

Article

An Efficient and Accurate Surface Defect Detection Method for Wood Based on Improved YOLOv8

Rijun Wang ^{1,2} , Fulong Liang ^{1,2}, Bo Wang ^{2,3,*}, Guanghao Zhang ^{1,2}, Yesheng Chen ^{1,2} and Xiangwei Mou ^{1,2,*}

¹ School of Teachers College for Vocational and Technical Education, Guangxi Normal University, Guilin 541004, China; rijunwang@mailbox.gxnu.edu.cn (R.W.); fl15350630563@163.com (F.L.); 18813209663@163.com (G.Z.); 17707734324@163.com (Y.C.)

² Key Laboratory of AI and Information Processing, Hechi University, Yizhou 546300, China

³ School of Artificial Intelligence and Smart Manufacturing, Hechi University, Yizhou 546300, China

* Correspondence: 05041@hcnu.edu.cn (B.W.); xwmou@mailbox.gxnu.edu.cn (X.M.)

Abstract: Accurate detection of wood surface defects plays a pivotal role in enhancing wood grade sorting precision, maintaining high standards in wood processing quality, and safeguarding forest resources. This paper introduces an efficient and precise approach to detecting wood surface defects, building upon enhancements to the YOLOv8 model, which demonstrates significant performance enhancements in handling multi-scale and small-target defects commonly found in wood. The proposed method incorporates the dilation-wise residual (DWR) module in the trunk and the deformable large kernel attention (DLKA) module in the neck of the YOLOv8 architecture to enhance the network's capability in extracting and fusing multi-scale defective features. To further improve the detection accuracy of small-target defects, the model replaces all the detector heads of YOLOv8 with dynamic heads and adds an additional small-target dynamic detector head in the shallower layers. Additionally, to facilitate faster and more-efficient regression, the original complete intersection over union (CIoU) loss function of YOLOv8 is replaced with the IoU with minimum points distance (MPDIoU) loss function. Experimental results indicate that compared with the YOLOv8n baseline model, the proposed method improves the mean average precision (mAP) by 5.5%, with enhanced detection accuracy across all seven defect types tested. These findings suggest that the proposed model exhibits a superior ability to detect wood surface defects accurately.

Keywords: deep learning; wood defects; small target detection; YOLOv8; attention mechanism



Citation: 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. <https://doi.org/10.3390/f15071176>

Academic Editor: Laurence Schimleck

Received: 7 June 2024

Revised: 28 June 2024

Accepted: 4 July 2024

Published: 6 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As human society increasingly emphasizes sustainable development, the market demand for environmentally friendly green building materials continues to grow. Wood, being a natural, renewable, and eco-friendly material, is gaining prominence. Its unique grain structure, excellent physical properties, and ease of processing make it highly valuable. Consequently, wood is extensively utilized not only in the construction industry but also in furniture manufacturing, shipbuilding, artistic sculpture, and other fields, underscoring its significant importance [1,2].

However, defects such as cracks, decay, knots, and discoloration inevitably occur as trees grow. These defects not only affect the aesthetics of the wood but also significantly reduce its service life and structural strength, adding a considerable degree of difficulty to its processing. In addition to efforts aimed at preventing defects during wood growth, accurate and efficient detection of defects that emerge during wood processing is crucial. This approach is vital for ensuring wood quality, improving wood utilization rates, and promoting the healthy development of the timber industry [3].

The development of wood defect detection technology not only helps identify and eliminate defective products promptly during wood processing and use, thereby reducing resource waste and environmental pollution; it also lays a scientific foundation for the

rational utilization and processing of wood. Advanced detection technologies allow for accurate assessment of the type, degree, and distribution of wood defects, offering strong support for the classification, grading, and rational use of wood [4].

Wood defect detection techniques can generally be classified into traditional non-destructive testing (NDT) methods and machine-learning-based computer vision techniques. Traditional NDT methods, such as stress waves [5–7], X-rays [8–10], and ultrasound [11,12], typically identify defects by detecting differences in the physical structure compared to the normal wood. However, these methods often fail to identify discoloration-type defects effectively because such defects do not involve significant changes in the wood's physical structure. Furthermore, traditional NDT techniques face several limitations when applied in industrial settings. First, the detection accuracy of these methods needs significant improvement, particularly for complex or hidden defects. Second, their detection efficiency is relatively low, which poses challenges for meeting the high-throughput demands of factories or enterprises. Additionally, these techniques require operators with a high level of expertise to make accurate judgments and interpretations, adding to the operational complexity and limiting their widespread application [13].

Machine-learning-based computer vision techniques include approaches utilizing both traditional machine learning algorithms and deep learning. Traditional machine learning algorithms often use manually extracted features for training and classification. Initially, wood images undergo preprocessing steps such as noise removal and contrast enhancement to enhance image quality. Subsequently, features such as color, texture, and geometry are extracted from the images using specific algorithms or techniques. These extracted feature datasets are then used to train various machine learning models, such as decision trees and support vector machines, enabling the models to learn how to accurately classify and localize wood defects [14]. This approach requires prior identification of defective regions and manual involvement in the feature selection and extraction process. Furthermore, its effectiveness is often constrained by the accuracy and comprehensiveness of the extracted features. Additionally, this method may heavily rely on artificial features and prior knowledge, utilizing traditional machine learning algorithms such as grayscale covariance matrix (GLCM) [15,16], image segmentation [17,18], support vector machines (SVM) [19,20], and wavelet neural networks [21,22]. Deep-learning-based methods utilize neural network models to automatically learn feature representations from raw wood images. Specifically, convolutional neural networks (CNNs), a type of deep learning model, excel at automatically extracting multi-level features from images. By training with large datasets, these models optimize their parameters to achieve accurate detection of wood defects. This approach eliminates the manual feature extraction step, enabling the automatic learning of complex feature representations. Moreover, deep learning methods demonstrate higher efficiency and accuracy, particularly when dealing with large-scale datasets [23]. Wang, M. et al. [24] proposed a feature fusion network model named TSW (i.e., triplet attention mechanism, small target detection head, Wise-IoU loss function)-YOLOv8n, which is based on the YOLOv8 algorithm. This model integrates an attention mechanism and loss function optimization to enhance wood defect detection. The proposed algorithm achieves a high recognition rate, with a mAP of 91.10% and an average detection time of 6 ms. This performance represents a 5.1% improvement in mAP and a reduction of 1 ms in average detection time compared to the original model. However, the detection accuracy of the model for Quartzzity defects, one of the common defects in wood, stands at 68.5%, indicating the necessity for further improvements. Cui, W. et al. [25] introduced a novel network named cascade center of gravity YOLOv7 (CCG-YOLOv7) designed to improve the accuracy of detecting small defects on wooden boards. This enhanced network effectively identifies surface imperfections such as knots, scratches, and mold. Additionally, the model exhibits robust performance in detecting these defects under different lighting conditions, which is crucial for industrial production. Lim et al. [26] introduced a set of lightweight and efficient CNN models based on the YOLOv4-Tiny architecture for rapid and precise wood defect detection. They emphasized the redundancy of large and complex

CNN models in industrial settings. Through an iterative pruning and recovery process, they reduced model parameters by 88% while maintaining accuracy comparable to state-of-the-art (SOTA) methods. As a result, the model operates efficiently on inexpensive general-purpose embedded processors, facilitating almost real-time inference without external hardware accelerators.

In summary, deep-learning-based target inspection algorithms have become the preferred method for batch wood defect detection in factories [27]. This approach does not rely on expensive equipment or require operators to possess extensive professional expertise; instead, it only necessitates inputting images of defective wood into a computer, which then processes these images through a pre-trained model to localize and classify defects automatically, outputting the processed images. This method simplifies the operation process while enhancing inspection efficiency and accuracy, thereby providing robust support for quality control and safe production in the wood processing industry.

In view of this, this paper focuses on the detection of seven common types of wood defects: Live_Knot, Marrow (in the field of wood processing, the term “pith” is more commonly used; however, since the dataset we employ refers to this particular defect as “Marrow”, we have chosen to retain the nomenclature “Marrow” for consistency in the present study.), Resin, Dead_Knot, Knot_with_crack, Knot_missing, and Crack. To address these complex detection challenges, we designed a specialized algorithm for wood defect detection based on the YOLOv8 framework. To address the issue of significant scale variations among different defects, even within the same defect type, in wood defect datasets, we integrated the DWR module and Deformable-LKA into the C2f module of YOLOv8, enhancing the model’s ability in multi-scale feature extraction and fusion. Furthermore, the introduction of a dynamic detection head and the MPD loss function significantly improved the model’s detection accuracy for small targets and regression efficiency. Our contributions are as follows:

1. An improved YOLOv8-based model for wood surface defects detection is proposed.
2. The C2f modules in the backbone and neck of YOLOv8 are, respectively, substituted with the C2f-DWR and C2f-DLKA modules to improve the network’s ability to extract multi-scale defects and to fuse defect feature information.
3. The detection head of YOLOv8 is upgraded to dynamic detection head, and a dynamic detection head is added in the shallower layer of the network to enhance the network’s ability to detect small targets.
4. The loss function is replaced with the MPD loss function to help the model achieve faster and more efficient regression.

2. Materials and Method

2.1. Wood Surface Defect Dataset

The dataset used in our study was a wood surface defect dataset produced by VSB-Technical University of Ostrava [28]. The original dataset consisted of 20,275 images, including 1992 images without any defects and 18,283 images with one or more surface defects. It covered the 10 most common types of wood defects. From the original dataset, 3600 images with a resolution of 2800×1024 were screened. Three defect types—Quartzity, Blue stain, and Overgrown—were removed due to their small sample sizes.

To improve the generalization and robustness of the model and to avoid overfitting during the training process, we performed a series of data enhancement processes on these 3600 images to increase the size and diversity of the dataset. To mimic the situation where the camera is not aligned with the wooden board, the images were subjected to rotation, shear, and crop operations. To simulate different lighting conditions, the images were subjected to random adjustments in brightness and saturation. Additionally, to further enhance the dataset and mitigate overfitting, we added 2% Gaussian noise pixels to the images and randomly combined various data enhancement methods. After data enhancement, the total number of images increased to 9456. The dataset was partitioned

into training, validation, and test sets in an 8:1:1 ratio. Examples of the enhanced images are illustrated in Figure 1 below.

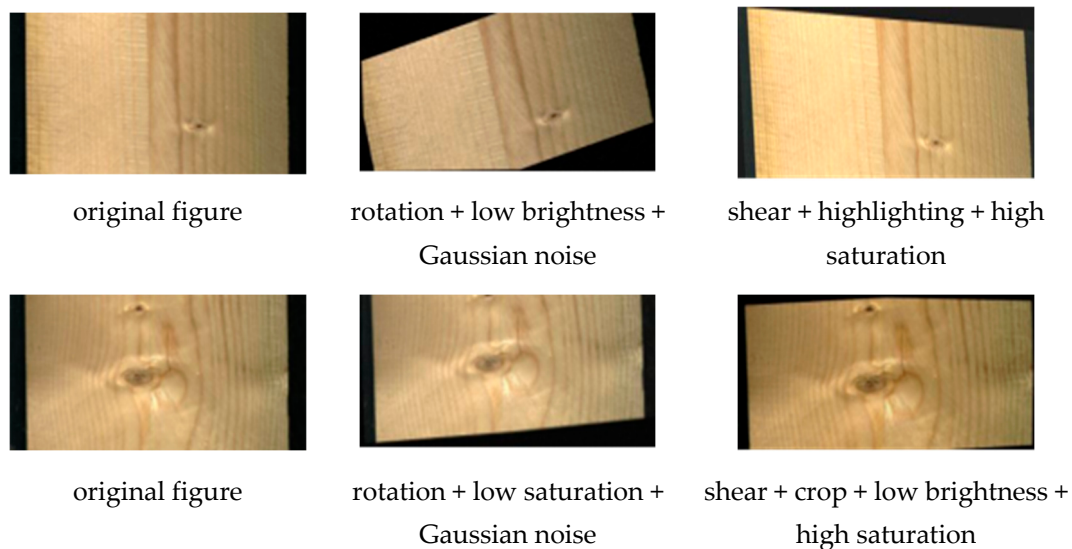


Figure 1. Examples of enhanced images of the wood surface defect dataset.

A detailed overview of the defect divisions in our screened and enhanced wood surface defect dataset is shown in Table 1.

Table 1. Overview of defect divisions in the enhanced wood surface defect dataset.

Defect Type	Number of Occurrences	Number of Images with the Defect	Over Occurrence in the Dataset
Live_Knot	10,112	5769	61
Marrow	525	490	5.2
Resin	1610	1301	13.8
Dead_Knot	7080	4686	49.6
Knot_with_crack	1256	996	10.5
Knot_missing	263	246	2.6
Crack	1306	959	10.1

2.2. Review of the YOLOv8 Algorithm

The YOLOv8 algorithm structure (as shown in Figure 2) incorporates design principles from YOLOv7 ELAN and utilizes the C2f structure from YOLOv5 in the backbone and neck networks to enhance gradient flow. In the head, YOLOv8 employs a decoupled head structure to separate the classification and detection tasks. Additionally, YOLOv8 uses the task aligned assigner strategy for positive sample allocation and introduces distribution focal loss as the regression loss. These improvements result in significant advancements in both loss computation and network structure.

The YOLOv8 algorithm performs well in detecting standard-sized objects [29]. However, its performance tends to decline in special scenarios where object sizes deviate from the standard range. Notably, other algorithms are designed specifically for complex object detection, such as algorithms for detecting small-sized objects and multi-scale objects [30,31], and these tend to outperform YOLOv8 in scenarios requiring the detection of complex targets. The wood surface dataset we used contains numerous small defects that exhibit significant variability in shape and size. Additionally, there is a high degree of similarity between different types of defects, and the difficulty of detection is further exacerbated by the overlap between defects and the multi-scale problem. These complex features pose a significant challenge to the algorithmic model for accurate defect detection.

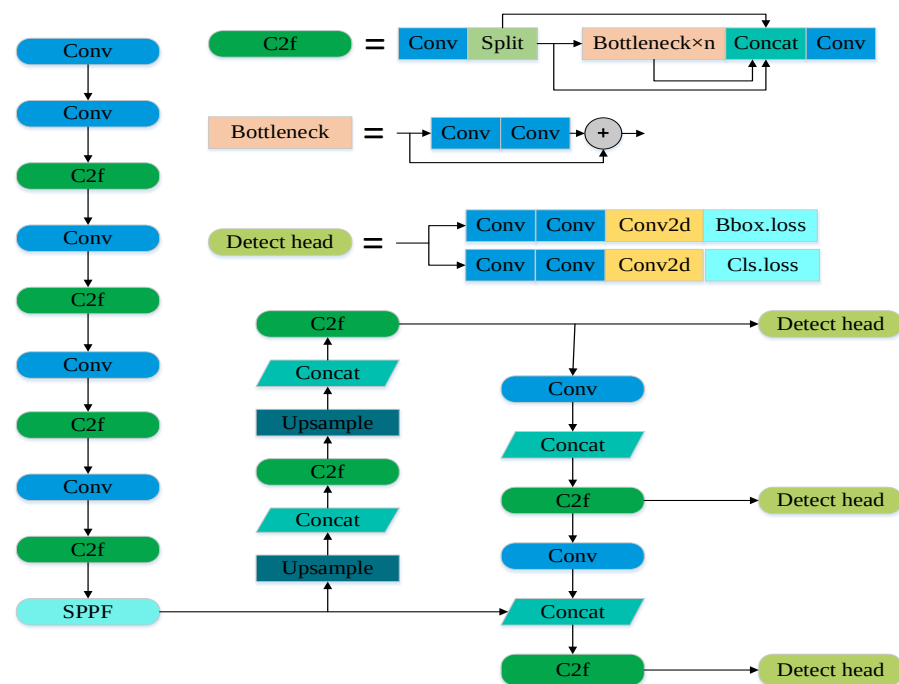


Figure 2. Structure of YOLOv8 algorithm.

2.3. C2f-DWR Module

To enhance the model's capability to extract contextual information at multiple scales, Haoran Wei proposed a robust DWR segmentation network [32], comprising two main modules: a dilated residual (DWR) module for higher-order networks and a simple inverted residual (SIR) module for lower-order networks. The DWR segmentation network divides the initial single-step feature extraction method into two phases: region residualization and semantic residualization. This methodology enhances the model's ability to achieve superior detection accuracy and faster detection rates.

The DWR module adopts a residual structure, as depicted in Figure 3. Within this framework, a two-step method is utilized to effectively extract multi-scale contextual information and then integrate the feature maps produced by multi-scale features. To simplify acquisition, the previous single-step method for acquiring multi-scale context information has been decomposed into a two-step approach.

Step I involves the generation of associated residual features from input features. In this phase, a series of compact feature maps from regions of varying sizes are created to serve as input for morphological filtering in the subsequent step. As illustrated in Figure 3, this is achieved by employing a 3×3 convolution for initial feature extraction, followed by a batch normalization (BN) layer and a ReLU layer.

Step II involves morphological filtering of features from regions of different sizes, a process referred to as semantic residualization, using a multi-rate extended depth-direction convolution. Each channel feature undergoes filtering with only a single desired receptive field in step II to avoid redundant receptive fields. In practice, the desired concise regional feature map learned in step I is efficiently matched with the receptive field size in step II for rapid processing. To carry out this step, the regional feature maps are first grouped, and then each group undergoes convolution with different expansion depths.

With the two-step region residualization–semantic residualization approach, multi-rate depth-direction convolution transitions from the demanding task of extracting extensive semantic information to the simpler task of morphological filtering with the desired receptive field on each succinctly expressed feature map. Region-based feature mapping simplifies the role of depth-direction dilation convolution to mere morphological filtering, thus streamlining the learning process. This approach allows for more efficient preservation of multiscale context information. Following the mapping of multiscale

context, multiple outputs are then aggregated. To elaborate, an aggregation process is initiated by concatenating all feature maps, subsequently followed by a batch normalization (BN) layer and point-wise convolution to integrate the features into definitive residual components. Ultimately, these consolidated residuals are superimposed onto the initial feature maps, fostering a strengthened and holistic feature portrayal. The DWR module facilitates the network in adapting more flexibly to features at different scales, thereby forming a multi-scale feature extraction mechanism capable of accurately recognizing and segmenting objects in images.

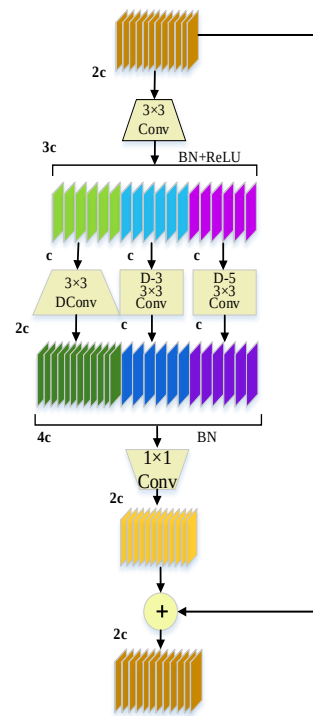


Figure 3. Structure of DWR module.

The C2f module is a crucial component in YOLOv8, representing one of the major advancements in the network structure. Compared to the C3 module in YOLOv5, the C2f module boasts fewer parameters while offering enhanced feature extraction capabilities.

To better suit the C2f module for learning the transformation of cross-scale feature points, we have made an adaptation. Therefore, in order to bolster the C2f module's multi-scale feature extraction ability, we have chosen to substitute the standard convolution in the bottleneck structure of the C2f architecture with a DWR module. The improved C2f-DWR configuration is illustrated in Figure 4.

C2f-DWR

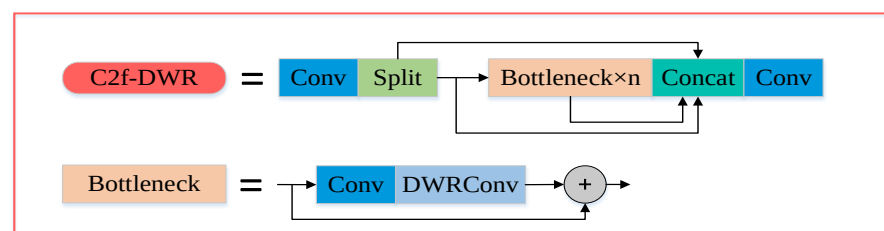


Figure 4. C2f-DWR module structure.

2.4. C2f-DLKA Module

Study [33] created a deformable large kernel attention (D-LKA attention), illustrated in Figure 5. This mechanism is a simplified attention mechanism utilizing a large convolu-

tional kernel, drawing primarily from the large kernel attention (LKA) architecture [34] and deformable convolution [35] and seamlessly integrating them. LKA decomposes a large kernel convolution into depth convolution, depth dilation convolution, and 1×1 convolution, effectively circumventing the significant computational overhead and parameter burden associated with directly employing large convolution kernels. While the direct application of large convolution kernels incurs substantial computational overhead and parameters, it also facilitates channel adaptivity and long-range dependencies, thereby effectively leveraging local context information.

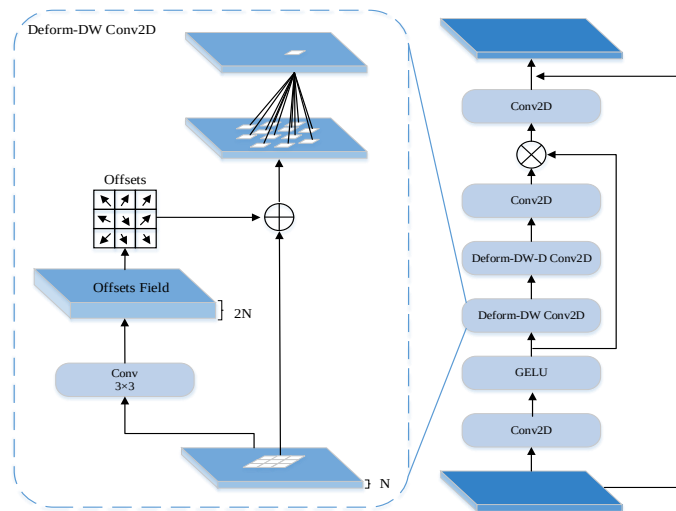


Figure 5. Structure of D-LKA attention.

The Deform-DW-D Conv2D and Deform-DW Conv2D modules within the D-LKA attention architecture apply the concept of deformable convolution to deep convolutional kernel depth dilation convolution. Deformable convolution (DCN) alters the convolutional receptive field from a fixed square to a shape closer to the actual object by introducing learnable offsets in the convolutional receptive field. Consequently, the convolutional region consistently encompasses the object's shape, enabling dynamic adjustments of the deformable convolution (DCN) based on the spatial and shape positions of the detected target. This allows for better capture of target features and adaptation to variations in target size dynamics.

D-LKA attention merges the broad receptive field of a large convolution kernel with the versatility of deformable convolution. This attention mechanism enhances the model's adaptability to irregularly shaped and multi-scale targets by dynamically adjusting the shape and size of the convolution kernel to align with various image features.

Given the challenges posed by multi-scale defects, similarities among different defect types, and defect overlap within the wood defect dataset, D-LKA attention offers several advantages. First, it leverages the self-attention capability of large kernel attention to enhance the model's comprehension of the relationships between targets, thereby addressing issues related to overlap and similarity among wood defects. Second, deformable convolution enables D-LKA attention to flexibly distort the sampling grid, facilitating the model's adaptation to multi-scale wood defects. Hence, to enhance YOLOv8's ability to fully utilize image feature information, we have chosen to further improve and optimize the C2f structure in the network. Specifically, we embed D-LKA attention into the C2f module, replacing the original position of regular convolution within the Darknet bottleneck. The resulting C2f-DLKA module structure is depicted in Figure 6.

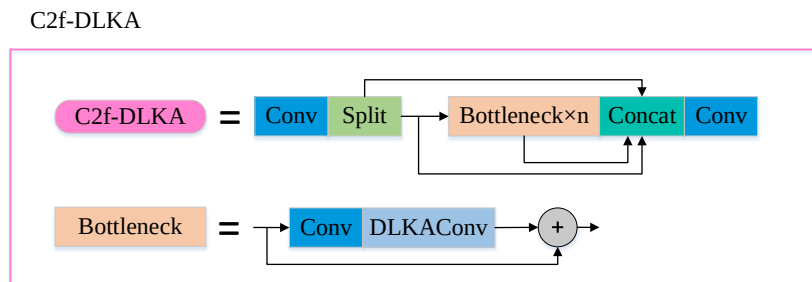


Figure 6. C2f-DLKA structure.

2.5. Dynamic-Head-Based Detection Head Module

2.5.1. Dynamic-Head Detection Head

The detection head, a pivotal component in YOLOv8, plays a crucial role in extracting both the location and category information of the target from the convolutional feature map. The YOLOv8 detection head module adopts the prevailing decoupled head structure and transitions from anchor-based to anchor-free methodology. This transition enhances the model's robustness and adaptability, allowing it to better accommodate various complex scenarios and target characteristics, thus providing a more dependable solution for practical applications. Nevertheless, it still lacks a sufficiently unified perspective to address the detection problem comprehensively.

The study in [36] introduced a dynamic head architecture that incorporates attention mechanisms into the object detection head. This design incorporates various self-attention mechanisms across feature layers to bolster scale sensitivity, spatial awareness across spatial locations, and task comprehension within output channels. This approach effectively augments the representation capabilities of the object detection head while maintaining computational efficiency.

The dynamic-head detection framework incorporates individualized attention mechanisms tailored for each feature dimension: horizontal, spatial, and channel-wise. The scale-oriented attention component specifically operates on the horizontal axis, distinguishing the significance of semantic levels to augment features tailored to individual objects' scales. The spatial-oriented attention module functions in the spatial plane (height $\times$ width), refining discriminative representations at distinct spatial positions. Concurrently, the task-oriented attention module is implemented channel-wise, allotting distinct functional channels to facilitate various tasks like classification, bounding box regression, and center/keypoint estimation. This strategy guarantees diverse convolutional kernel responses tailored to individual objects. Despite their isolated application to distinct feature dimensions, these attention mechanisms synergistically complement each other, integrating the detection framework with attention mechanisms in a unified manner.

The dynamic-head detection can be represented as follows:

$$W(\mathcal{F}) = \pi_C(\pi_S(\pi_L(\mathcal{F}) \cdot \mathcal{F}) \cdot \mathcal{F}) \cdot \mathcal{F},$$

where $\pi_L(\cdot)$, $\pi_S(\cdot)$, and $\pi_C(\cdot)$ are three different attention functions applying on dimension L , S , and C , respectively.

Figure 7 illustrates the frame structure of the dynamic-head detection head.

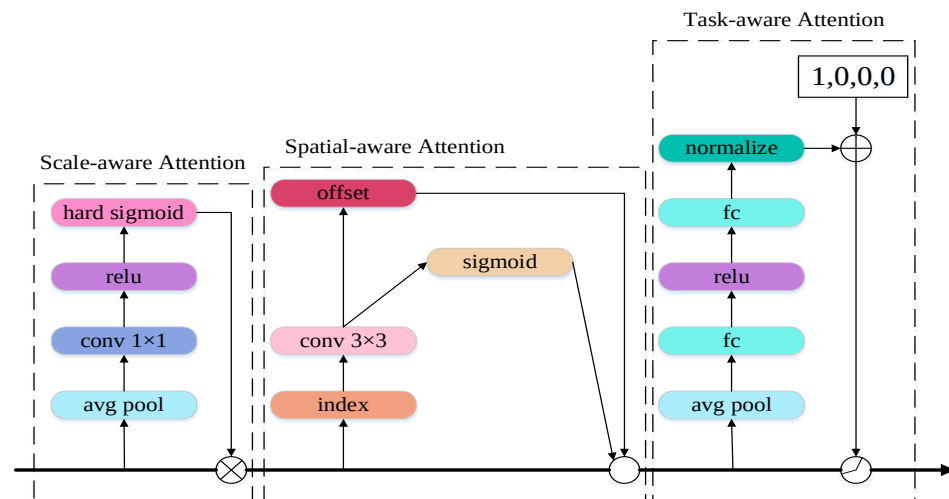


Figure 7. Structure of the dynamic-head detection head.

2.5.2. A Small Target Layer in the Detection Header

Given the prevalence of small targets within the wood defect dataset, there are inherent challenges associated with detecting such objects. These challenges stem from factors such as fewer discernible features, subtle semantic characteristics, and susceptibility to feature flooding due to constant convolution. Within the YOLO algorithm series, the relatively high downsampling multiples often hinder the accurate extraction of feature information for small objects from deeper feature maps. Moreover, the original YOLO architecture incorporates three distinct detection heads (P3, P4, P5), tailored for detecting objects of varying scales: small, medium, and large. Nevertheless, this configuration tends to compromise detection accuracy for small targets, stemming from limitations in capturing their intricate features.

The COCO dataset defines a small target as one with a size of (32, 32), while for the P3 detector head, the size of the feature layer is typically (80, 80), downsampled by a factor of 8 compared to the original input feature map size. This downsampling results in feature maps smaller than (4, 4) for small targets, leading to insufficient features and poor detection ability of the P3 detector head. To address this limitation, we opt to add a detection head in the P2 layer. First, the P2 layer, being in the shallower layer of the network, can capture more fine-grained features, which are essential for discerning the shape and texture of small targets. Second, the P2 layer usually has higher resolution, providing more spatial information, which is advantageous for detecting small targets by leveraging both shallow and deep feature information [37]. Integrating the small target detection layer enables YOLOv8 to prioritize the feature information of small targets, thereby enhancing its sensitivity in detecting such objects. This enhancement ultimately contributes to enhanced detection accuracy, as it enables the model to become more proficient in precisely recognizing and pinpointing small targets within the dataset.

To further enhance the detection of small targets, we opted to replace all detection heads in YOLOv8 with dynamic heads. With the addition of a detection head in the P2 layer, YOLOv8 now incorporates four dynamic heads. The dynamic head employs a dynamic routing mechanism, facilitating the integration of contextual information and the extraction of multi-scale features. By leveraging self-attention mechanisms, the dynamic head achieves a unified approach to scale-aware attention, spatial-aware attention, and task-aware attention. This allows for better integration of contextual information and dynamic adjustment of weights across different feature layers, thereby facilitating the extraction and recognition of multi-scale features. Consequently, the model is better equipped to detect and classify targets within wood defect datasets, particularly those comprising multi-scale and small-target defects with complex morphology.

2.6. Improvement of YOLOv8 Algorithm

To address the issue of large-scale variation between different defects in the wood defects dataset, or even among the same type of defects, we embedded the DWR (dynamic weighting receptive field) module into the C2f module in the YOLOv8 backbone. This enhancement improves the network's ability to extract multi-scale features. Second, to enable the network to better utilize the features extracted from the backbone layer and adapt to multi-scale wood defects, we introduced deformable large kernel attention (Deformable-LKA) into the C2f module at the neck of the network. Deformable-LKA combines the advantages of self-attention with the broad receptive field of large kernel attention and incorporates flexible deformable convolution. This allows for the sampling grid to be flexibly adjusted, enabling the network to better fuse feature information and helping the detector head to capture the complex features and different scale variations of wood surface defects with greater accuracy.

Next, we replaced the original detection head of YOLOv8 with a dynamic detection head. This dynamic detection head unifies the object detection head and attention across three dimensions: scale, space, and task. This helps the model reduce the false detection rate and the miss rate of small-target samples. To augment the model's proficiency in detecting diminutive objects, we incorporated a dynamic detection module into the shallower P2 layer of YOLOv8. This head combines shallow feature information with deeper features for more comprehensive detection.

Finally, we replaced the CIoU loss function used in YOLOv8 with the InnerMPD loss function. This replacement aims to help the model achieve faster and more effective regression. In summary, the method proposed in our study is shown in Figure 8.

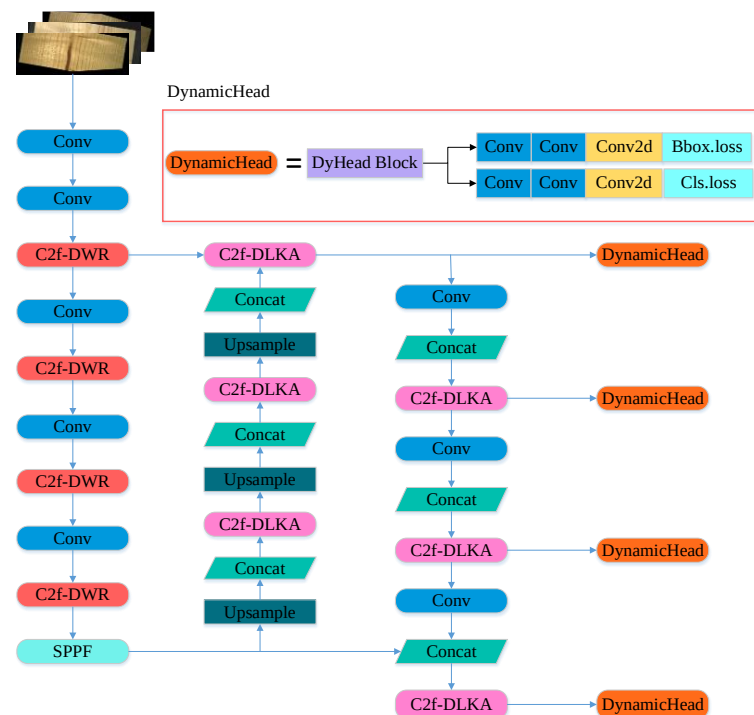


Figure 8. The improved YOLOv8 algorithm structure.

2.7. MPDIoU Loss Function Based on Inner Ideas

The bounding box regression loss function plays a critical role in target detection as it enables the model to learn to predict the precise location of the bounding box, closely aligning it with the actual bounding box of the detected target. This information is essential for accurately determining the location and area of the detected target. Although YOLOv8 utilizes CIoU [38] as its bounding box regression loss function, there are two main limitations associated with CIoU. First, in scenarios where the predicted bounding box

and the ground truth bounding box exhibit a similar aspect ratio but varying width and height dimensions, the efficacy of the CIoU loss function tends to decrease. This constraint can hinder the swift convergence and precision of the model's performance [39]. Second, the CIoU framework lacks flexibility in accommodating varying detectors and detection scenarios, thereby affecting its resilience in generalizing to new conditions [40].

Drawing inspiration from the geometric characteristics of horizontal rectangles [39], introduces a novel similarity metric called MPDIoU, which is rooted in the minimum point distance between bounding boxes. This metric comprehensively accounts for crucial factors, including overlapping and non-overlapping regions, the distance between centroids, and deviations in width and height, while offering a streamlined computational approach. Based on this, the MPDIoU-based bracketed box regression loss function, LMPDIoU, is proposed. This new loss function aims to address the limitations of CIoU by providing a more efficient and comprehensive approach to bounding box regression, ultimately improving the convergence speed, accuracy, and generalization ability of the model. The Inner-IoU loss introduced in [40] advances the conventional IoU loss by incorporating an auxiliary bounding box into its computation and introducing a scaling ratio to modulate the dimensions of this auxiliary box, thereby enhancing its effectiveness. This approach effectively speeds up the bracketed box regression process. Inner-IoU offers a more precise evaluation of overlapping regions, providing a more accurate assessment. Experiments demonstrate that Inner-IoU exhibits better generalization than traditional IoU on various datasets. Applying Inner-IoU to existing IoU-based loss functions results in faster and more efficient regression.

Therefore, in order to enable the model to perform faster and more effective regression, we choose to apply Inner-IoU to the MPD loss function and replace the original CIoU loss function in YOLOv8 to further enhance the detection accuracy, improve the accuracy of the model, help the model better adapt to various shapes and sizes of objects, reduce the leakage rate of the special defective targets in the wood images, and make the model have a better generalization performance. The calculation of InnerMPDIoU is summarized in Algorithm 1.

Algorithm 1. InnerMPDIoU as bounding box losses

Parameters: $\mathcal{B}_{prd}$ —predicted box; $\mathcal{B}_{gt}$ —ground truth (GT) box;

$\mathcal{B}_{prd} = (x_1^{prd}, y_1^{prd}, x_2^{prd}, y_2^{prd})$ —predicted box coordinates; $\mathcal{B}_{gt} = (x_1^{gt}, y_1^{gt}, x_2^{gt}, y_2^{gt})$ —GT box coordinates; (x_c^{gt}, y_c^{gt}) —center point of the GT box and the inner GT box; (x_c^{prd}, y_c^{prd}) —center point of the anchor and the inner anchor; w_{gt} —width of the GT box; h_{gt} —height of the GT box; w_{prd} —width of the anchor; h_{prd} —height of the anchor; b^{gt} —auxiliary predicted box; b —auxiliary GT box; $b^{gt} = (b_l^{gt}, b_t^{gt}, b_r^{gt}, b_b^{gt})$ —auxiliary predicted box coordinates; $b = (b_l, b_t, b_r, b_b)$ —auxiliary GT box coordinates; $ratio$ —scaling factor, typically within the range of values [0.5, 1.5]; w —width of input image; h —height of input image; $\mathcal{L}_{MPDIoU}$ —loss function of MPDIoU.

Input: $\mathcal{B}_{prd} = (x_1^{prd}, y_1^{prd}, x_2^{prd}, y_2^{prd})$, $\mathcal{B}_{gt} = (x_1^{gt}, y_1^{gt}, x_2^{gt}, y_2^{gt})$, $ratio, w, h$.

Output: $\mathcal{L}_{InnerIoU}, \mathcal{L}_{MPDIoU}$

1. For the predicted box $\mathcal{B}_{prd}$, ensuring $x_2^{prd} > x_1^{prd}$ and $y_2^{prd} > y_1^{prd}$.
 2. $x_c^{gt} = \frac{x_1^{gt} + x_2^{gt}}{2}$, $y_c^{gt} = \frac{y_1^{gt} + y_2^{gt}}{2}$, $x_c^{prd} = \frac{x_1^{prd} + x_2^{prd}}{2}$, $y_c^{prd} = \frac{y_1^{prd} + y_2^{prd}}{2}$.
 3. $w_{gt} = x_2^{gt} - x_1^{gt}$, $h_{gt} = y_2^{gt} - y_1^{gt}$, $w_{prd} = x_2^{prd} - x_1^{prd}$, $h_{prd} = y_2^{prd} - y_1^{prd}$.
 4. $b_l^{gt} = x_c^{gt} - \frac{w_{gt} * ratio}{2}$, $b_r^{gt} = x_c^{gt} + \frac{w_{gt} * ratio}{2}$, $b_t^{gt} = y_c^{gt} - \frac{h_{gt} * ratio}{2}$, $b_b^{gt} = y_c^{gt} + \frac{h_{gt} * ratio}{2}$.
 5. $b_l = x_c^{prd} - \frac{w_{prd} * ratio}{2}$, $b_r = x_c^{prd} + \frac{w_{prd} * ratio}{2}$, $b_t = y_c^{prd} - \frac{h_{prd} * ratio}{2}$, $b_b = y_c^{prd} + \frac{h_{prd} * ratio}{2}$.
 6. $inter = (\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l)) * (\min(b_b^{gt}, b_b) - \max(b_t^{gt}, b_t))$.
 7. $union = (w_{gt} * h_{gt}) * (ratio)^2 + (w_{prd} * h_{prd}) * (ratio)^2 - inter$.
 8. $IoU^{inner} = \frac{inter}{union}$.
 9. $InnerMPDIoU = IoU^{inner} - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2}$.
 10. $d_1^2 = (x_1^{prd} - x_1^{gt})^2 + (y_1^{prd} - y_1^{gt})^2$, $d_2^2 = (x_2^{prd} - x_2^{gt})^2 + (y_2^{prd} - y_2^{gt})^2$.
 11. $\mathcal{L}_{IoU} = 1 - IoU$, $\mathcal{L}_{MPDIoU} = 1 - MPDIoU$.
-

In view of the above analysis, to enable the model to perform faster and more effective regression, we applied Inner-IoU to the MPD loss function, replacing the original CIoU loss function in YOLOv8. This enhancement aims to improve detection accuracy, enhance model precision, and help the model better to adapt more successfully to defects of various shapes and sizes. It reduces the leakage rate of special defective targets in wood images and provides the model with better generalization performance. The calculation of InnerMPDIoU is summarized in Algorithm 1.

3. Results and Discussion

The experimental environment configuration is outlined in Table 2. Specifically, the training parameters used in this experiment are as follows: (1) input image size = 640 pixels; (2) iteration period (the iteration period is the number of times the model traverses the entire dataset once) = 200; (3) batch size (the batch size is the number of samples used in each model weight update) = 4; (4) initial learning rate (the learning rate is used to control the step size of the optimization algorithm when updating the model parameters) = 0.001; (5) weight decay coefficient (weight decay is a regularization technique that serves to suppress model overfitting as a way to improve model generalization) = 0.0005. Two performance evaluation criteria—average precision (AP) and mean average precision (mAP)—were employed to evaluate the accuracy and the effectiveness of our method.

Table 2. Experimental environment configuration.

Configuration	Version Parameter
System Environment	Windows 11
Central Processor	9th Gen Intel® Core™ i5-9300 CPU 2.40 GHz (It is manufactured by Intel Corporation, headquartered in Santa Clara, CA, USA.)
Graphics Processor	NVIDIA GeForce GTX 1650 (It is manufactured by NVIDIA Corporation, headquartered in Santa Clara, CA, USA.)
Graphics Processor Accelerator Library	CUDA 11.7.0, CUDNN8500
Random Access Memory	8.0 GB
Deep Learning Environment	Pytorch 1.13.1
Deep Learning Frameworks	Python 3.8.19

3.1. Ablation Experiments

To assess the impact of various enhancements on the YOLOv8 model, the following variants were developed and tested:

1. YOLOv8-C2f-DWR: This replaces the regular convolution in the bottleneck structure of all neck C2f modules with the C2f-DWR module.
2. YOLOv8-C2f-DLKA: This replaces the regular convolution in the bottleneck structure of all neck C2f modules with the C2f-DLKA module.
3. YOLOv8-Dynamichead: This changes the original three detection heads of YOLOv8 to four dynamic detection heads.
4. YOLOv8-InnerMPD: This replaces the original CIoU loss function in YOLOv8 with the InnerMPD loss function.
5. Ours: Incorporates all of the above improvements into YOLOv8.

The ablation experiment results are shown in Table 3.

Through ablation experiments, the effectiveness of the proposed series of improvements was verified. Ultimately, the optimized model achieved a significant performance improvement of 5.5% compared to the baseline model. As evident from the bold text data in Table 3, the introduction of the dynamic head significantly improves the detection performance for the four defect types: Live_Knot, Morrow, Resin, and Knot_with_crack. Furthermore, the addition of the DLKA module specifically enhances the model's detection capabilities for Dead_Knot and Crack defects. Additionally, the integration of the DWR

module notably increases the detection accuracy for the Knot_missing defect. Although the optimization of the InnerMPD loss function did not achieve the highest AP value for all seven defects, it resulted in a significant improvement in detection accuracy for five of the defects, excluding Resin and Dead_Knot. Ultimately, the proposed algorithm, which integrates these optimization measures, not only attains the highest mAP value but also demonstrates an enhancement in AP values across various defect types. Specifically, the mAP values for the detection of Resin, Knot_with_crack, Knot_missing, and Crack defects were improved by 9.9%, 6.8%, 5.4%, and 11.9%, respectively, compared to the baseline model. The results fully demonstrate that the optimized network has significantly improved the accurate recognition of small-target defects (e.g., Knot_missing), defects with large-scale variations (e.g., Resin), as well as new types of defects formed by overlapping defects (e.g., Knot_with_crack) and similar defects (e.g., Crack and Resin). The improvements illustrate that our method can more effectively handle the complex and diverse characteristics of wood surface defects, leading to better performance in practical applications.

Table 3. Ablation experiment results for YOLOv8 variants.

	mAP%	AP%						
		Live_Knot	Morrow	Resin	Dead_Knot	Knot_with_Crack	Knot_Missing	Crack
YOLOv8	0.722	0.797	0.937	0.753	0.859	0.413	0.771	0.523
YOLOv8-C2f-DWR	0.752	0.801	0.977	0.758	0.853	0.472	0.891	0.511
YOLOv8-C2f-DLKA	0.749	0.797	0.957	0.748	0.866	0.490	0.757	0.627
YOLOv8-Dynamichead	0.755	0.816	0.989	0.774	0.859	0.520	0.761	0.565
YOLOv8-InnerMPD	0.741	0.800	0.964	0.706	0.842	0.506	0.802	0.564
Ours	0.777	0.814	0.958	0.852	0.868	0.481	0.825	0.642

Note: AP = average precision; mAP = mean AP.

3.2. Comparative Experiments with Other Algorithms

The performance evaluation of YOLOv5, YOLOv7, YOLOv8, and our proposed model in detecting seven defects is illustrated via precision–recall (P-R) curves in Figure 9. Notably, the P-R curve serves as a crucial indicator of a model’s performance, wherein a greater area under the curve signifies superior performance.

As evident in Figure 9, our proposed methodology exhibits notable superiority in defect detection capabilities when contrasted with the other three models. A comparative visualization of the detection efficacy of these models is depicted in Figure 10. It is evident that our method not only locates and detects all wood defects more accurately but also provides more-precise prediction frames. Furthermore, our approach outperforms other models in terms of significantly mitigating both false positives and false negatives in the detection process.

Figure 11 presents the gradient-weighted class activation map (Grad-CAM) comparison results between YOLOv8 and our proposed model. It is evident from the heatmaps that our method generates more pronounced red highlights at the defect locations, showing higher correspondence between the highlighted areas and the actual defect areas compared to YOLOv8. This outcome indicates that our method excels in detecting and localizing defect locations, leading to more accurate identification and labeling of defect areas. This underscores the effectiveness of our method in defect detection tasks.

The experimental outcomes detailed in Table 4 offer additional confirmation that our proposed algorithm surpasses other prominent defect detection methods, namely, SSD, YOLOv5, YOLOX, YOLOv7, YOLOv8, and RetinaNet, in terms of detection accuracy. Moreover, as indicated by the boldface data in the table, the mean average precision of our method outperforms other algorithms. Except for a slightly lower detection accuracy than YOLOv5 and RetinaNet in detecting Morrow and Knot_with_crack defects, respectively, the proposed algorithm achieves better results than the other six algorithms in detecting the remaining five defects.

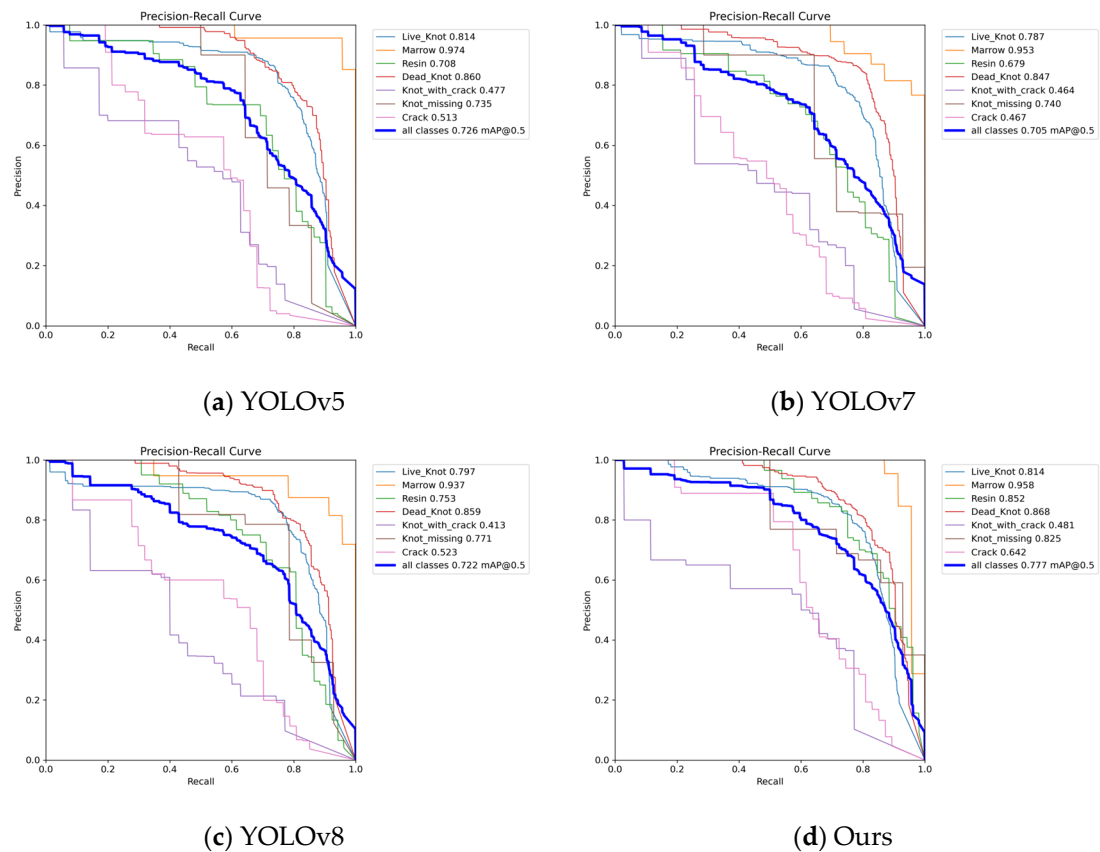


Figure 9. Precision–recall (P-R) curves.

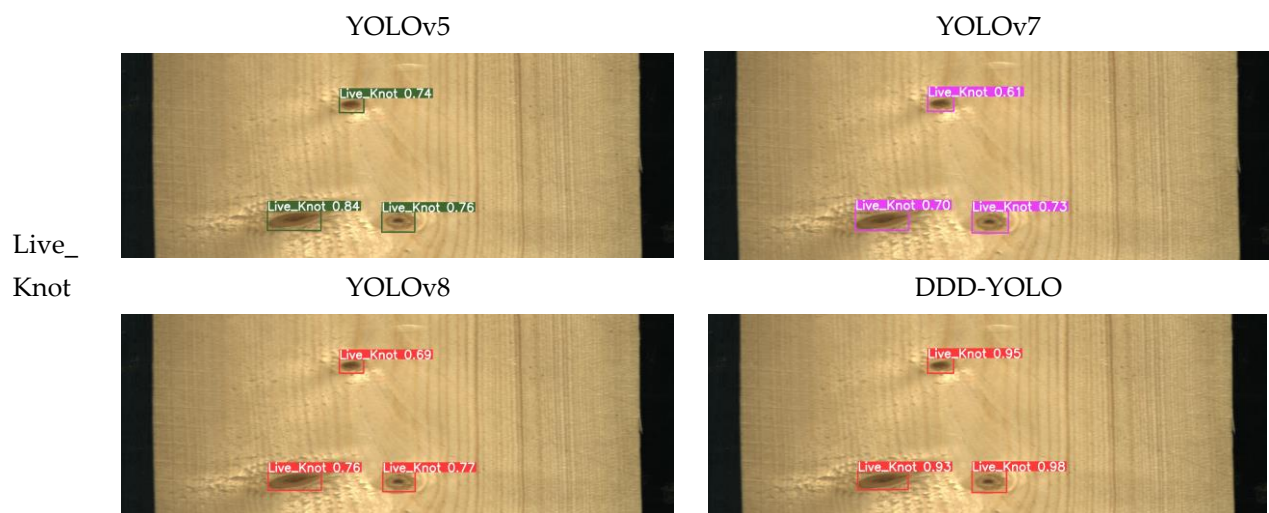


Figure 10. Cont.

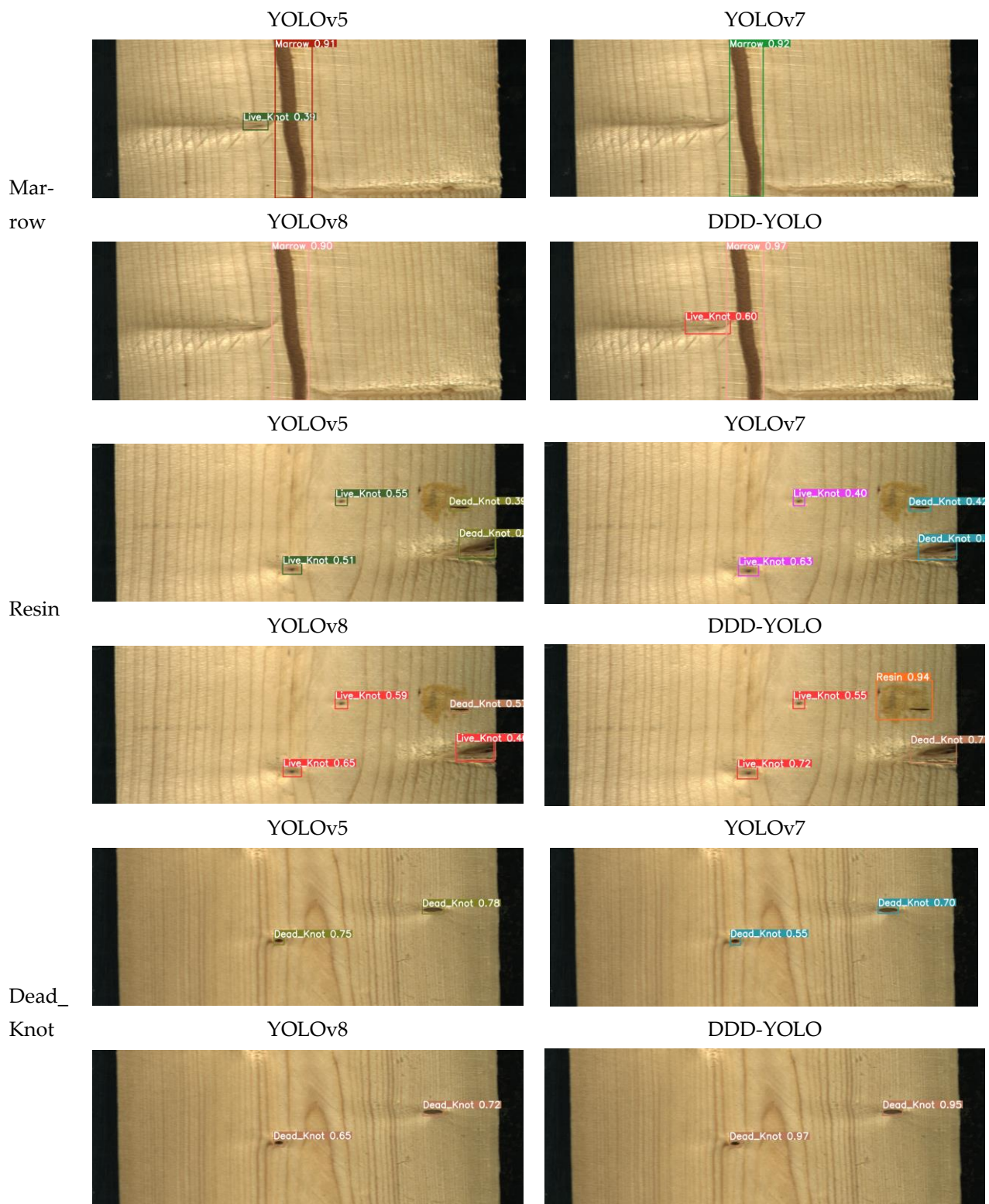


Figure 10. Cont.

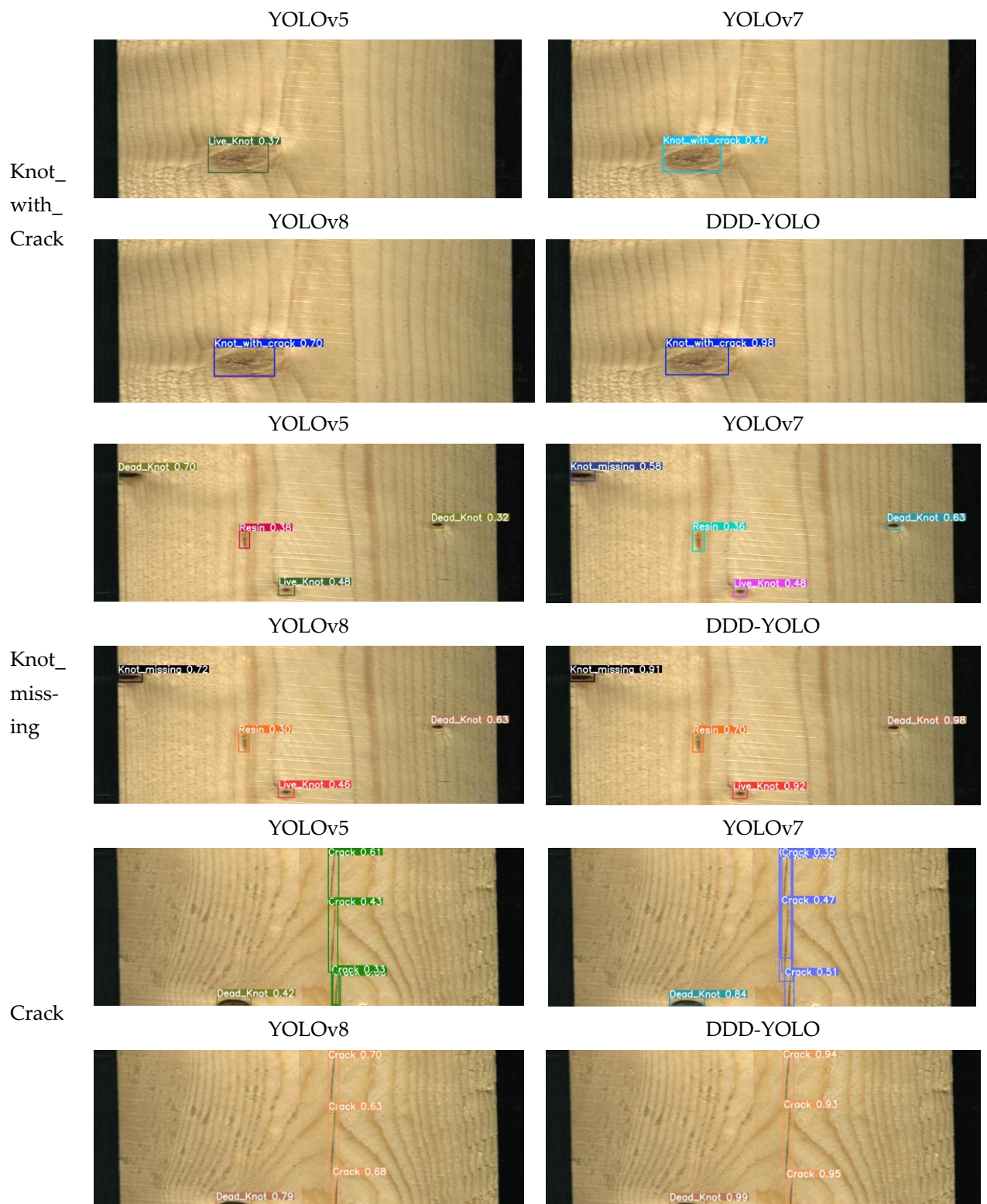


Figure 10. Comparison of detection results.

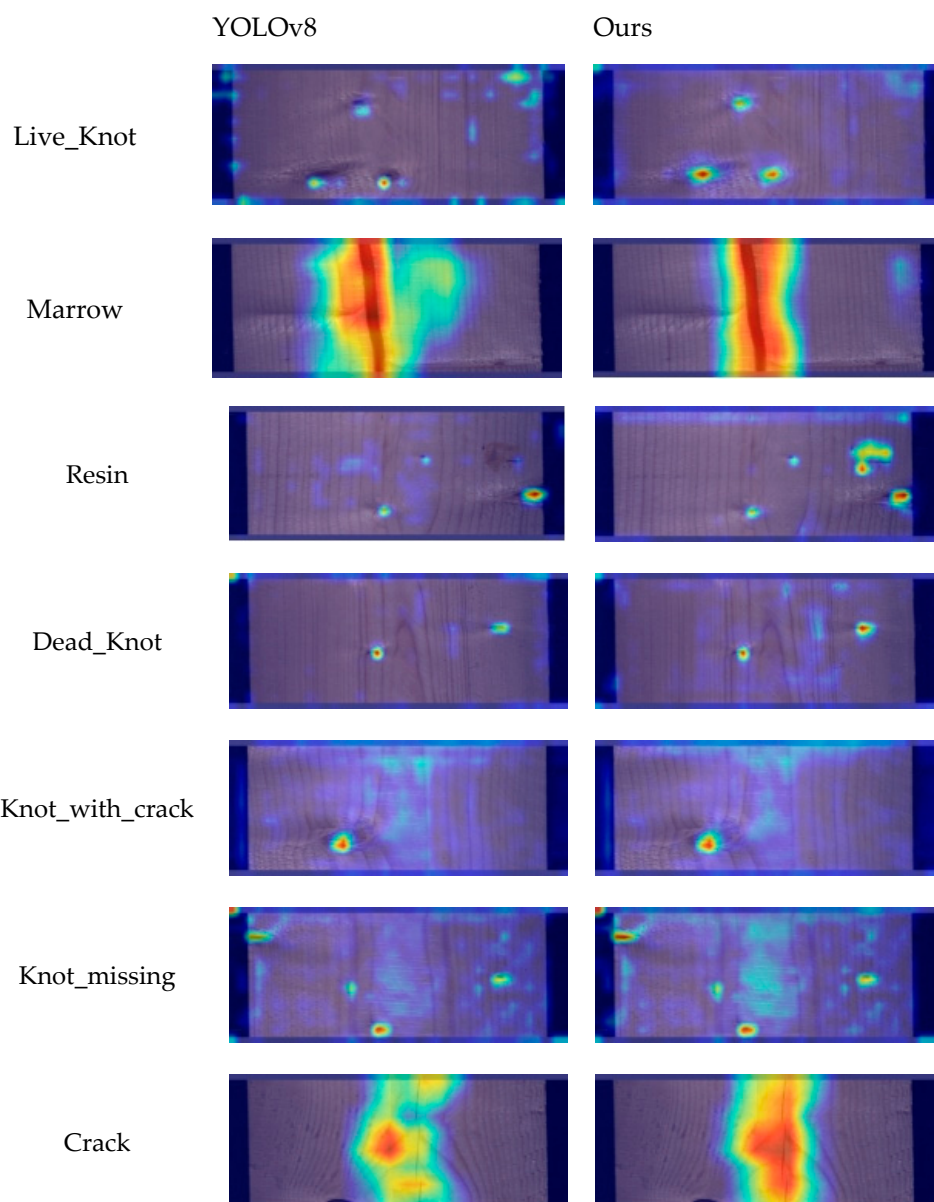


Figure 11. Grad-CAM comparison results.

Table 4. Comparison results with other algorithms.

	mAP%	AP%						
		Live_Knot	Morrow	Resin	Dead_Knot	Knot_with_Crack	Knot_Missing	Crack
YOLOv8	0.722	0.797	0.937	0.753	0.859	0.413	0.771	0.523
YOLOv7	0.705	0.787	0.953	0.679	0.847	0.464	0.740	0.467
YOLOv5	0.726	0.814	0.974	0.708	0.860	0.477	0.735	0.513
YOLOX	0.632	0.738	0.769	0.721	0.684	0.455	0.534	0.526
SSD	0.590	0.585	0.617	0.510	0.703	0.511	0.627	0.578
RetinaNet	0.515	0.673	0.439	0.728	0.651	0.519	0.433	0.161
Ours	0.777	0.814	0.958	0.852	0.868	0.481	0.825	0.642

4. Conclusions

In this study, we propose a wood surface defect detection algorithm based on improved YOLOv8, which effectively addresses the challenges posed by multi-scale and small-target defects, as well as target overlapping issues in wood defect detection. By integrating the DWR module and the DLKA module into the C2f of YOLOv8, we enhance the

network’s capability to handle multi-scale and overlapping defects, while also improving its generalization ability across different defect types. Furthermore, the incorporation of a dynamic detection head and the addition of a small target detection layer significantly enhance the network’s ability to detect small-target defects. Additionally, the utilization of the InnerMPD loss function accelerates the convergence of the network. Experimental results demonstrate that the improved model achieves a 5.5% increase in mAP accuracy compared to the original model and outperforms current mainstream algorithms.

Author Contributions: R.W.: Supervision, Writing—review and editing, Funding acquisition, Investigation, Methodology; F.L.: Writing—original draft, Software, Methodology; B.W.: Conceptualization, Validation, Supervision, Methodology, Data curation, Funding acquisition; G.Z.: Data curation, Conceptualization, Software; Y.C.: Software, Data curation; X.M.: Resources, Writing—review and editing, 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 industry-university-research innovation fund projects of China University in 2021 (No. 2021ITA10018); the fund project of the Key Laboratory of AI and Information Processing (No. 2022GXZDSY101); the Natural Science Foundation Project of Guangxi, China (No. 2018GXNSFAA050026); 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).

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

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

References

1. Mai, C.; Schmitt, U.; Niemz, P. A brief overview on the development of wood research. *Holzforschung* **2022**, *76*, 102–119. [\[CrossRef\]](#)
2. Lian, M.; Huang, Y.; Liu, Y.; Jiang, D.; Wu, Z.; Li, B.; Xu, Q.; Murugadoss, V.; Jiang, Q.; Huang, M.; et al. An overview of regenerable wood-based composites: Preparation and applications for flame retardancy, enhanced mechanical properties, biomimicry, and transparency energy saving. *Adv. Compos. Hybrid Mater.* **2022**, *5*, 1612–1657. [\[CrossRef\]](#)
3. Chen, Y.; Sun, C.; Ren, Z.; Na, B. Review of the Current State of Application of Wood Defect Recognition Technology. *BioResources* **2022**, *18*, 2288–2302. [\[CrossRef\]](#)
4. Meng, W.; Yuan, Y. SGN-YOLO: Detecting Wood Defects with Improved YOLOv5 Based on Semi-Global Network. *Sensors* **2023**, *23*, 8705. [\[CrossRef\]](#)
5. Li, X.; Qian, W.; Cheng, L.; Chang, L. A Coupling Model Based on Grey Relational Analysis and Stepwise Discriminant Analysis for Wood Defect Area Identification by Stress Wave. *BioResources* **2020**, *15*, 1171–1186. [\[CrossRef\]](#)
6. Du, X.; Li, J.; Feng, H.; Chen, S. Image Reconstruction of Internal Defects in Wood Based on Segmented Propagation Rays of Stress Waves. *Appl. Sci.* **2018**, *8*, 1778. [\[CrossRef\]](#)
7. Liao, C.; Zhang, H.; Song, X.; Chen, T.; Huang, S. The Screening Method of the Internal Defects in Wood Members of the Ancient Architectures by Hammering Sound. *BioResources* **2017**, *12*, 2711–2720. [\[CrossRef\]](#)
8. Beaulieu, J.; Dutilleul, P. Applications of Computed Tomography (CT) Scanning Technology in Forest Research: A Timely Update and Review. *Can. J. For. Res.* **2019**, *49*, 1173–1188. [\[CrossRef\]](#)
9. Wedvik, B.; Stein, M.; Stornes, J.M.; Mattsson, J. On-Site Radioscopic Qualitative Assessment of Historic Timber Structures: Identification and Mapping of Biological Deterioration of Wood. *Int. J. Archit. Herit.* **2016**, *10*, 646–662. [\[CrossRef\]](#)
10. Wang, Q.; Liu, X.; Yang, S. Predicting Density and Moisture Content of *Populus Xiangchengensis* and *Phyllostachys Edulis* Using the X-Ray Computed Tomography Technique. *For. Prod. J.* **2020**, *70*, 193–199. [\[CrossRef\]](#)
11. Palma, S.S.A.; Goncalves, R.; Trinca, A.J.; Costa, C.P.; Reis, M.N.D.; Martins, G.A. Interference from Knots, Wave Propagation Direction, and Effect of Juvenile and Reaction Wood on Velocities in Ultrasound Tomography. *BioResources* **2018**, *13*, 2834–2845. [\[CrossRef\]](#)
12. 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. [\[CrossRef\]](#)
13. Zhang, Q.; Liu, L.; Yang, Z.; Yin, J.; Jing, Z. WLSD-YOLO: A Model for Detecting Surface Defects in Wood Lumber. *IEEE Access* **2024**, *12*, 65088–65098. [\[CrossRef\]](#)
14. Wang, Q.; Zhan, X.; Wu, Z.; Liu, X.; Feng, X. The applications of machine vision in raw material and production of wood products. *BioResources* **2022**, *17*, 5532–5556. [\[CrossRef\]](#)
15. Hashim, U.R.; Hashim, S.Z.M.; Muda, A.K. Performance Evaluation of Multivariate Texture Descriptor for Classification of Timber Defect. *Optik* **2016**, *127*, 6071–6080. [\[CrossRef\]](#)

16. Wang, D.; Liu, Z.; Cong, F. Wood Surface Quality Detection and Classification Using Gray Level and Texture Features. In *Advances in Neural Networks—ISNN 2015*; Hu, X., Xia, Y., Zhang, Y., Zhao, D., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 9377, pp. 248–257, ISBN 978-3-319-25392-3. [\[CrossRef\]](#)
17. Chang, Z.; Cao, J.; Zhang, Y. A Novel Image Segmentation Approach for Wood Plate Surface Defect Classification through Convex Optimization. *J. For. Res.* **2018**, *29*, 1789–1795. [\[CrossRef\]](#)
18. Zhang, Y.; Liu, S.; Cao, J.; Li, C.; Yu, H. A Rapid, Automated Flaw Segmentation Method Using Morphological Reconstruction to Grade Wood Flooring. *J. For. Res.* **2014**, *25*, 959–964. [\[CrossRef\]](#)
19. Xu, Z.; Jia, P.; Wu, N.; Xu, K. Research on wood defect recognition technology. *Transducer Microsyst. Technol. Based GA-SVM* **2019**, *38*, 153–156. [\[CrossRef\]](#)
20. Zhang, S.; Huang, H.; Huang, Y.; Cheng, D.; Huang, J. A GA and SVM Classification Model for Pine Wilt Disease Detection Using UAV-Based Hyperspectral Imagery. *Appl. Sci.* **2022**, *12*, 6676. [\[CrossRef\]](#)
21. Yadav, A.R.; Anand, R.S.; Dewal, M.L. Binary wavelet transform-based completed local binary pattern texture descriptors for classification of microscopic images of hardwood species. *Wood Sci. Technol.* **2017**, *51*, 909–927. [\[CrossRef\]](#)
22. Zhang, Y.; Liu, S.; Cao, J.; Li, C.; Yu, H. Wood Board Image Processing Based on Dual-Tree Complex Wavelet Feature Selection and Compressed Sensing. *Wood Sci. Technol.* **2016**, *50*, 297–311. [\[CrossRef\]](#)
23. Sun, P. Wood Quality Defect Detection Based on Deep Learning and Multicriteria Framework. *Math. Probl. Eng.* **2022**, *2022*, 4878090. [\[CrossRef\]](#)
24. Wang, M.; Li, M.; Cui, W.; Xiang, X.; Duo, H. TSW-YOLO-V8n: Optimization of Detection Algorithms for Surface Defects on Sawn Timber. *BioResources* **2023**, *18*, 8444–8457. [\[CrossRef\]](#)
25. Cui, W.; Li, Z.; Duanmu, A.; Xue, S.; Guo, Y.; Ni, C.; Zhu, T.; Zhang, Y. CCG-YOLOv7: A Wood Defect Detection Model for Small Targets Using Improved YOLOv7. *IEEE Access* **2024**, *12*, 10575–10585. [\[CrossRef\]](#)
26. Lim, W.-H.; Bonab, M.B.; Chua, K.H. An Aggressively Pruned CNN Model with Visual Attention for Near Real-Time Wood Defects Detection on Embedded Processors. *IEEE Access* **2023**, *11*, 36834–36848. [\[CrossRef\]](#)
27. Wang, Y.; Zhang, W.; Gao, R.; Jin, Z.; Wang, X.-H. Recent advances in the application of deep learning methods to forestry. *Wood Sci. Technol.* **2021**, *55*, 1171–1202. [\[CrossRef\]](#)
28. Kodytek, P.; Bodzas, A.; Bilik, P. A Large-Scale Image Dataset of Wood Surface Defects for Automated Vision-Based Quality Control Processes. *F1000Research* **2022**, *10*, 581. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Zhao, E.-X. Traffic Signals Recognition Based on YOLOv8. *AIRR* **2023**, *12*, 246–254. [\[CrossRef\]](#)
30. Liu, H.; Duan, X.; Chen, H.; Lou, H.; Deng, L. DBF-YOLO: UAV Small Targets Detection Based on Shallow Feature Fusion. *IEEJ Trans. Electr. Electron. Eng.* **2023**, *18*, 605–612. [\[CrossRef\]](#)
31. Liu, H.; Sun, F.; Gu, J.; Deng, L. SF-YOLOv5: A Lightweight Small Object Detection Algorithm Based on Improved Feature Fusion Mode. *Sensors* **2022**, *22*, 5817. [\[CrossRef\]](#)
32. Wei, H.; Liu, X.; Xu, S.; Dai, Z.; Dai, Y.; Xu, X. DWRSeg: Rethinking Efficient Acquisition of Multi-Scale Contextual Information for Real-Time Semantic Segmentation. *arXiv* **2023**, arXiv:2212.01173.
33. Azad, R.; Niggemeier, L.; Huttemann, M.; Kazerouni, A.; Aghdam, E.K.; Velichko, Y.; Bagci, U.; Merhof, D. Beyond Self-Attention: Deformable Large Kernel Attention for Medical Image Segmentation. *arXiv* **2023**, arXiv:2309.00121.
34. Guo, M.-H.; Lu, C.-Z.; Liu, Z.-N.; Cheng, M.-M.; Hu, S.-M. Visual Attention Network. *arXiv* **2022**, arXiv:2202.09741. [\[CrossRef\]](#)
35. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable Convolutional Networks. *arXiv* **2017**, arXiv:1703.06211.
36. Dai, X.; Chen, Y.; Xiao, B.; Chen, D.; Liu, M.; Yuan, L.; Zhang, L. Dynamic Head: Unifying Object Detection Heads with Attentions. *arXiv* **2021**, arXiv:2106.08322.
37. Dong, H.; Song, K.; He, Y.; Xu, J.; Yan, Y.; Meng, Q. PGA-Net: Pyramid Feature Fusion and Global Context Attention Network for Automated Surface Defect Detection. *IEEE Trans. Ind. Inf.* **2020**, *16*, 7448–7458. [\[CrossRef\]](#)
38. Zheng, Z.; Wang, P.; Ren, D.; Liu, W.; Ye, R.; Hu, Q.; Zuo, W. Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation. *arXiv* **2021**, arXiv:2005.03572. [\[CrossRef\]](#)
39. Siliang, M.; Yong, X. MPDIoU: A Loss for Efficient and Accurate Bounding Box Regression. *arXiv* **2023**, arXiv:2307.07662.
40. Zhang, H.; Xu, C.; Zhang, S. Inner-IoU: More Effective Intersection over Union Loss with Auxiliary Bounding Box. *arXiv* **2023**, arXiv:2311.02877.

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.