forests

MDPI

Type of the Paper (Article, Review, Communication, etc.)

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
[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; xwmou@mailbox.gxnu.edu.cn

Abstract: Accurate detection of wood surface defects is crucial for improving the accuracy of wood grade sorting, ensuring the quality of wood processing, and protecting forest resources. This paper proposes an efficient and accurate surface defect detection method for wood based on improved You Only Look Once version 8（YOLOv8), 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 DLKA (Deformable Large Kernel Attention) 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 Intersection over Union 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 types of defects tested. These findings suggest that the proposed model exhibits a superior ability to accurately detect wood surface defects.

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

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

Copyright: © 2024 by the authors. Submitted for possible open access publication 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, as 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 holds significant importance in furniture manufacturing, shipbuilding, artistic sculpture, and other fields[1-3].

However, defects such as cracks, decay, knots and discoloration inevitably occur during the growth of wood. 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 processing. In addition to efforts aimed at preventing defects during the wood's 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 [4].

The development of wood defect detection technology not only helps to identify and eliminate defective products in a timely manner during wood processing and use, thereby

reducing resource waste and environmental pollution, but also provides a scientific basis for the rational utilization and deep 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 [5,6].

Wood defect detection techniques can be broadly categorized into traditional non-destructive testing (NDT) methods and machine learning-based computer vision techniques. Traditional NDT methods, such as stress waves [7–9], X-rays [10–12], and ultrasound [13,14], typically identify defects by detecting differences in the physical structure compared to the normal wood. However, these methods often fail to effectively identify discoloration-type defects, as 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 [15].

Machine learning-based computer vision techniques encompass methods based on traditional machine learning algorithms as well as those based on deep learning. Traditional machine learning algorithms typically rely on manually extracted features for training and classification purposes. 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 utilized to train machine learning models, including decision trees and support vector machines, enabling the models to learn how to accurately classify and localize wood defects [16]. This approach necessitates 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) [17,18], image segmentation [19,20], support vector machines (SVM) [21,22], and wavelet neural networks [23,24]. Deep learning-based methods leverage deep 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 [25]. Wang, M. et al. [26] proposed a feature fusion network model called TSW(Triplet Attention mechanism, Small target detection head, Wise-Iou loss function) -YOLOv8n, based on the YOLOv8 algorithm, which incorporates an attention mechanism and loss function optimization for detecting wood defects. The proposed algorithm achieves a high recognition rate, with the 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 Quartzity defects, one of the common defects in wood, stands at 68.5%, indicating the necessity for further improvements.Cui, W. et al. [27] proposed a novel network called the Cascade Center of Gravity YOLOv7 (CCG-YOLOv7) to enhance the accuracy of detecting small defects on wood floors. This improved network effectively identifies small defects on the surface of wooden boards, including knots, scratches, and mold. The proposed model also demonstrates robust performance in detecting surface defects under varying illumination conditions, which is of great significance for industrial production. Lim. et al. [28] proposed a

Commented [1]: A sample of wood defect Q can be seen in literature 26 as follows:



Quartzity

set of lightweight and efficient CNN models based on the YOLOv4-Tiny architecture for fast and accurate wood defect detection, acknowledging that large and complex CNN models are unnecessary in an industrial environment. By implementing an iterative pruning and recovery process, they reduced the model parameters by 88% while maintaining accuracy comparable to most state-of-the-art (SOTA) methods. Consequently, the model can run on a low-cost general-purpose embedded processor, enabling near real-time inference without the need for external hardware accelerators.

In summary, deep learning-based target inspection algorithms have become the preferred method for batch wood defect detection in factories. [29]This approach does not rely on expensive equipment or require operators to have 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 automatically localize and classify defects, 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, 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.

Commented [2]: Marrow is a wood defect, samples of which can be seen in the literature30 : as shown below. Since it was named Marrow by the authors in both the original paper and the dataset, we have chosen to continue to use this name.
(I) Marrow



Commented [3]: Since the YOLOv8 has three detector heads to be exact, and we replaced them all with dynamic detector heads. To avoid ambiguity, we thought it better to leave it unchanged.

2. Materials and Method

2.1. Wood surface defect Dataset

The dataset used in our study was a dataset of wood surface defect produced by VSB-Technical University of Ostrava [30]. The original dataset consisted of 20,275 images, including 1,992 images without any defects and 18,283 images with one or more surface defects. It covered the 10 most common types of wood defects: Live_Knot, Dead_Knot, Quartzity, Knot_with_crack, Knot_missing, Crack, Overgrown, Resin, Marrow, and Blue stain. From the original dataset, 3,600 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 3,600 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 9,456. The dataset was then divided into training, validation, and test sets in a ratio of 8:1:1. Figure 1 below shows examples of the enhanced images.





original figure	Rotation + Low brightness + Gaussian noise	Shear + Highlighting + High Saturation





original figure	Rotation + Low saturation + Gaussian noise	Shear + Crop + Low Brightness + High Saturation

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 as 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	10112	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 [31]. 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 [32,33], 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.

Commented [4]: To avoid ambiguity, we have deleted this section

Commented [5]: To avoid ambiguity, we have deleted this section

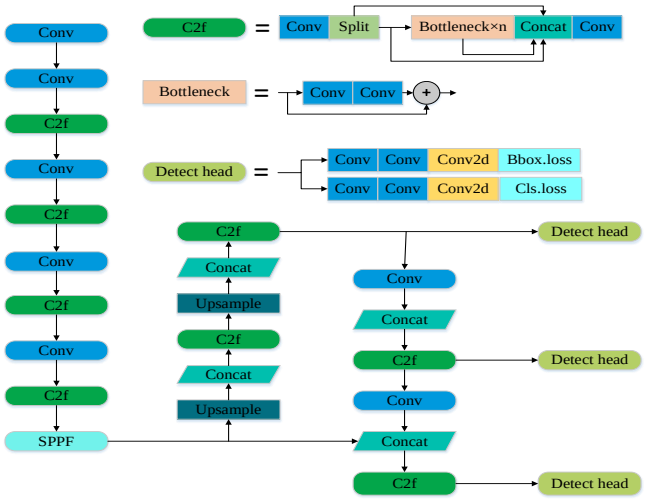


Figure 2. Structure of YOLOv8 algorithm

2.3. C2f-DWR module

To enhance the model's capability to extract contextual information at multiple scales, Haoran Wei proposes a robust DWR segmentation network [34], 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 decomposes the original single-step feature extraction method into two steps: region residualization and semantic residualization. This approach enables the model to achieve higher detection accuracy and faster detection rates.

The DWR module is structured in a residual manner, as illustrated in Figure 3. Within the residual, a two-step method is employed to efficiently extract multi-scale contextual information and subsequently fuse the feature maps generated by multi-scale features. Specifically, the previous single-step approach for acquiring multi-scale context information is decomposed into a two-step method to alleviate the complexity of acquisition.

Step I involves the generation of associated residual features from input features. In this step, a series of concise feature maps in the form of regions with different sizes are generated as material for morphological filtering in the second step. As illustrated in Figure 3, this is achieved by employing a 3x3 convolution for initial feature extraction, followed by a batch normalization (BN) layer and a ReLU layer.

Step II involves morphological filtering of features of different sizes of regions, 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 one is efficiently matched with the receptive field size in step two for rapid processing. To execute this step, the region feature maps are initially divided into groups, and then each group undergoes convolution with varying rates of expansion depth.

With the two-step region residualization-semantic residualization approach, the role of multirate depth-direction dilation convolution shifts from laboriously obtaining as much complex semantic information as possible to simply performing morphological filtering using a desired receptive field on each succinctly expressed feature map. Feature

mapping in the form of regions reduces the role of depth-direction dilation convolution to morphological filtering, thereby streamlining the learning process. Consequently, multiscale context information can be preserved more efficiently. After mapping the multi-scale context information, multiple outputs are aggregated. Specifically, all feature maps are concatenated, followed by batch normalization (BN) and point-by-point convolution to merge the features into the final residuals. Finally, these final residuals are added to the input feature maps to establish a more robust and comprehensive feature representation. 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.

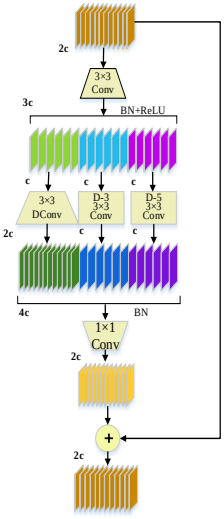


Figure 3. Structure of DWR module

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.

This adaptation is more fitting for the C2f module to learn the transformation of cross-scale feature points. Consequently, to enhance the multi-scale feature extraction capability of the C2f module, we opt to replace the regular convolution in the Bottleneck structure of the C2f architecture with a DWR module. The enhanced C2f-DWR structure is depicted in Figure 4.

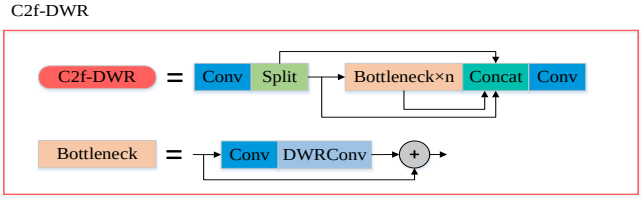


Figure 4. C2f-DWR module structure

2.4. C2f-DLKA module

Literature [35] created to a deformable large kernel attention (D-LKA Attention), illustrated in Figure 5. This mechanism is a simplified attention mechanism utilizing a large convolutional kernel, drawing primarily from the Large Kernel Attention (LKA) architecture [36] and deformable convolution [37], 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.

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. Firstly, 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. Secondly, 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.

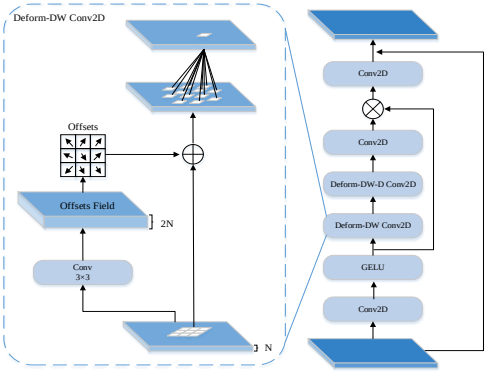


Figure 5. Structure of D-LKA Attention

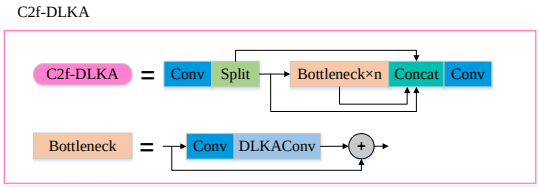


Figure 6. C2f-DLKA structure

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 unified perspective to address the detection problem comprehensively.

Literature [38] proposed a novel dynamic head framework that integrates the object detection head with attention mechanisms. This framework seamlessly combines multiple self-attention mechanisms across feature layers to enhance scale perception, between spatial locations for spatial perception, and within the output channel for task perception. This approach notably enhances the representation of the object detection head without introducing any additional computational overhead.

The Dynamic-head detection head employs the attention mechanism individually in each specific dimension of the feature, namely the horizontal, spatial, and channel directions. The scale-aware attention module is exclusively deployed on the horizontal dimension, where it discerns the relative importance of various semantic levels to enhance features at appropriate levels for individual objects based on their scale. The spatial-aware attention module operates on the spatial dimension (i.e., height × width), learning coherent discriminative representations at spatial locations. Meanwhile, the task-aware attention module is deployed on channels, directing different functional channels to support various tasks such as classification, box regression, and center/keypoint learning, respectively. This ensures diverse object-based convolutional kernel responses. Despite being applied separately to different dimensions of the feature tensor, these attention mechanisms effectively complement each other, resulting in the unified integration of the detection head and attention mechanisms.

The dynamic-head detection can be represented as:

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

Where, $\pi_L(\,)$, $\pi_S(\,)$, and $\pi_C(\,)$ 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.

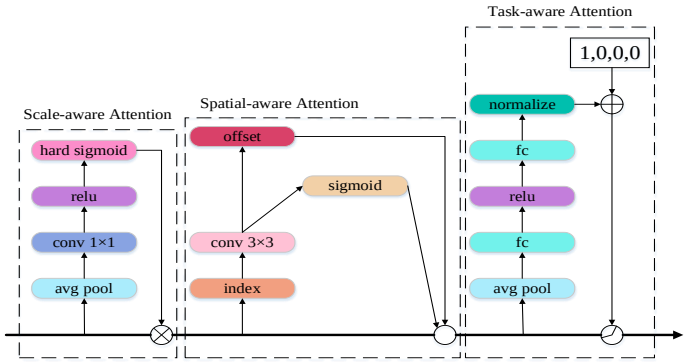


Figure 7. Structure of the dynamic-head detection head

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. In the YOLO family of algorithms, the downsampling multiples are typically larger, which poses difficulties in learning feature information for small targets with deeper feature maps. Additionally, the original structure of YOLO comprises three detection heads (P3, P4, P5), representing detection capabilities for small-scale, medium-scale, and large-scale objects, respectively. However, this structure often results in poorer detection accuracy for small targets due to limitations in effectively capturing their 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. Firstly, 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. Secondly, 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 [39]. 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 leads to improved detection accuracy, as the model becomes more adept at accurately identifying and localizing 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 the sampling grid to be flexibly adjusted, enabling the network to better fuse feature information and helping the detector head to capture 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 further enhance the model's capability of detecting small targets, we also added a dynamic detection head in 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 MPD loss function. This replacement aims to help the model achieve faster and more effective regression. In summary, the proposed method in our study is shown in Figure 8.

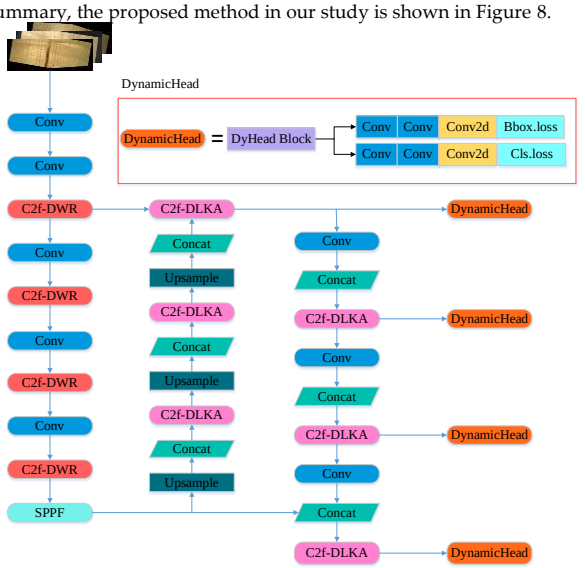


Figure 8. the improved YOLOv8 algorithm structure

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 [40] as its bounding box regression loss function, there are two main limitations associated with CIoU. Firstly, when the predicted bounding box and the

ground truth bounding box have the same aspect ratio but different width and height values, the effectiveness of the CIoU loss function diminishes. This limitation can impede the convergence speed and accuracy of the model [41]. Secondly, CIoU lacks the ability to adapt to different detectors and detection tasks, leading to a lack of robust generalization ability [42].

Inspired by the geometric features of horizontal rectangles, literature [41] proposes a new minimum point distance-based bracket box similarity comparison metric, MPDIoU. This metric encompasses all the relevant factors considered in existing loss functions, such as overlapping or non-overlapping areas, centroid distance, and width-height deviation, while simplifying the computation process. 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. Literature [42] proposes the Inner-IoU loss, which enhances the traditional IoU loss by calculating it through an auxiliary bounding box and introducing a scaling factor, ratio, to control the size of this auxiliary bounding box. 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, in order 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.

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 adapt to various shapes and sizes of defects. 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.

Algorithm1. Inner-IoU and MPDIoU as bounding box losses

Input: Predicted $\mathcal{B}_{prd}$ and ground truth $\mathcal{B}_{gt}$ bounding box coordinates. $\mathcal{B}_{prd} = (x_1^{prd}, y_1^{prd}, x_2^{prd}, y_2^{prd})$, $\mathcal{B}_{gt} = (x_1^{gt}, y_1^{gt}, x_2^{gt}, y_2^{gt})$ width and height of input image: *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. $x_c^{gt} = \frac{x_1^{gt} + x_2^{gt}}{2}, y_c^{gt} = \frac{y_1^{gt} + y_2^{gt}}{2}, y_c^{prd} = \frac{y_1^{prd} + y_2^{prd}}{2}, x_c^{prd} = \frac{x_1^{prd} + x_2^{prd}}{2},$
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. $\mathcal{L}_{IoU} = 1 - IoU, \mathcal{L}_{MPDIoU} = 1 - MPDIoU.$

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= 200, (3) Batch size = 4, (5) initial learning rate = 0.001, (6) weight decay coefficient: 0.0005, (7) intersection over Union (IoU) threshold for testing= 0.7. Two performance evaluation criteria: average precision (AP) and mean average accuracy (mAP) were 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(R) Core(TM) i5-9300 CPU 2.40 GHz
Graphics Processor	NVIDIA GeForce GTX 1650
Graphics Processor Accelerator Library	CUDA 11.7.0, CUDNN8500
Random Access Storage	8.0 GB
Deep Learning Environment	Pytorch 1.13.1
Deep Learning Frameworks	Python3.8.19

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: Replaces the regular convolution in the Bottleneck structure of all neck C2f modules with the C2f-DWR module.

2. YOLOv8-C2f-DLKA: Replaces the regular convolution in the Bottleneck structure of all neck C2f modules with the C2f-DLKA module.

3. YOLOv8-Dynamicchead: Changes the original three detection heads of YOLOv8 to four dynamic detection heads.

4. YOLOv8-InnerMPD: Replaces the original CIoU loss function in YOLOv8 with the InnerMPD loss function.

5. Ours: Incorporates all the above improvements into YOLOv8.

The ablation experiment results are shown in Table 3.

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

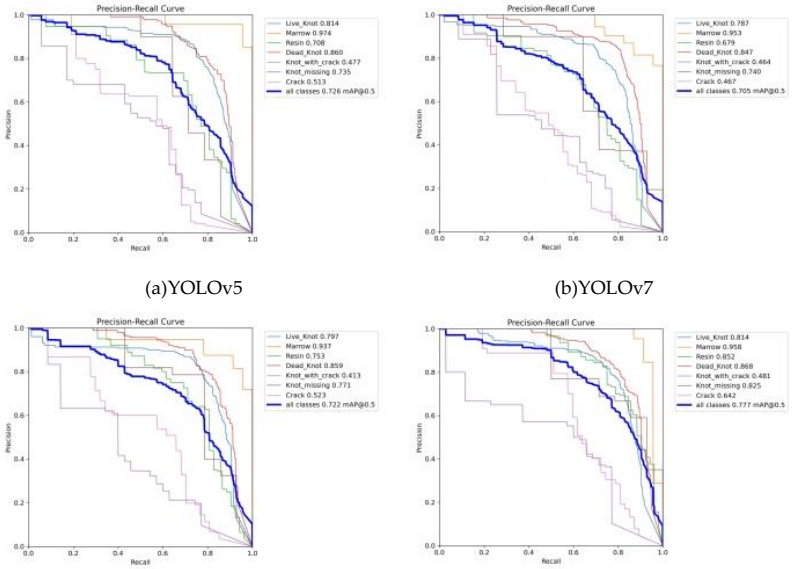
Note: AP = Average precision; mAP = Mean AP.

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. This enhancement is reflected not only in the overall recognition accuracy of our method for all types of defects but also in the mean Average Precision (mAP) values for specific defect types. Specifically, the mAP values for the detection of Resin, Knot_with_crack, Knot_missing, and Crack defects were improved by 0.099, 0.068, 0.054, and 0.119, 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.

3.2. Comparative experiments with other algorithms

The precision-recall (P-R) curves of YOLOv5, YOLOv7, YOLOv8, and Ours models for detecting seven defects are shown in Figure. 9. The P-R curve is an important measure of the performance of a model, where the larger the area under the curve, the better the performance.

From Figure 9, it can be seen that our proposed method demonstrates superior defect detection performance compared to the other three models. The comparison of the visualization results of the detection effects of these models is shown 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. Additionally, compared to the other models, our method significantly reduces false detections and missed detections.

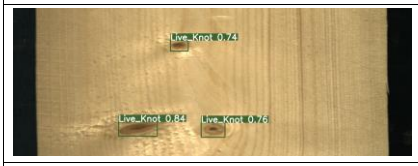
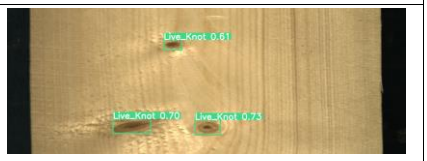
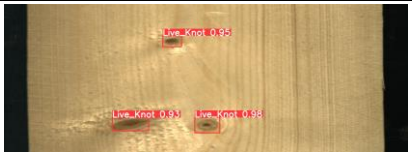


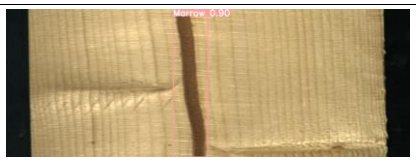
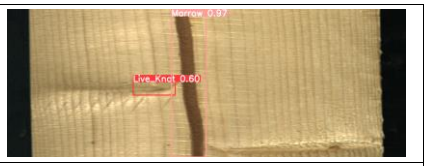
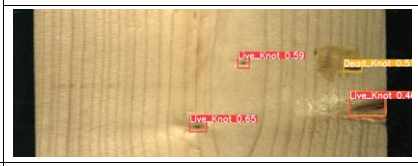
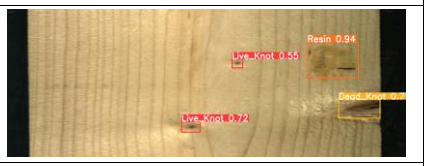
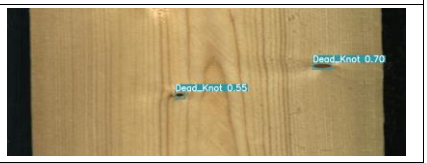
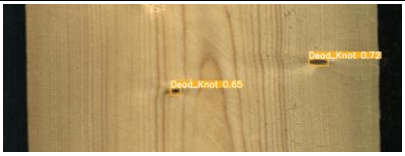
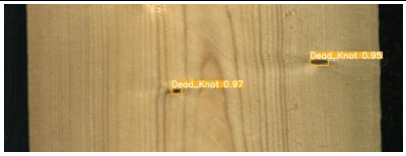
(a)YOLOv5 (b)YOLOv7

(c)YOLOv8 (d)Ours

Figure 9. Precision−recall (P−R) curves: (a) YOLOv5; (b) YOLOv7; (c) YOLOv8; (d) Ours.

458

Live_ Knot	YOLOv5	YOLOv7
		
	YOLOv8	DDD-YOLO
		
Mar- row	YOLOv5	YOLOv7
		
	YOLOv8	DDD-YOLO

		
Resin	YOLOv5	YOLOv7
		
	YOLOv8	DDD-YOLO
		
Dead_ Knot	YOLOv5	YOLOv7
		
	YOLOv8	DDD-YOLO
		
Knot_ with_C rack	YOLOv5	YOLOv7
		
	YOLOv8	DDD-YOLO
		
	YOLOv5	YOLOv7

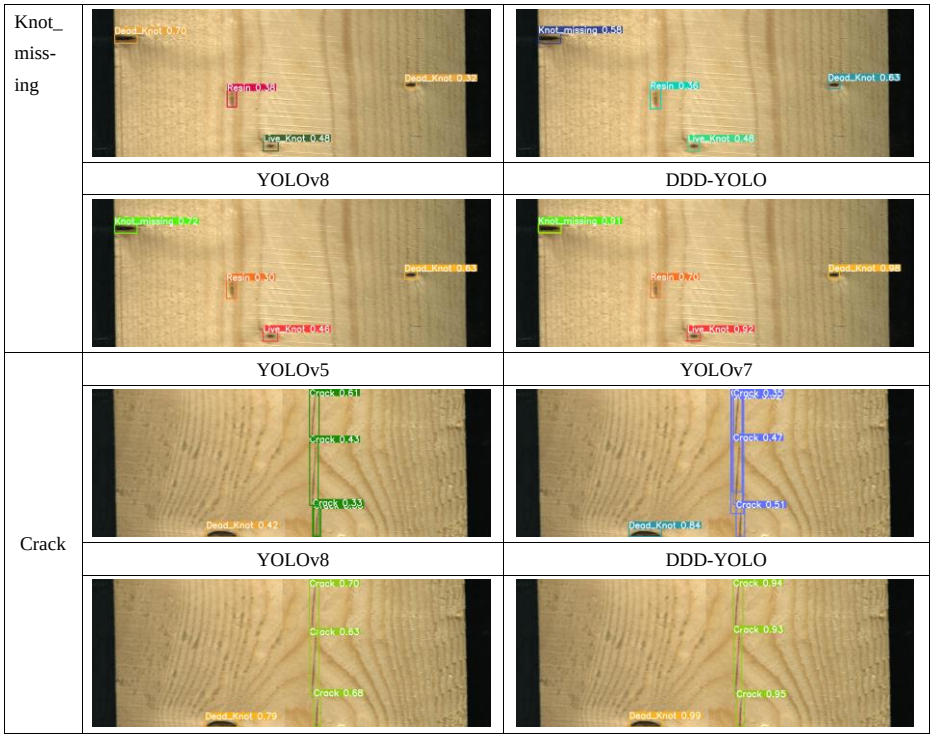
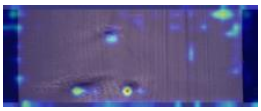
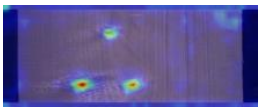
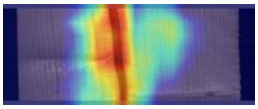
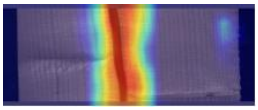


Figure 10. Comparison of detection results

Figure 11 presents the Gradient-weighted Class Activation Map (Grad-CAM) comparison results between YOLOv8, and Ours. It's 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.

	YOLOv8	Ours
Live_Knot		
Marrow		

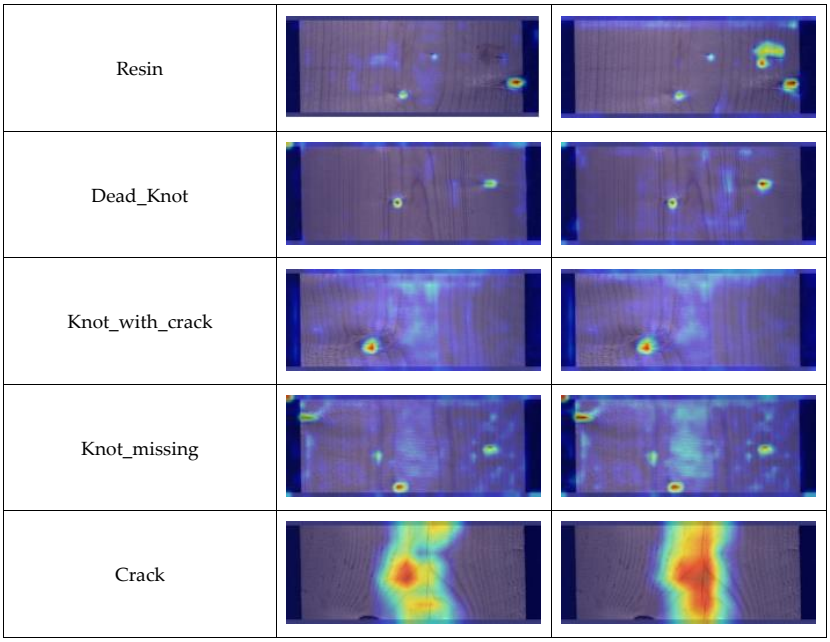


Figure 11. Grad-CAM comparison results

The experimental results, as presented in Table 4, provide further verification that the proposed algorithm achieves better detection accuracy compared to other mainstream defect algorithms, including SSD, YOLOv5, YOLOX, YOLOv7, YOLOv8, and RetinaNet. The comparative experimental results also demonstrate that our method achieves higher accuracy compared to other classical algorithms.

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, it enhances 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: Rijun WANG: Supervision, Writing-review & editing, Funding acquisition, Investigation, Methodology. Fulong LIANG: Writing-original draft, Software, Methodology. Bo WANG: Conceptualization, Validation, Supervision, Methodology, Data curation, Funding acquisition. Guanghao Zhang: Data curation, Conceptualization, Software. Yesheng Chen: Software, Data curation. Xiangwei MOU: Resources, Writing-review, Investigation, Funding acquisition.

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); 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. and Niemz, P. A brief overview on the development of wood research. Holzforschung 2022, 76, 102-119. https://doi.org/10.1515/hf-2021-0155.
2. Lian, M., Huang, Y., Liu, Y. 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 5, 1612–1657 (2022). https://doi.org/10.1007/s42114-022-00475-6.
3. Wang, R.; Zhang, G.; Liang, F.; Wang, B.; Mou, X.; Chen, Y.; Sun, P.; Wang, C. WPS-Dataset: A Benchmark for Wood Plate Segmentation in Bark Removal Processing. arXiv 2024,arXiv: 2404.11051.
4. 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. doi:10.15376/biores.18.1.Chen.
5. Meng, W.; Yuan, Y. SGN-YOLO: Detecting Wood Defects with Improved YOLOv5 Based on Semi-Global Network. Sensors 2023, 23, 8705, doi:10.3390/s23218705.
6. Wang, R.; Liang, F.; Wang, B.; Mou, X. ODCA-YOLO: An Omni-Dynamic Convolution Coordinate Attention-Based YOLO for Wood Defect Detection. Forests 2023, 14, 1885, doi:10.3390/f14091885.
7. 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, doi:10.15376/biores.15.1.1171-1186.
8. Du, X.; Li, J.; Feng, H.; Chen, S. Image Reconstruction of Internal Defects in Wood Based on Segmented Propagation Rays of Stress Waves. Applied Sciences 2018, 8, 1778, doi:10.3390/app8101778.
9. 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, doi:10.15376/biores.12.2.2711-2720.
10. 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, doi:10.1139/cjfr-2018-0537.
11. 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. International Journal of Architectural Heritage 2016, 10, 646–662, doi:10.1080/15583058.2015.1077905.
12. Wang, Q.; Liu, X.; Yang, S. Predicting Density and Moisture Content of Populus Xiangchengensis and Phyllostachys Edulis Using the X-Ray Computed Tomography Technique. Forest Products Journal 2020, 70, 193–199, doi:10.13073/FPJ-D-20-00001.
13. 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, doi:10.15376/biores.13.2.2834-2845.
14. 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. Annals of Forest Science 2020, 77, 68, doi:10.1007/s13595-020-00971-z.
15. 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, doi: 10.1109/ACCESS.2024.3395623.

16. 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(3), 5532-5556. doi: 10.15376/biores.17.3.Wang
17. 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, doi:10.1016/j.ijleo.2016.04.005.
18. 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, 2015; Vol. 9377, pp. 248–257 ISBN 978-3-319-25392-3. doi: 10.1007/978-3-319-25393-0_28
19. 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, doi:10.1007/s11676-017-0572-7.
20. Zhang, Y.; Liu, S.; Cao, J.; Li, C.; Yu, H. A Rapid, Automated Flaw Segmentation Method Using Morphological Reconstruction to Grade Wood Flooring. Journal of Forestry Research 2014, 25, 959–964, doi:10.1007/s11676-014-0543-1.
21. XU, Z.; JIA, P.; WU, N.; XU, K. Research on wood defect recognition technology. Transducer and Microsystem Technologies based on GA-SVM 2019, 38, 153-156, doi:10.13873 /J.1000-9787(2019) 09-0153-04.
22. Gu, I.Y.-H.; Andersson, H.; Vicen, R. Wood Defect Classification Based on Image Analysis and Support Vector Machines. Wood Sci Technol 2010, 44, 693–704, doi:10.1007/s00226-009-0287-9.
23. Wang, L.; Li, L.; Qi, W.; Yang, H. Pattern Recognition and Size Determination of Internal Wood Defects Based on Wavelet Neural Networks. Computers and Electronics in Agriculture 2009, 69, 142–148, doi:10.1016/j.compag.2009.07.019.
24. 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, doi:10.1007/s00226-015-0776-y.
25. Sun, P. Wood Quality Defect Detection Based on Deep Learning and Multicriteria Framework. Mathematical Problems in Engineering 2022, 2022, 1–9, doi:10.1155/2022/4878090.
26. 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, doi:10.15376/biores.18.4.8444-8457.
27. 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, doi:10.1109/ACCESS.2024.3352445.
28. 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, doi:10.1109/ACCESS.2023.3266737.
29. 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. https://doi.org/10.1007/s00226-021-01309-2
30. 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. doi:10.12688/f1000research.52903.2
31. Zhao, E.-X. Traffic Signals Recognition Based on YOLOv8. AIRR 2023, 12, 246–254, doi:10.12677/AIRR.2023.123028.
32. Haiying, L.; Xuehu, D.; Haonan, C.; Haitong, L.; Lixia, D. DBF-YOLO: UAV Small Targets Detection Based on Shallow Feature Fusion. IEEJ Transactions on Electrical and Electronic Engineering 2023, 18, 605-612,doi: 10.1002/tee.23758.
33. Haiying, L.; Fengqian, S.; Jason, G.; Lixia, D. SF-YOLOv5: A Lightweight Small Object Detection Algorithm Based on Improved Feature Fusion Mode. Sensors 2022, 22 , 5817. doi:10.3390/s22155817.
34. 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.
35. 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.
36. Guo, M.-H.; Lu, C.-Z.; Liu, Z.-N.; Cheng, M.-M.; Hu, S.-M. Visual Attention Network. arXiv 2022,arXiv: 2202.09741.
37. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable Convolutional Networks. arXiv 2017,arXiv: 1703.06211.
38. 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.
39. 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, doi:10.1109/TII.2019.2958826.
40. 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.
41. Siliang, M.; Yong, X. MPDIoU: A Loss for Efficient and Accurate Bounding Box Regression. arXiv 2023,arXiv: 2307.07662.
42. Zhang, H.; Xu, C.; Zhang, S. Inner-IoU: More Effective Intersection over Union Loss with Auxiliary Bounding Box. arXiv 2023,arXiv: 2311.02877.

See the guidelines

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.