which has the same dimensions as the input feature maps used in the fusion process.

$$X_{out} = \mathrm{Conv}(\mathrm{ReLU}(\textstyle\sum_{i=0}^{n} X_i^{in} \cdot v_i \cdot \omega_i)) \tag{11}$$

$$v_i = \frac{v_j}{\sum_{j=0}^{n} v_j + \delta}, \quad i, j = 1, 2, \cdots, n \tag{12}$$

where v_i represents the learnable parameter in the global weighting, δ is a small constant, and w_i is the weight obtained from the group weighting.

3. Experiment and Results

3.1. Experimental Configuration and Evaluation Indicators

The experiments in this study were conducted using an NVIDIA GeForce RTX 4060 Laptop 8G GPU on the Windows 11 operating system. The experimental environment was configured with CUDA 11.0, Python 3.8, and PyTorch 2.1.0. The model was constructed, trained, and tested using the PyCharm 2023.2.1 deep learning framework. In addition, the hyperparameter settings for training and testing across all models were kept consistent. The detailed hyperparameter settings are provided in Table 2.

Table 2. Hyperparameter information of each model in the experiment.

Hyperparameters	Values
image size	640
batch size	8
epoch	200
learning rate	0.01
Momentum	0.937
Weight_decay	0.0005

To evaluate the performance of the model, Average Precision (AP) and mean Average Precision (mAP) were selected as the evaluation metrics for the experiments. The calculations are as follows:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{13}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{14}$$

$$AP = \int_0^1 P(R)dR \quad (15)$$

$$mAP = \frac{\sum_{i=1}^{c} AP_i}{C} \quad (16)$$

where True Positives (TP) represent the number of samples correctly identified by the model as containing wood defects. False Positives (FP) refer to the number of samples incorrectly classified by the model as having defects when no defects are present. False Negatives (FN) denote the number of samples where The model fails to detect wood defects, misclassifying them as background. AP evaluates the balance between precision and recall, specifically represented by the area under the precision-recall (P-R) curve. mAP is the average of AP values across all wood defect categories and serves as a comprehensive metric to assess the overall performance of the model.

3.2. Ablation Experiments

To validate the effectiveness of the designed modules in improving the YOLOv8 algorithm for detecting wood defects, this section conducts a series of ablation experiments. To ensure experimental validity, only the introduced optimization methods were adjusted, while the remaining settings and parameters were kept constant throughout the process.

In this experimental stage, we designed four groups of comparative experiments by progressively integrating the optimization methods into the original network. The improvements brought to the YOLOv8 algorithm by the optimized models were evaluated using the evaluation metrics. The experimental results are shown in Table 3, where:

1. YOLOv8 represents the original model.
2. YOLOv8+C2f-FA denotes replacing the C2f module in the original backbone with the C2f-FA module.
3. YOLOv8+DSPPF+CSAM indicates replacing the SPPF module with the DSPPF module and adding the CSAM attention mechanism afterward.
4. YOLOv8+DCWF replaces the original Concat operation with the DCWF structure.
5. FDD-YOLO is the detection model proposed in this study.

Table 3. Results of the ablation experiment.

Methods	mAP	AP						
		Live_Knot	Marrow	Resin	Dead_Knot	Knot_with_Crack	Knot_Missing	Crack
YOLOv8	0.780	0.824	0.862	0.721	0.841	0.592	0.837	0.780
YOLOv8+C2f-FA	0.794	0.829	0.851	0.777	0.832	0.611	0.878	0.780
YOLOv8+DSPPF +CSAM	0.793	0.835	0.816	0.805	0.847	0.560	0.915	0.771
YOLOv8+DCWF	0.798	0.824	0.892	0.692	0.851	0.570	0.972	0.784
FDD-YOLO	0.823	0.865	0.896	0.923	0.890	0.669	0.800	0.715

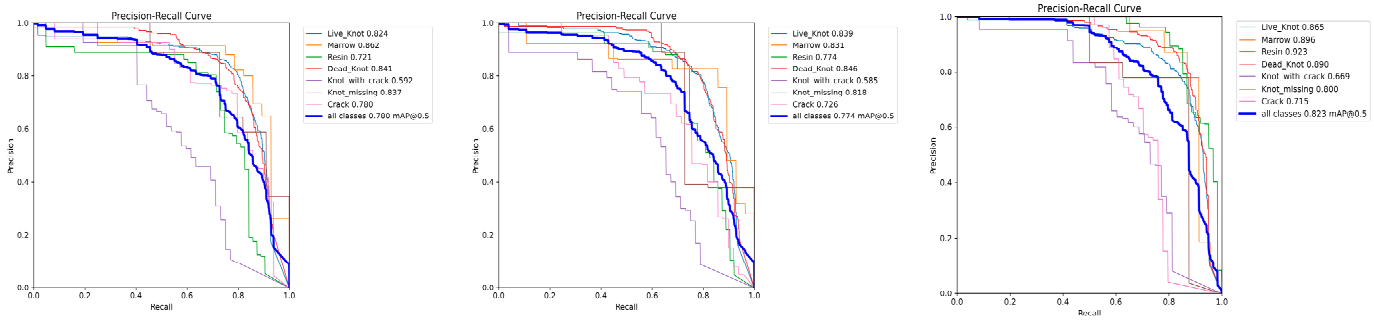
The experimental results show that the original YOLOv8 model achieved a mAP value of 0.780. When the C2f-FA module DSPPF and CSAM module and the DCWF module were respectively introduced or replaced in the network, the mAP values increased by 1.4%, 1.3%, and 1.8%, respectively. It is worth mentioning that the DCWF module significantly improves the detection of defects with large scale variations, such as Knot_Missing and Crack. For Knot_Missing, the DCWF increases AP from 83.7% (baseline) to 97.2% (+13.5%), as it enhances the model's ability to detect subtle crack patterns. For Crack, DCWF just improves AP by 0.4%, but with the highest AP, as it effectively combines multi-scale features to handle irregular missing knot shapes. Finally, when these modules were simultaneously integrated into the original YOLOv8 model, forming the proposed FDD-

YOLO model, its mAP value improved by 4.3%. Additionally, for each defect category, except for Knot_Missing and Crack, the FDD-YOLO model achieved the highest AP values. These experimental results collectively demonstrate the rationality and effectiveness of the proposed optimization strategies.

3.3. Comparative Experiments with Other Models

In the comparative experiment phase, we conducted an in-depth comparison between FDD-YOLO and several mainstream models in the field of object detection. All experiments were carried out under the same configuration environment and parameter settings to ensure the fairness and reliability of the experimental results.

The precision-recall (P-R) curve is commonly used to evaluate the classification performance of models on defects. It visually represents the AP value for each defect category, showing the relationship between precision and recall. The area under the P-R curve (AUC-PR) is a key indicator of overall model performance, with a higher AUC-PR value indicating that the model can maintain high precision even with a high recall rate. By comparing the P-R curves of different models, we can intuitively observe the balance between precision and recall, allowing us to assess which model performs better in classification. Figure 8 presents the P-R curves of the original YOLOv8 model, the latest detection model YOLO11, and our proposed model. In terms of AP values, FDD-YOLO demonstrated significant improvements over YOLOv8 and YOLO11 in the majority of defect detection tasks. While the precision for certain categories slightly decreased for key defect categories such as Live_Knot, Marrow, and Resin, Dead-Knot, and Knot-with-Crack, the AP values improved by 2.6%, 6.5%, 14.9%, 4.4%, and 8.4%, respectively, compared to YOLO11. Furthermore, in terms of the overall metric mAP, the FDD-YOLO model also showed excellent performance, with improvements of 4.3% and 4.9% over YOLOv8 and YOLO11, respectively. These results clearly demonstrate that FDD-YOLO not only offers significant improvements in detection accuracy for specific defect categories but also achieves higher overall balance and accuracy in performance.



(a) (b) (c)

Figure 8. Precision-recall (P-R) curves: (**a**) YOLOv8; (**b**) YOLO11; (**c**) FDD-YOLO.

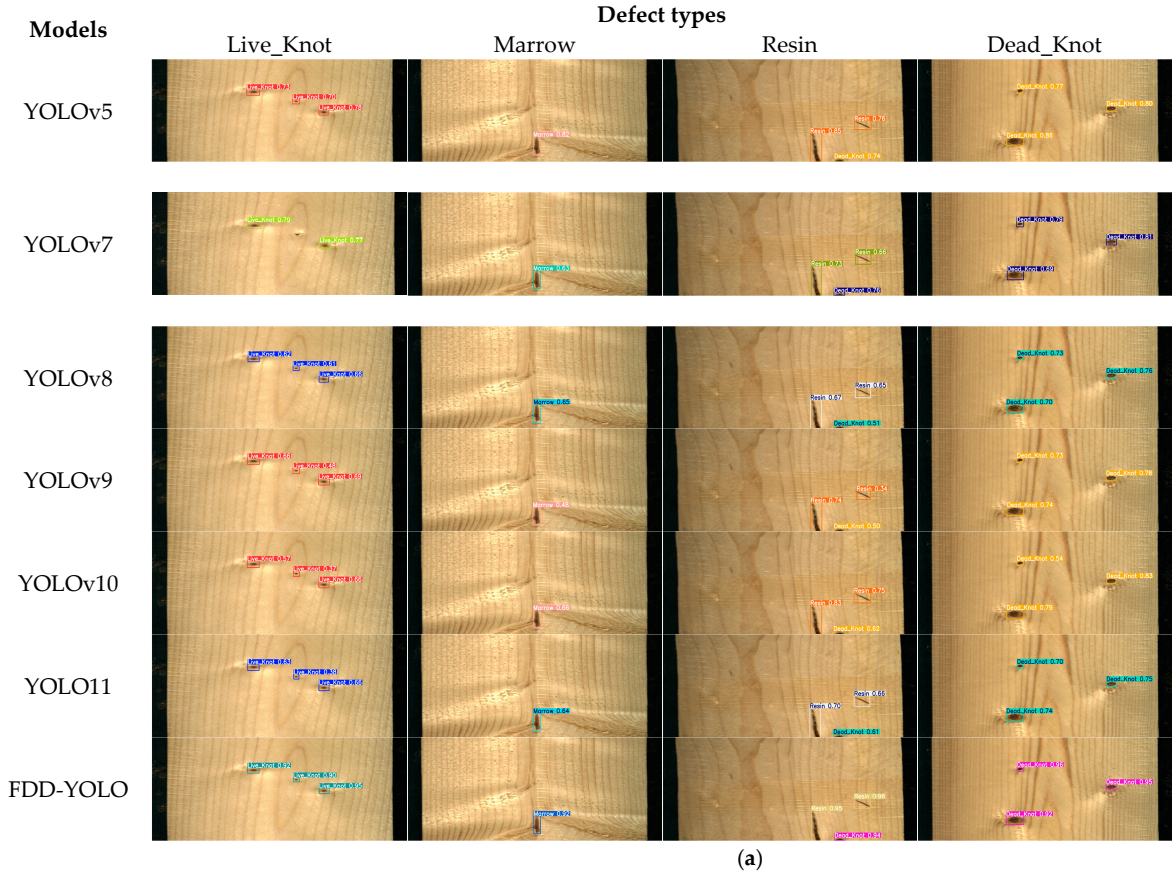
Comparison experiments with several mainstream YOLO series models (including v5, v7, v8, v9, v10, and YOLO11) were conducted; the experimental results are shown in Table 4. For AP, the FDD-YOLO model demonstrates the best performance in detecting defects such as Live_Knot, Resin, Dead_Knot, and Knot_with_Crack. For mAP, compared with YOLOv5, v7, v8, v9, v10, and YOLO11, the mAP value of the FDD-YOLO model shows improvements of 4.7%, 6.2%, 4.3%, 2.7%, 6.0%, and 4.9%, respectively. Furthermore, the paired t-test statistics were less than 0.005 (10 times) for all five groups (v5, v7, v8, v10, and YOLO11), except for YOLOv9. This result confirms that the performance improvement is statistically significant and not due to random chance.

Table 4. Comparison of various detection models.

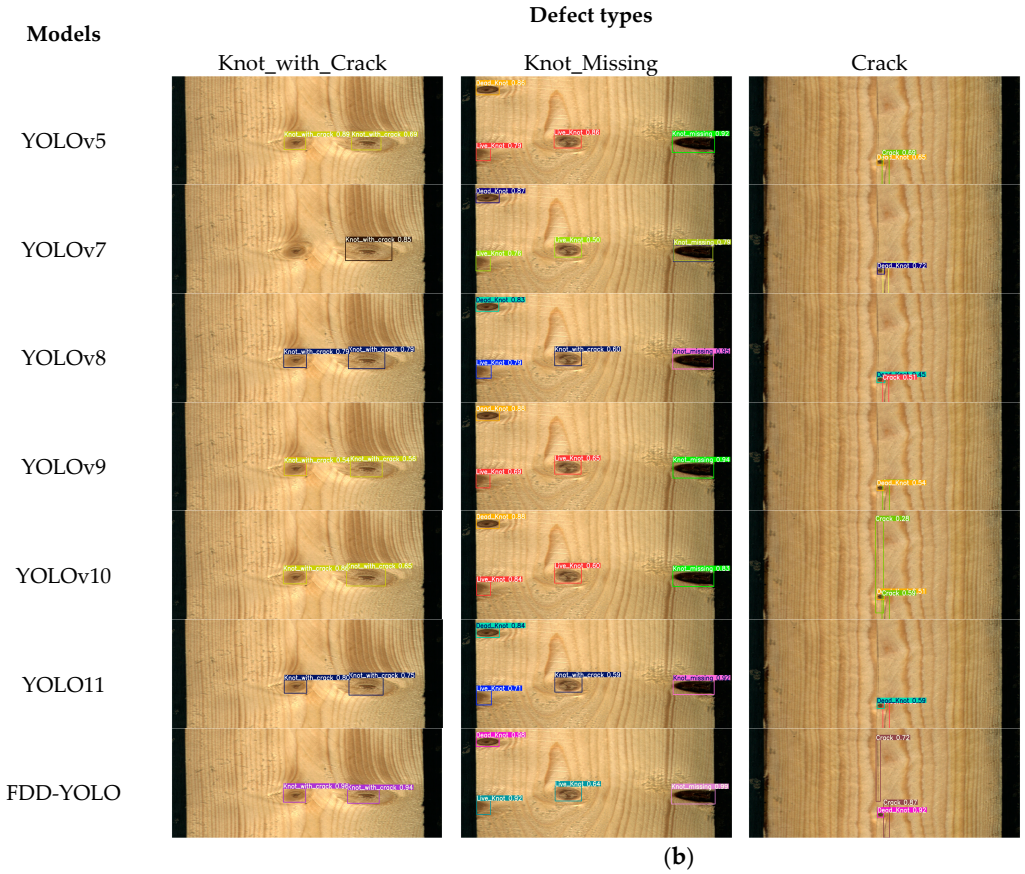
Methods	mAP	AP						
		Live_Knot	Marrow	Resin	Dead_Knot	Knot_with_Crack	Knot_Missing	Crack
YOLOv5 [37]	0.776	0.807	0.905	0.762	0.817	0.574	0.858	0.706
YOLOv7 [44]	0.761	0.807	0.876	0.782	0.84	0.563	0.708	0.755
YOLOv8 [36]	0.780	0.824	0.862	0.721	0.841	0.592	0.837	0.780
YOLOv9 [45]	0.796	0.832	0.799	0.81	0.847	0.610	0.894	0.780
YOLOv10 [46]	0.763	0.814	0.814	0.732	0.812	0.579	0.870	0.723
YOLO11 [47]	0.774	0.839	0.831	0.774	0.846	0.585	0.818	0.726
FDD-YOLO	0.823	0.865	0.896	0.923	0.890	0.669	0.800	0.715

Figure 9 presents example results of the visual comparison for seven types of defects. In Figure 9, each detection box is accompanied by a confidence score, which indicates the model's certainty about the detection result. The confidence score ranges from 0 to 1, with higher values representing greater confidence in the detection and lower values indicating more uncertainty in the model's prediction. When the confidence score is close to 1, the model shows a high level of trust in the presence of the target object within the detection box. Conversely, a confidence score of 0 means the model has completely excluded the possibility of the target's presence. The experimental results demonstrate that FDD-YOLO has a significant advantage over other models in the wood defect detection task. For instance, for the Dead_Knot defect, the detection confidence of the FDD-YOLO model reached 0.92, 0.96, and 0.95, while the corresponding confidence scores for the original YOLOv8 model were only 0.70, 0.73, and 0.76, with improvements of 0.22, 0.23, and 0.19, respectively. Additionally, FDD-YOLO did not produce any misclassifications or missed detections in all tasks, which indicates its high reliability and accuracy in wood defect detection.

Figure 10 shows the visual results generated using Grad-CAM for YOLOv8 and FDD-YOLO in wood defect detection tasks. Grad-CAM is a visualization method used for model interpretability. By capturing the gradient information from the model, it generates corresponding heatmaps that highlight the areas of the input image most important for predicting a specific class. This technique provides an intuitive way to demonstrate the model's attention distribution and explain its decision-making process. From Figure 10, it is clear that FDD-YOLO focuses its attention more precisely on the key areas of defects during wood defect detection. The generated heatmap has deeper colors, indicating a stronger response to the target region, with attention concentrated on the critical features of the defect. In contrast, YOLOv8's heatmap displays a more scattered color distribution, suggesting that its attention spans a wider area, possibly including irrelevant regions. This difference reflects the significant advantage of the FDD-YOLO model in feature extraction and defect localization, enabling more accurate identification and capture of wood defect details, thus improving detection accuracy and reliability.

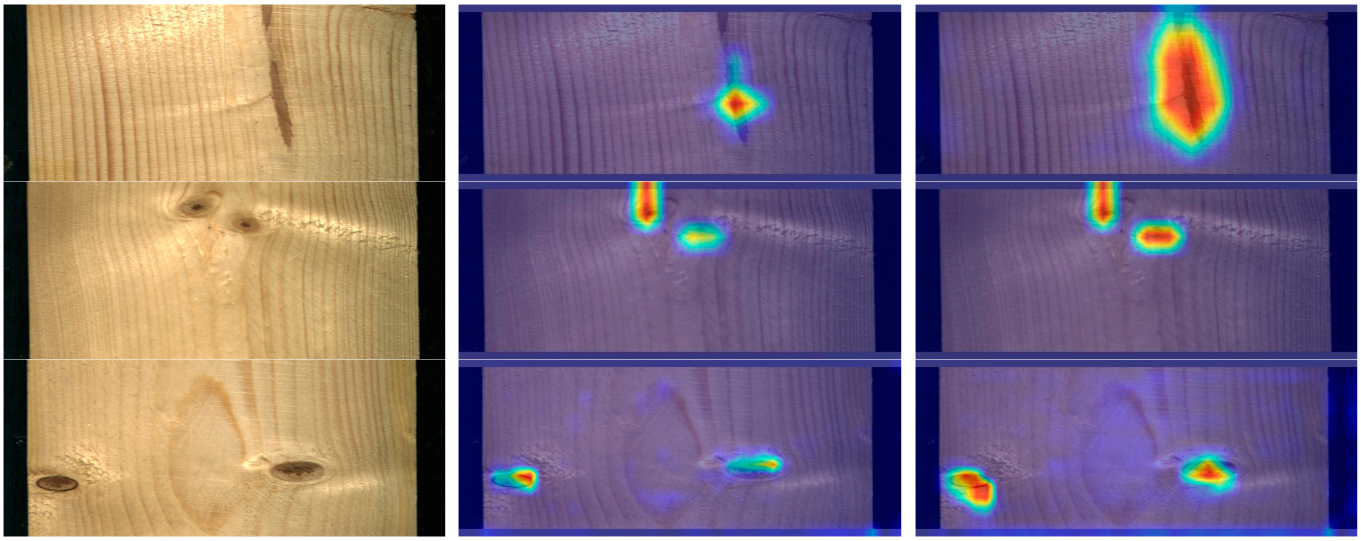


(**a**)



(**b**)

Figure 9. Example results of the visual comparison for seven types of defects: (**a**) primary four defect types; (**b**) remaining three defect types.



(a) (b) (c)

Figure 10. Visual results of Grad-CAM for YOLOv8 and FDD-YOLO: (**a**) original image; (**b**) YOLOv8; (**c**) FDD-YOLO.

4. Conclusions

The detection of surface defects in wood is crucial for automated wood processing. To address challenges caused by the diversity, multi-scale nature, and small target characteristics of wood surface defects, this study proposes a novel detection model, FDD-YOLO, based on YOLOv8 with enhancements in feature extraction and feature fusion stages. In the feature extraction stage, the C2f-FA module was developed to strengthen the model's ability to extract multi-scale defect features. This module captures local extrema of small target defects by leveraging pixel-to-surrounding relationships, inter-channel connections, and global information preservation. Additionally, the DSPPF module, featuring a dual-branch structure with max pooling and average pooling, was designed to aggregate local salient features and global contextual features. The CSAM module further enhances detection by capturing global long-range dependencies, enabling differential representation of defect features, and improving the detection of small defects that are easily confused with the background. In the feature fusion stage, the DCWF module was introduced to fully utilize the channel-wise information of input feature maps while minimizing the impact of redundant information. By employing a weighted summation approach and refining cross-scale fusion, the DWCF module effectively integrates deep and shallow features, addressing variations in defect scales. The effectiveness of the proposed model was validated through ablation and comparative experiments. Ablation studies demonstrated that the inclusion of the C2f-FA module, DSPPF module, CSAM, and DCWF module led to significant improvements in mAP values. When all modules were integrated, the mAP value increased by 4.3 percentage points over the baseline YOLOv8, confirming the model's effectiveness. Comparative experiments with mainstream detection models revealed that FDD-YOLO achieved superior performance in detecting defects such as Live_Knot, Resin, Dead_Knot, and Knot_with_Crack, with significant improvements in mAP values across these categories. Visualization experiments further demonstrated the robustness of FDD-YOLO, particularly in detecting crack defects, where other YOLO series models (v5, v7, v8, v9, and YOLO11) exhibited missed detections. In contrast, FDD-YOLO consistently achieved accurate detections, addressing issues of missed and false detections. In summary, the proposed FDD-YOLO model effectively tackles challenges posed by scale variations and small target characteristics in wood surface defect detection, significantly enhanc-

ing detection accuracy and reliability. It is acknowledged that the wood surface defect dataset was acquired under specific external conditions, such as controlled lighting, which may limit the model's robustness. Therefore, future research will focus on incorporating factors that could influence the detection process, including variations in external light sources and inherent changes in the wood material itself. Additionally, we will explore strategies to mitigate these factors, aiming to enhance the model's detection accuracy in real-world conditions.

Author Contributions: B.W.: Writing—original draft, Funding acquisition, Investigation, Methodology. R.W.: Conceptualization, Writing—review and editing, Supervision, Methodology, Funding acquisition. Y.C.: Data curation, Conceptualization, Software. C.Y.: Software, Data curation. X.T.: Software, Data curation. P.S.: Resources, Investigation, Funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This study was co-supported by the Science and Technology Planning Project of Guangxi Province, China (No. 2022AC21012); the fund project of the Key Laboratory of AI and Information Processing (No. 2022GXZDSY101); the China University Industry University Research Innovation Fund-New Generation Information Technology Innovation Project (No. 2022IT018); the Scientific Research Project of Hechi University (Grant No. 2022YLXK002); the National Natural Science Foundation of China (No. 62105305), and the Fundamental Research Program of Shanxi Province (No. 20210302123068).

Data Availability Statement: Data will be made available on request.

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

References

1. Chen, Y.; Sun, C.; Ren, Z.; Na, B. Review of the Current State of Application of Wood Defect Recognition Technology. *BioResources* **2023**, *18*, 2288–2302. [CrossRef]
2. Singh, T.; Singh, A.P. A review on natural products as wood protectant. *Wood Sci. Technol.* **2012**, *46*, 851–870. [CrossRef]
3. Hashim, U.R.; Hashim, S.Z.; Muda, A.K. Automated vision inspection of timber surface defect: A review. *J. Teknol.* **2015**, *77*, 127–135. [CrossRef]
4. Li, C.; Zhang, Y.; Tu, W.; Jun, C.; Liang, H.; Yu, H. Soft measurement of wood defects based on LDA feature fusion and compressed sensor images. *J. For. Res.* **2017**, *28*, 1285–1292. [CrossRef]
5. Chun, T.H.; Hashim, U.R.; Ahmad, S.; Salahuddin, L.; Choon, N.H.; Kanchymalay, K. A review of the automated tim-ber defect identification approach. *Int. J. Electr. Comput. Eng.* **2023**, *13*, 2156–2166. [CrossRef]
6. Fan, J.N.; Liu, Y.; Yang, Y.T.; Gou, B.L. Research progress in the application of machine vision to wood defect detec-tion. *World For. Res.* **2020**, *33*, 32–37. [CrossRef]
7. Espinosa, L.; Brancheriau, L.; Cortes, Y.; Prieto, F.; Lasaygues, P. Ultrasound computed tomography on standing trees: Accounting for wood anisotropy permits a more accurate detection of defects. *Ann. For. Sci.* **2020**, *77*, 68. [CrossRef]
8. Fang, Y.; Lin, L.; Feng, H.; Lu, Z.; Emms, G.W. Review of the use of air-coupled ultrasonic technologies for nonde-structive testing of wood and wood products. *Comput. Electron. Agric.* **2017**, *137*, 79–87. [CrossRef]
9. Yang, H.; Yu, L. Feature extraction of wood-hole defects using wavelet-based ultrasonic testing. *J. For. Res.* **2017**, *28*, 395–402. [CrossRef]
10. Wang, X.P.; Thomas, E.D.; Xu, F.; Liu, Y.F.; Brashaw, B.K.; Ross, R.J. Defect Detection And Quality Assessment of Hardwood Logs: Part 2-Combined Acoustic And Laser Scanning System. *Wood Fiber Sci.* **2018**, *50*, 310–322. [CrossRef]
11. Hu, C.S.; Tanaka, C.; Ohtani, T. Locating and identifying splits and holes on sugi by the laser displacement sensor. *J. Wood Sci.* **2023**, *49*, 492–498. [CrossRef]
12. Wang, J.D.; Zhao, Q.; Liu, Y.; Zhai, F.F.; Zhao, T. Design of laser scanning and detection devices for wood sur-face defects. *For. Mach. Woodwork. Equip.* **2016**, *44*, 24–27.
13. Yu, H.; Liang, Y.; Liang, H.; Zhang, Y.Z. Recognition of wood surface defects with near infrared spectroscopy and machine vision. *J. For. Res.* **2019**, *30*, 2379–2386. [CrossRef]
14. Cao, J.; Liang, H.; Lin, X.; Zhang, Y. Potential of near-infrared spectroscopy to detect defects on the surface of solid wood boards. *BioResources* **2017**, *12*, 19–28. [CrossRef]

15. Yang, Z.; Zhang, M.M.; Li, K.; Chen, L. Rapid detection of knot defects on wood surface by near infrared spec-troscopy coupled with partial least squares discriminant analysis. *BioResources* **2016**, *11*, 2557–2567. [CrossRef]
16. Kline, D.E.; Surak, C.; Araman, P.A. Automated hardwood lumber grading utilizing a multiple sensor machine vision technology. *Comput. Electron. Agric.* **2003**, *41*, 139–155. [CrossRef]
17. Xiao, X. A Multiple Sensors Approach to Wood Defect Detection. Ph.D. Thesis, Virginia Polytechnic Institute & State University, Blacksburg, VA, USA, 2004.
18. Zhang, Y.; Liu, S.; Tu, W.; Yu, H.; Li, C. Using computer vision and compressed sensing for wood plate surface detection. *Opt. Eng.* **2015**, *54*, 103102. [CrossRef]
19. Qayyum, R.; Kamal, K.; Zafar, T.; Mathavan, S. Wood defects classification using GLCM based features and PSO trained neural network. In Proceedings of the 22nd International Conference on Automation and Computing, ICAC 2016: Tackling the New Challenges in Automation and Computing, Colchester, UK, 7–8 September 2016; pp. 273–277. [CrossRef]
20. Gao, M.; Qi, D.; Mu, H.; Chen, J. A transfer residual neural network based on ResNet-34 for detection of wood knot defects. *Forests* **2021**, *12*, 212. [CrossRef]
21. Liu, B.; Zhang, X.; Gao, Z.; Li, Z. Weld Defect Images Classification with VGG16-Based Neural Network. In *In-ternational Forum on Digital TV and Wireless Multimedia Communications*; Springer: Singapore, 2017. [CrossRef]
22. Li, Z.; Tian, X.; Liu, X.; Liu, Y.; Shi, X. A Two-Stage Industrial Defect Detection Framework Based on Im-proved-YOLOv5 and Optimized-Inception-ResnetV2 Models. *Appl. Sci.* **2022**, *12*, 834. [CrossRef]
23. An, H.; Liang, Z.; Qin, M. Wood defect detection based on the CWB-YOLOv8 algorithm. *J. Wood Sci.* **2024**, *70*, 26. [CrossRef]
24. Wang, R.; Liang, F.; Wang, B.; Zhang, G.; Chen, Y.; Mou, X. An Efficient and Accurate Surface Defect Detection Method for Wood Based on Improved YOLOv8. *Forests* **2024**, *15*, 1176. [CrossRef]
25. Yi, L.P.; Akbar, M.F.; Rosdi, B.A.; Aminudin, M.F.C.; Fauthan, M. Wood Defect Inspection on Dead Knots and Pinholes Using YOLOv5x Algorithm. In Proceedings of the 12th International Conference on Robotics, Vision, Signal Processing and Power Applications. RoViSP 2021; Lecture Notes in Electrical Engineering. Springer: Singapore, 2024; Volume 1123. [CrossRef]
26. Qiu, Y.; Ai, Z.; Lin, Y.; Xu, Z.; Liu, X. Detecting Defects of Wooden Boards by Improved YOLOv4-tiny Algo-rithm. In Proceedings of the 2021 Chinese Intelligent Systems Conference; Lecture Notes in Electrical Engineering; Springer: Singapore; 2022; Volume 805. [CrossRef]
27. Urbonas, A.; Raudonis, V.; Maskeliūnas, R.; Damaševičius, R. Automated identification of wood veneer surface defects using faster region-based convolutional neural network with data augmentation and transfer learning. *Appl. Sci.* **2019**, *9*, 4898. [CrossRef]
28. Shi, J.; Li, Z.; Zhu, T.; Wang, D.; Ni, C. Defect Detection of Industry Wood Veneer Based on NAS and Mul-ti-Channel Mask R-CNN. *Sensors* **2020**, *20*, 4398. [CrossRef]
29. Ding, F.; Zhuang, Z.; Liu, Y.; Jiang, D.; Yan, X.; Wang, Z. Detecting defects on solid wood panels based on an improved SSD algorithm. *Sensors* **2020**, *20*, 5135. [CrossRef] [PubMed]
30. Wang, X.; Zhang, Z.; Liu, H. Deep Learning Based Robot Detection and Grinding System for Veneer Defects. In *Transactions on Intelligent Welding Manufacturing (RWIA 2020)*; Springer: Singapore, 2022. [CrossRef]
31. Wieruszewski, M.; Turbański, W.; Mydlarz, K.; Sydor, M. Economic Efficiency of Pine Wood Processing in Fur-niture Production. *Forests* **2023**, *14*, 688. [CrossRef]
32. Hu, J.; Song, W.; Zhang, W.; Zhao, Y.; Yilmaz, A. Deep learning for use in lumber classification tasks. *Wood Sci. Technol.* **2019**, *53*, 505–517. [CrossRef]
33. He, T.; Liu, Y.; Xu, C.; Zhou, X.; Hu, Z.; Fan, J. A fully convolutional neural network for wood defect location and identification. *IEEE Access* **2019**, *7*, 123453–123462. [CrossRef]
34. Sun, P. Wood Quality Defect Detection Based on Deep Learning and Multicriteria Framework. *Math. Probl. Eng. Theory Methods Appl.* **2022**, *2022*, 4878090. [CrossRef]
35. Kodytek, P.; Bodzas, A.; Bilik, P. A Large-Scale Image Dataset of Wood Surface Defects for Automated Vi-sion-Based Quality Control Processes. *F1000Research* **2021**, *10*, 581. [CrossRef] [PubMed]
36. Terven, J.; Cordova-Esparza, D. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *arXiv* **2024**, arXiv:2304.00501v6.
37. Jocher, G.; Chaurasia, A.; Qiu, J.; Stoken, N. YOLOv5 v6.0. GitHub Repository. 2021. Available online: https://github.com/ultralytics/yolov5/releases/tag/v6.0 (accessed on 17 December 2024).
38. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nash-ville, TN, USA, 11–15 June 2020; pp. 11531–11539. [CrossRef]
39. Varghese, R.; Sambath, M. YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. In Proceedings of the 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, 18–19 April 2024; pp. 1–6. [CrossRef]

40. Dai, J.F.; Qi, H.Z.; Xiong, Y.W.; Li, Y.; Zhang, G.D.; Hu, H. Deformable Convolutional Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 764–773. [CrossRef]
41. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bot-tlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520. [CrossRef]
42. Hendrycks, D.; Gimpel, K. Gaussian Error Linear Units (GELUs). *arXiv* **2023**, arXiv:1606.08415v5.
43. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 10778–10787. [CrossRef]
44. 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 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 18–22 June 2023; pp. 7464–7475. [CrossRef]
45. Wang, C.Y.; Yeh, I.H.; Mark Liao, H.Y. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. In *Computer Vision—ECCV 2024*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2024; Volume 15089. [CrossRef]
46. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. YOLOv10: Real-Time End-to-End Object Detec-tion. *arXiv* **2024**, arXiv:2405.14458v2.
47. Jocher, G.; Qiu, J. Ultralytics YOLO11. 2024. Available online: https://github.com/ultralytics/ultralytics (accessed on 10 December 2024).

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.