

# WD-SEG: A Deep Learning Framework for Delicate and Accurate Wood Defect Segmentation

Junlin Qu,<sup>a,c</sup> Yan Pang,<sup>b,c</sup> and Zhongwei Wang<sup>b,c,\*</sup>

Precise segmentation of subtle wood defects is crucial for optimizing wood utilization and product value. Despite the prevalence of deep learning in wood defect detection, its deployment in real-world forestry environments is impeded by three primary challenges: 1: The limited capacity of traditional models to represent low-contrast, faint defect features; 2: feature ambiguity caused by complex background interference; and 3: entrapment in local optima because of insufficient global feature integration. To surmount these obstacles, this study proposes WD-SEG (Wood Defect Segmentation), a high-performance model tailored for complex forestry scenarios. The architecture integrates three core modules: an Augmented Feature Network (AFN) to mitigate spatial information loss; a Threshold Filtering Network (TFN), which leverages cosine similarity to adaptively suppress background noise; and a novel Interstellar Collision Optimization (ICO) algorithm to accelerate convergence and bypass local optima. Experimental evaluations on the wood defect training dataset demonstrate that WD-SEG outperforms state-of-the-art models, achieving an Intersection over Union (IoU) of 87.97% and an accuracy of 90.02%. Furthermore, generalization tests on wood defect datasets confirm the model's robustness, yielding an IoU of 86.50%. By introducing a novel "Enhance-Filter-Accelerate" framework, this study provides a precise, robust solution for automated wood quality inspection in resource-constrained environments.

DOI: 10.15376/biores.21.2.2925-2947

**Keywords:** *Wood surface defect segmentation; Enhance-filter-accelerate framework; Deep learning; Subtle defect features*

**Contact information:** a: College of Electronic Information and Physics, Central South University of Forestry and Technology, CS 410004 China; b: College of Economics and Management, Central South University of Forestry and Technology, CS 410004 China; c: Hunan Key Laboratory of Intelligent Logistics Technology, CS 410004 China; \*Corresponding author: t19900777@csuft.edu.cn

## INTRODUCTION

As a fundamental industrial material, wood is susceptible to surface and internal defects—such as cracks, knots, decay, and wormholes—during its growth and processing. These wood defects not only significantly reduce wood utilization rates and product value, but they also compromise its structural safety and service life (Chen *et al.* 2023). Consequently, achieving rapid and precise segmentation of wood defects is of critical importance for enhancing production quality and control efficiency in the forestry industry (Dou *et al.* 2025).

Early defect identification used manual inspection, which was inefficient and subjective. With the advent of machine vision, image processing-based detection methods have been progressively adopted. Traditional approaches typically employ color space conversions (*e.g.*, HSV, Lab) combined with threshold segmentation, or utilizing texture

features (e.g., GLCM, LBP) to construct classification models. In recent years, the rapid advancement of computer vision and deep learning has empowered Convolutional Neural Network (CNN)-based methods, such as U-Net (Ronneberger *et al.* 2015) and DeepLab (Chen *et al.* 2017), to demonstrate significant potential in automatically extracting features and localizing defect regions.

However, as a traditional industry, wood processing presents unique challenges for defect segmentation due to diverse material textures and variable environmental conditions. Faint surface defects, including knots, cracks, and pest damage, frequently blend seamlessly with natural grain patterns (Dong *et al.* 2025). Consequently, existing methods exhibit considerable limitations when detecting subtle defects (e.g., minute cracks, scratches) and low-contrast targets. Subtle defects are often characterized by fine textures, low contrast, and irregular shapes (Zhuo *et al.* 2022), while low-contrast defects—constrained by lighting, material properties, or imaging equipment—often exhibit blurred edges and high noise interference. These factors frequently lead to missed detections or false positives in traditional segmentation approaches (Dong *et al.* 2019).

Furthermore, existing deep learning-based methods often demonstrate limited generalization capabilities. Traditional CNNs possess a limited capacity to represent the features of low-contrast defects, making it difficult to capture targets with blurred edges or subtle grayscale variations. Moreover, complex background interference can cause feature confusion, degrading model performance in noisy environments. Most models are trained on single-species datasets, limiting their adaptability to multi-species scenarios and varying environmental conditions (e.g., differences in lighting and viewing angles). Additionally, the high computational resource consumption during training restricts their deployment in resource-constrained real-world scenarios (Chen *et al.* 2020). Therefore, balancing high accuracy with computational efficiency and robust generalization remains a core challenge in wood defect segmentation.

To address these challenges, this study proposes an innovative wood defect segmentation model, WD-SEG. By integrating mechanisms for feature enhancement, noise filtering, and intelligent optimization, the model significantly improves segmentation precision, generalization, and efficiency for subtle and low-contrast defects. The primary contributions of this study are as follows:

1. In this work, a hierarchical encoder-decoder structure is proposed for wood defect segmentation, named WD-SEG (Wood Defect Segmentation). This model employs an “Enhance-Filter-Accelerate” integrated framework to resolve the difficulty of accurately segmenting faint defects on wood surfaces against complex forestry backgrounds.
2. An Augmented Feature Network (AFN) is developed that leverages spatial and channel attention mechanisms to effectively reinforce the feature response of subtle defects characterized by blurred edges and minimal grayscale differences. This module demonstrates significant advantages in extracting features from subtle and low-contrast defects.
3. A Threshold Filtering Network (TFN) is developed based on directional consistency modeling and an adaptive binarization mask generation mechanism. This network exhibits superior selectivity in suppressing background noise and pseudo-responses, effectively filtering out irrelevant information.

4. The Interstellar Collision Optimization (ICO) algorithm is developed. It simulates celestial gravitational and collision mechanisms to achieve global optimization. This algorithm accelerates model convergence and reduces reliance on computational resources, overcoming the tendency of traditional gradient descent methods to become trapped in local optima.

## Related Work

### *Efficient attention methods*

Attention mechanisms have emerged as pivotal techniques for enhancing the performance of deep learning models. Within the domain of computer vision, particularly for dense prediction tasks such as image segmentation, the efficient capture of both local and global contextual information is of paramount importance. For instance, the CSWin Transformer (Dong *et al.* 2022) introduces cross-shaped window self-attention, which performs self-attention within local windows and incorporates horizontal and vertical window cross-branching. This design effectively balances local feature interaction with global information flow, thereby yielding improved model accuracy without a prohibitive increase in computational complexity. Regarding multi-scale feature fusion, the Channel and Spatial Attention Fusion Network (CSAFNet), proposed by Lei *et al.* (2022), leverages synergistic channel and spatial attention to bolster feature representations across disparate encoding streams, thereby refining detail-capturing capabilities. Furthermore, Bi-Level Routing Attention (Liu *et al.* 2022) employs a dual-level routing mechanism to achieve dynamic, content-aware allocation of computational resources, offering an efficient solution for high-resolution image processing. Collectively, these advancements provide a robust foundation for developing efficient attention modules tailored for tasks requiring meticulous local comprehension and broad global context, such as wood defect segmentation. From the perspective of engineering optimization, Flash Attention 2 (Dao 2023) significantly accelerates computational throughput and minimizes memory footprints by optimizing parallelization and tiling strategies for GPU memory access, all while preserving the original mathematical definition of attention. Such optimizations facilitate the deployment of sophisticated attention-based models in resource-constrained environments. Moreover, various approaches focused on mathematical reconstruction—such as Optimized Attention and Efficient Attention—further demonstrate the potential for enhancing the efficiency of standard attention mechanisms (Hosseini 2024).

### *Optimization algorithm*

In wood defect segmentation, the complex noise patterns often lead traditional gradient-based methods into local optima. The selection of an optimization algorithm governs the convergence rate, final performance, and generalization capability of a model (Chen *et al.* 2024). Within the realm of adaptive optimizers, AdamW (Loshchilov and Hutter 2017) mitigates the overfitting tendencies inherent in the original Adam optimizer by decoupling weight decay from gradient updates. This modification significantly bolsters generalization, establishing AdamW as the predominant choice for training large-scale architectures, such as Vision Transformers (ViTs).

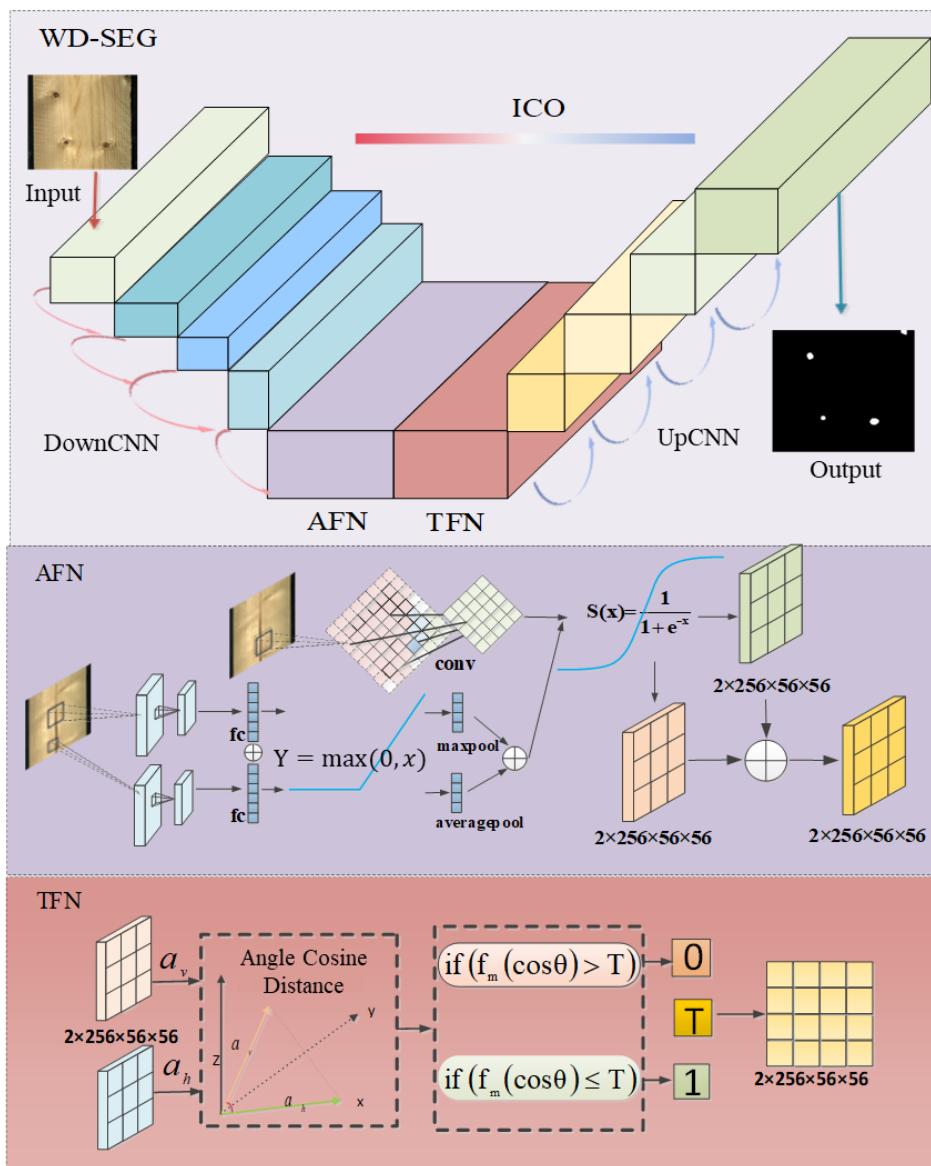
Recently, several efficient optimization paradigms tailored for large-scale training have emerged. Specifically, the Sophia optimizer (Liu *et al.* 2023) introduces a lightweight second-order approach that employs diagonal Hessian estimates as a preconditioner. By integrating an element-wise clipping mechanism to constrain update magnitudes in worst-case scenarios, Sophia demonstrates accelerated convergence over AdamW in language

modeling tasks. From a theoretical perspective, Cohen *et al.* (2021) characterized the “Edge of Stability” phenomenon, where gradient descent typically operates in a regime where loss surface sharpness adapts dynamically. This finding provides a novel lens through which the practical dynamics of modern neural network optimizers can be understood.

## EXPERIMENTAL

### Proposed Methods

The WD-SEG model developed in this study is designed to address critical limitations in existing wood defect segmentation methods, specifically their insufficient accuracy, limited sensitivity to subtle and low-contrast defects, and low computational efficiency.



**Fig. 1.** Overall framework diagram of WD-SEG wood defect segmentation model

The core philosophy of the proposed method involves the integration of novel feature extraction and enhancement mechanisms, adaptive strategies for filtering irrelevant information (noise), and an intelligent algorithm combining multi-dimensional search with global optimization. These components work synergistically to achieve precise segmentation and effective discrimination of defect regions, with a particular focus on subtle and low-contrast defects characterized by blurred edges, minimal grayscale variation, and fine texture changes. The WD-SEG architecture comprises three pivotal modules: the AFN, TFN, and ICO algorithm. The overall framework is illustrated in Fig. 1.

### Amplified Feature Network

In semantic segmentation, the encoder-decoder architecture uses convolution and down sampling to extract high-dimensional features. However, the inevitable loss of spatial detail during this process poses a significant challenge for identifying subtle wood defects, which requires high spatial sensitivity. Furthermore, standard convolution operations treat all feature channels uniformly, failing to account for the varying importance of information encoded across different channels. For low-contrast defects, faint signals are typically concentrated in only a few key channels. If these informative channels are not adaptively enhanced while irrelevant background features are suppressed, the defect signals risk being obscured, thereby compromising the model's overall generalization capability.

While existing attention mechanisms offer potential solutions to these issues, they are typically implemented as lightweight, add-on modules. In contrast, the AFN proposed in this study is designed as a robust, deeply integrated module. The core philosophy of the AFN is to synergistically perform non-linear transformations and self-calibration on features across both spatial and channel dimensions along the critical feature propagation path. This approach significantly amplifies defect-related information while suppressing noise and redundancy. It is worth noting that despite its robust design, the AFN's integration into the deeper layers ensures that its computational footprint remains minimal due to reduced spatial dimensions. This allows the model to prioritize high-fidelity feature extraction without significantly compromising real-time performance. Specifically, the AFN takes the feature tensor down-sampled from the fourth layer of the encoder as input, which is formulated as shown in Eq. 1:

$$X_{out} = F_{channel}(X; \Theta_{channel}) \oplus F_{spatial}(X; \Theta_{spatial}) \quad (1)$$

In Eq. 1,  $F_{channel}$  and  $F_{spatial}$  represent the channel and spatial feature extraction networks, respectively,  $\Theta$  denotes their parameters, and  $\oplus$  indicates the element-wise addition operation. Prior to the summation, the outputs of the two sub-networks are passed through an optional convolutional layer to adjust the channel dimensions, thus ensuring a consistent output. The network described above compensates for the spatial information loss caused by down-sampling by modeling dependencies in the spatial dimension, thereby capturing the global contextual information of the entire feature map. The detailed forward propagation process is as follows:

### *Dual pooling and embedding*

Global Average Pooling (GAP) and Global Max Pooling (GMP) are applied simultaneously to the input  $X$ , mapping it from a high-dimensional space to a highly abstract spatial descriptor. The operations can be formulated as Eqs. 2 and 3, respectively.

$$Z_{avg} = GAP(X) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X(i, j) \quad z_{avg} \in R^{C_{in}} \quad (2)$$

$$Z_{max} = GMP(X) = \max_{i=1:H, j=1:W} X(i, j) \quad z_{max} \in R^{C_{in}} \quad (3)$$

Here, GAP extracts the statistical mean of global features, which is sensitive to the overall context, while GMP captures the most salient local features, which is sensitive to potential subtle defect points. Their combination thus provides complementary spatial information.

#### *Non-linear dimensionality reduction and transformation*

Both descriptors are then passed through a shared-parameter Multi-Layer Perceptron (MLP). This MLP employs a bottleneck structure designed to introduce non-linearity and reduce computational complexity. Specifically, the first linear layer (parameterized by  $W_1$ ) reduces the dimension from  $C_{in}$  to  $C_{in}/r$ , and the second linear layer (parameterized by  $W_2$ ) subsequently restores it to  $C_{in}$ , as detailed in Eqs. 4 and 5.

$$S_{avg} = W_2 \cdot \delta(W_1 \cdot z_{avg} + b_1) + b_2 \quad (4)$$

$$S_{max} = W_2 \cdot \delta(W_1 \cdot z_{max} + b_1) + b_2 \quad (5)$$

In these equations,  $W_1$  and  $W_2$  are the weight matrices,  $b_1$  and  $b_2$  are the bias terms, and  $\delta$  denotes the ReLU activation function. The weights  $W_1$  and  $W_2$  enable the network to process information derived from the two distinct pooling operations with a consistent transformation.

#### *Spatial attention weight generation*

The two transformed feature vectors are summed, and a spatial importance weight vector  $S$  is generated through a Sigmoid activation function. Each element in corresponds to a comprehensive importance score for a specific channel across all spatial locations, as detailed in Eq. 6.

$$S = \sigma(S_{avg} + S_{max}) \quad (6)$$

Here,  $\sigma$  represents the Sigmoid function, which maps values to the interval  $(0, 1)$ . The generated spatial weight vector  $S$  is then multiplied with the original input feature map  $X$  in an element-wise, channel-wise manner, yielding the spatially enhanced feature  $X_{spatial}$ , as specified in Eq. 7.

$$X_{spatial} = X \otimes S \quad (7)$$

In this process, for the channel feature  $X_{spatial}$ ,  $X$  is adaptively scaled for each channel by combining it with the weight  $S$ , thereby enhancing the channels that contain important spatial contextual information. GAP and GMP are applied to compress spatial dimensions, resulting in two distinct channel descriptors, as detailed in Eqs. 8 and 9, respectively.

$$u_{avg} = GAP_{spatial}(X) \in R^{C_{in}} \quad (8)$$

$$u_{max} = GMP_{spatial}(X) \in R^{C_{in}} \quad (9)$$

Subsequently,  $u_{avg}$  and  $u_{max}$  are processed by another MLP (which shares the same bottleneck structure as the one in the spatial network but with independent, non-shared parameters), as detailed in Eqs. 10 and 11, respectively.

$$C_{avg} = W'_2 \cdot \delta(W'_1 \cdot u_{avg} + b'_1) + b'_2 \quad (10)$$

$$C_{max} = W'_2 \cdot \delta(W'_1 \cdot u_{max} + b'_1) + b'_2 \quad (11)$$

The outputs of the two MLPs are summed and then passed through a Sigmoid function to generate the channel attention weight vector  $c$ , as specified in Eq. 12.

$$C = \sigma(c_{avg} + c_{max}) \quad (12)$$

The channel weight vector  $C$  then multiplied with the original input  $X$ , yielding the channel-enhanced feature  $X_{channel}$ . This operation achieves adaptive selection across different feature channels, amplifying the key feature channels associated with defects. The process is detailed in Eq. 13.

$$X_{channel} = X \otimes C \quad (13)$$

The spatially enhanced feature  $X_{spatial}$  and the channel-enhanced feature  $X_{channel}$  are fused, as shown in Eq. 14.

$$X_{out} = X_{spatial} + X_{channel} \quad (14)$$

To allow for more flexible control over the fusion process, a learnable weighting parameter  $\alpha$  (initialized to 0) is introduced. This enables the network to adaptively determine from which sub-network to learn more. Subsequently, a small convolutional network is employed to automatically learn the fusion weights, as formulated in Eq. 15.

$$X_{out} = \alpha \cdot X_{spatial} + (1 - \alpha) \cdot X_{channel} \quad (15)$$

$X_{out}$  represents the final output of the AFN, which is a refined feature where both spatial and channel information are synergistically enhanced.

In summary, the AFN is integrated into the WD-SEG framework primarily to preserve critical defect feature information. Addressing the loss of spatial details caused by convolutional downsampling—which hinders the identification of dispersed wood defects—we incorporated the AFN as a dedicated branch.

### Threshold Filtering Network

AFN-enhanced feature maps are dense, and they contain irrelevant information, which can degrade performance during up sampling and increase computational cost. This issue is particularly critical when processing low-contrast defects, where faint signals are highly susceptible to noise interference. To address this challenge, the Threshold Filtering Network (TFN) is introduced in this work.

The core premise of the TFN is to formulate feature filtering as a binary decision problem. By employing an adaptive threshold learning mechanism, the TFN transforms continuous feature maps into sparse binary attention masks, explicitly distinguishing between “salient” and “secondary” feature regions. Functioning as a gating mechanism, this ensures that only significant features propagate through the network. The specific procedure is as follows:

(1) Feature tensor reshaping: The input feature tensor  $X_{\text{out}}$  is flattened along the spatial dimension to form a set of  $h$  feature vectors. Each vector encapsulates all channel information corresponding to a specific spatial coordinate  $(i, j)$ .

(2) Spatial gradient calculation: To capture the spatial variation trends of each feature point, the first-order differences (approximate gradients) in both horizontal and vertical directions were computed. Specifically, the horizontal feature vector  $G_h(i, j)$  is derived by calculating the difference between each position and its right-adjacent neighbor, as defined in Eq. 16:

$$G_h(i, j) = X_{\text{out}}(i, j+1) - X_{\text{out}}(i, j) \quad (16)$$

To preserve the dimensions, zero-padding is applied to the rightmost column. The vertical feature vector is then obtained by calculating the difference between each position and its neighbor below, as detailed in Eq. 17:

$$G_v(i, j) = X_{\text{out}}(i, j+1) - X_{\text{out}}(i, j) \quad \text{for } j = 1, \dots, H-1 \quad (17)$$

The gradient vectors,  $G_h(i, j)$  and  $G_v(i, j)$ , jointly describe the direction and magnitude of the fastest feature change at point  $(i, j)$ . The cosine value between the horizontal and vertical gradient vectors is computed, and its absolute value is used as a measure of feature consistency at that point, as specified in Eq. 18:

$$\cos \theta_{(i, j)} = \frac{|\langle G_h(i, j), G_v(i, j) \rangle|}{\|G_h(i, j)\|_2 \cdot \|G_v(i, j)\|_2 + \varepsilon} \quad (18)$$

In this formulation, the term  $\cos$  denotes the cosine similarity,  $\|*\|_2$  represents the  $L_2$  norm, and  $\varepsilon$  is a small constant introduced to prevent division-by-zero errors. A cosine value approaching 1 indicates that the gradient vectors are nearly parallel ( $0^\circ$  or  $180^\circ$ ).

This parallelism suggests consistent feature variation trends across orthogonal directions, implying a high probability that the corresponding location represents a salient defect feature. Conversely, a value approaching 0 implies that the gradient vectors are nearly orthogonal ( $90^\circ$ ). This directional inconsistency typically characterizes background noise or complex textures rather than structural defects.

Upon generating the importance score map, a thresholding operation is required to facilitate binary decision-making. While conventional methods often rely on manually preset hyper-parameters, this approach lacks flexibility. To address this, a learnable parameter is introduced, enabling the network to adaptively optimize the threshold level during training *via* global loss backpropagation. The final binary mask  $M$  is generated using a step function approximated by a Straight-Through Estimator (STE), as defined in Eq. 19:

$$M(i, j) = \mathbb{I}[S(i, j) > thr] = \begin{cases} 1 & \text{if } S(i, j) > thr \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

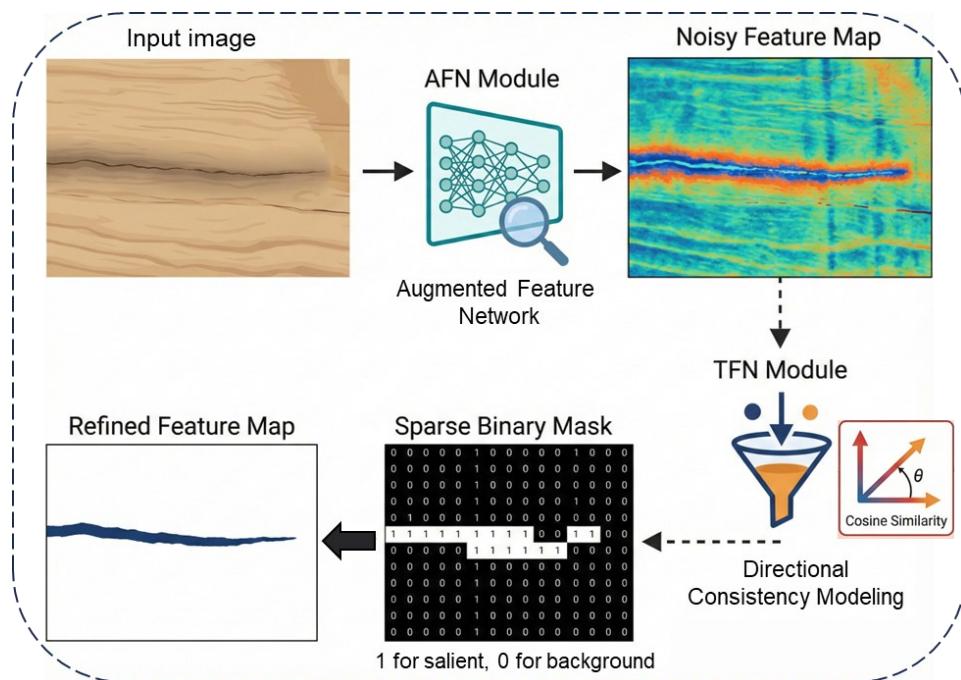
In this formulation,  $\mathbb{I}$  denotes the indicator function. During backpropagation, the Straight-Through Estimator (STE) propagates gradients directly from the output to the input layer, effectively bypassing the non-differentiability of the step function. The utilization of cosine similarity offers two primary architectural advantages. First, it enhances scale robustness; by normalizing features *via* the  $L_2$  norm, the model becomes invariant to feature magnitude, prioritizing feature direction over absolute amplitude.

In summary, the TFN is designed to refine the enhanced feature matrix by pruning task-irrelevant information inherited from the AFN, effectively categorizing features into

“salient” and “secondary” tiers. Operatively, the TFN maps the composite feature matrix to the spatial domain and evaluates the orthogonality of gradient vectors *via* an attention mechanism. A cosine similarity approaching zero (implying a 90° angle) indicates a lack of directional consensus, which is characteristic of background noise rather than structural defects. By imposing a learnable threshold  $thr$ , the network selectively enhances informative features while suppressing noise. Specifically, feature responses with similarity scores falling below  $thr$ —indicating low directional consistency—are suppressed (set to 0), while those exceeding  $thr$  are retained (set to 1).

To further elucidate the internal mechanism of the Threshold Filtering Network (TFN), the intermediate feature maps were visualized to demonstrate its pruning efficacy. As illustrated in Fig. 2, the feature map produced by the Augmented Feature Network (AFN) effectively captures the faint signals of subtle defects but inherently retains significant background interference, such as natural wood grain and uneven lighting.

Upon processing by the TFN, the model calculates the directional consistency of feature gradients and generates an adaptive binary attention mask. This mask explicitly distinguishes between “salient” defect regions and “secondary” background textures. By applying this sparse gating mechanism, task-irrelevant noise is effectively suppressed while the high-fidelity structural features of the defects are preserved.



**Fig. 2.** Visual interpretability of the AFN and TFN module

### Interstellar Collision Optimization Algorithm

Intelligent optimization algorithms have demonstrated remarkable efficacy in addressing complex combinatorial optimization problems. However, these methods often necessitate manual customization based on domain-specific expertise and lack transferability across different problem instances, resulting in computational inefficiency. Consequently, designing a tailored intelligent optimization algorithm specifically for wood defect recognition is of paramount importance.

The training of deep neural networks constitutes a high-dimensional, non-convex optimization challenge. Traditional gradient-based methods, such as Stochastic Gradient Descent (SGD) and Adam, are prone to entrapment in local optima and exhibit high sensitivity to hyperparameters, particularly the learning rate.

To overcome these limitations, the Interstellar Collision Optimization (ICO) algorithm is introduced here. Conceptualizing the solution space as a universe and candidate solutions as celestial bodies, ICO precisely simulates two fundamental interactions: gravitation and collision. This mechanism effectively equilibrates the trade-off between global search and local refinement, enabling the model to escape local optima and converge efficiently towards the global optimum. The fundamental principles of ICO are detailed below.

### (1) Quality Mapping Function

Assuming the objective is to minimize  $f(x)$ , this study employs the following quality mapping function, as shown in Eq. 20,

$$M_i(t) = \frac{1}{1 + f(X_i(t)) - f_{best}} \quad (20)$$

where  $X_i(t)$  denotes the position vector of the  $i$ -th celestial body at time  $t$ ,  $M_i(t)$  represents the mass of the  $i$ -th celestial body at time  $t$ , and  $f_{best}$  is the optimal function value found in the current population. This function ensures that the mass is always positive and that better solutions possess greater mass.

The gravitational force causes a celestial body  $i$  to be attracted by another body  $j$ , which in turn alters the velocity of body  $i$ . In the ICO algorithm, instead of computing the gravitational forces between all pairs of celestial bodies, each body is only attracted by the current best solution (best) and another randomly selected high-quality solution, as specified in Eqs. 21 and 22, respectively.

$$F_i^{best}(t) = G \cdot \frac{M_i(t) \cdot M_{best}(t)}{\|X_{best}(t) - X_i(t)\|^2 + \varepsilon} \cdot (X_{best}(t) - X_i(t)) \quad (21)$$

$$F_i^{rand}(t) = G \cdot \frac{M_i(t) \cdot M_k(t)}{\|X_k(t) - X_i(t)\|^2 + \varepsilon} \cdot (X_k(t) - X_i(t)) \quad (k \neq i, k \neq best) \quad (22)$$

In these equations,  $G$  is the gravitational constant, which decays as the number of iterations increases to enhance convergence in the later stages. The velocity update is determined by the acceleration caused by gravitational forces, as detailed in Eq. 23:

$$V_i(t + 1) = \omega \cdot V_i(t) + \frac{C_i}{M_i(t)} \cdot (F_i^{best}(t) + F_i^{rand}(t)) \quad (23)$$

Collision is the key mechanism for ICO to escape local optima. In this study, the collision condition is defined as follows: if the distance between two celestial bodies  $i$  and  $j$  is less than the collision radius  $r_{collide}$  and the difference in their fitness values is greater than a threshold  $\delta$ , a collision event is triggered, as specified in Eq. 24:

$$\|X_i(t) - X_j(t)\| < R_{collide} \text{ and } |f(X_i(t)) - f(X_j(t))| > \delta \quad (24)$$

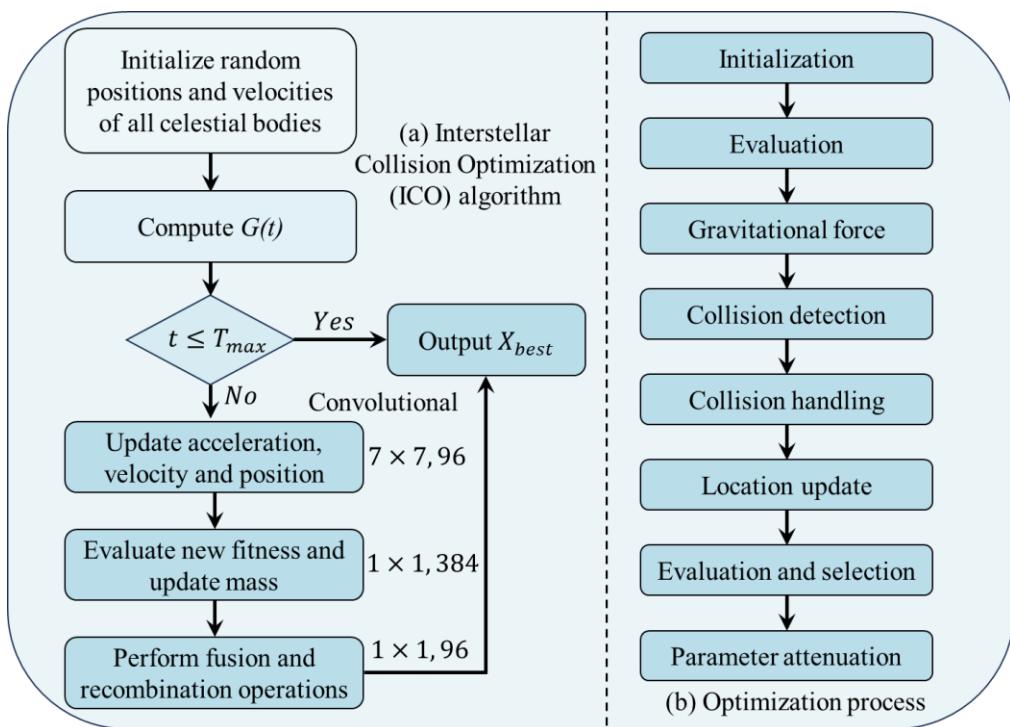
After a collision, the two celestial bodies do not annihilate but rather merge and recombine to produce one or more new offspring bodies, which replace the poorer solution

among the parents. For celestial bodies that do not collide, their positions are updated in the standard manner, as shown in Eq. 25:

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (25)$$

The new solutions generated from collisions are directly inserted into the population, typically replacing one or several of the current worst solutions in terms of mass. The flowchart of the Planetary Collision Optimization algorithm and its key algorithm are illustrated in Fig. 3.

In summary, the ICO algorithm is employed to enhance the training efficiency of the WD-SEG model and facilitate the search for global optima. Within the optimization process, the collision mechanism models stochastic perturbations during exploration. These events induce significant shifts in the solution space, effectively enabling the algorithm to escape entrapment in local optima.

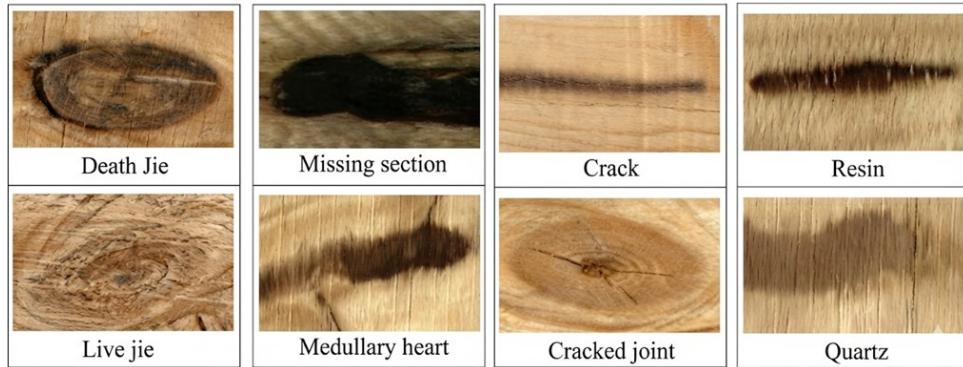


**Fig. 3.** (a) Interstellar Collision Optimization (ICO) algorithm iteration process. (b) Schematic diagram of the optimization process of Interstellar Collision Optimization algorithm

## RESULTS AND DISCUSSION

### Wood Defect Training Dataset

The Wood Defect Training Dataset (Kodytek *et al.* 2022) is derived from real-world wood imagery sourced from an open-source repository. As a subset of a large-scale wood surface defect database, it comprises over 8,000 images covering common defect categories. For the purpose of training the WD-SEG model, the dataset was randomly partitioned into training and validation sets at an 8:2 ratio. Representative examples of these defect categories are illustrated in Fig. 4.

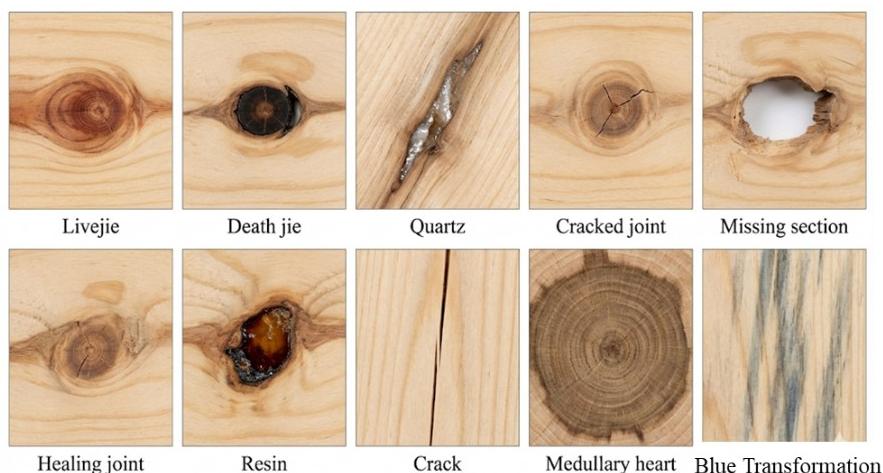


**Fig. 4.** Example of wood defect categories in the wood defect training dataset

Analysis of the dataset reveals a hierarchical structure in defect distribution, characterized by significant variations in prevalence. Missing Knots constitute the predominant category, accounting for 26% of the dataset. This prevalence is approximately 1.3 times that of the second-highest category, Live Knots (20%), and 6.5 times that of the rarest category, Insect Damage (4%). The second cluster is formed by Live Knots and Dead Knots (15%), which collectively represent 35% of the data. The third tier comprises Quartz (12%) and Cracks (10%); while the difference between them is only 2% points, the prevalence of Cracks is merely 38% of that of Missing Knots. Pith (8%) and Resin (5%) occupy the fourth tier, with a 3% differential, where Resin accounts for only 19% of the Missing Knot volume. Finally, Insect Damage (4%) occupies the lowest tier, representing only 15% of Missing Knots, 20% of Live Knots, and 27% of Dead Knots. This imbalance indicates that the dataset accurately reflects the natural distribution of wood defects in real-world environments.

## Wood Defect Dataset

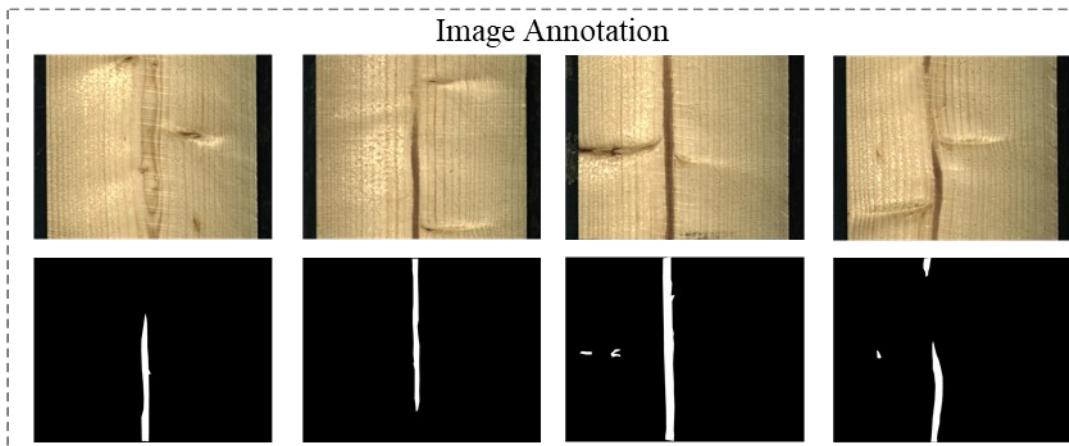
Similarly, the Wood Defect Dataset (Pavel *et al.* 2021), also derived from real-world open-source imagery, includes categories such as cracks, dead knots, live knots, and pith. Samples from this dataset are presented in Fig. 5.



**Fig. 5.** Example of wood defect categories in the wood defect dataset

## Data Processing

All images in the Wood Defect Training dataset were manually annotated using the Labelme software. Special attention was paid to complex defects that may encapsulate non-defect regions; to avoid mislabeling background textures as defects, images were annotated at high magnification to ensure meticulous contour tracing. Upon completion, the generated JSON annotation files were converted into binary masks (PNG format) to align with the model's input requirements, as illustrated in Fig. 6.



**Fig. 6.** Example of wood defects after LabelMe software

## Evaluation Criteria

To ensure a fair evaluation of the segmentation model's performance, this study employed the current mainstream evaluation metrics, including Intersection over Union (IoU), Precision, Recall, and Accuracy, to assess the model's efficacy.

The confusion matrix components presented are defined as follows:

- (1) True Positives (TP): Defective regions correctly classified as defective;
- (2) False Negatives (FN): Defective regions incorrectly classified as non-defective;
- (3) False Positives (FP): Non-defective regions incorrectly classified as defective;
- (4) True Negatives (TN): Non-defective regions correctly classified as non-defective.

Intersection over Union (IoU): Defined as the ratio of the intersection to the union of the predicted and ground truth sets, IoU serves as the primary metric for evaluating semantic segmentation performance. It quantifies the spatial overlap between the predicted mask and the ground truth, directly reflecting the model's capability to accurately delineate target boundaries. The calculation is defined in Eq. 26:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \quad (26)$$

Precision measures the proportion of pixels predicted as positive that truly belong to the target class. In binary segmentation, high precision indicates the model's reliability in suppressing false positives, ensuring that identified regions are relevant. The formula for Precision is given in Eq. 27:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (27)$$

Recall (or Sensitivity) calculates the proportion of actual positive pixels that are correctly identified by the model. This metric assesses the completeness of the segmentation. Recall is calculated as shown in Eq. 28:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (28)$$

Accuracy represents the ratio of correctly classified pixels (both defective and non-defective) to the total pixel count. It provides a global measure of the model's classification performance across the entire image. The formula is provided in Eq. 29:

$$\text{Accuracy} = \frac{\text{TP} + \text{FN}}{\text{TP} + \text{FN} + \text{FP} + \text{FN}} \quad (29)$$

### Implementation and Training Protocols

All experiments in this study were conducted on a Windows-based platform using the PyTorch framework. To ensure rigorous benchmarking across different methods and datasets, a unified training, testing, and evaluation framework was established. This was inspired by the architecture of the Segment Anything Model (SAM).

For hyper-parameters, the initial learning rate was set to  $1e^{-3}$  with a weight decay of 0.01. To guarantee fair comparison, WD-SEG utilized the same data processing pipeline as the baseline U-Net model. A polynomial learning rate decay schedule was employed to provide fine-grained control over learning rate adjustments, ensuring efficient and stable convergence. The training duration was set to 500 epochs with a global batch size of 4. To account for potential variations arising from stochastic initialization and to ensure experimental robustness, each experiment was conducted five times independently. All reported performance metrics represent the mean values accompanied by their respective standard deviations ( $\text{mean} \pm \text{std}$ ). Regarding input resolution, images from the Wood Defect dataset were resized to  $2800 \times 1024$  pixels. Detailed hardware and software configurations for the experimental environment are provided in Table 1.

**Table 1.** Configuration of Software and Hardware Used in the Experiment

Name	Category	Version Information
Processing Platform	GPU	NVIDIA GeForce RTX4060
Operating System	OS	Windows 11
Deep Learning Framework	Pytorch	1.9
Programming Language	Python	3.8
Code Compiler	IDE	Pycharm

Furthermore, the proposed ICO algorithm for model optimization were employed. Sharing underlying principles with AdamW, the ICO optimizer incorporates enhanced momentum acceleration and adaptive learning rate mechanisms. To validate the effectiveness of the proposed ICO algorithm, we compared the performance of ICO with SGD, Adam, and AdamW under strictly identical training parameters, including the same learning rate schedule, batch size, and number of epochs, using the Wood Defect dataset and with WD-SEG as the fixed backbone network. The experimental results are presented in Table 2. It can be seen that the ICO algorithm achieves the best performance across all evaluation metrics.

**Table 2.** Performance Comparison of Different Optimizers

Method	Recall (%)	IoU (%)	Precision (%)	Accuracy (%)
SGD	81.45 $\pm$ 0.21	80.12 $\pm$ 0.18	82.56 $\pm$ 0.24	83.10 $\pm$ 0.15
Adam	84.32 $\pm$ 0.16	84.50 $\pm$ 0.20	87.12 $\pm$ 0.22	87.55 $\pm$ 0.19
AdamW	85.10 $\pm$ 0.15	85.01 $\pm$ 0.17	88.22 $\pm$ 0.20	87.25 $\pm$ 0.16
ICO (Ours)	88.96 $\pm$ 0.12	87.97 $\pm$ 0.15	89.98 $\pm$ 0.10	90.02 $\pm$ 0.11

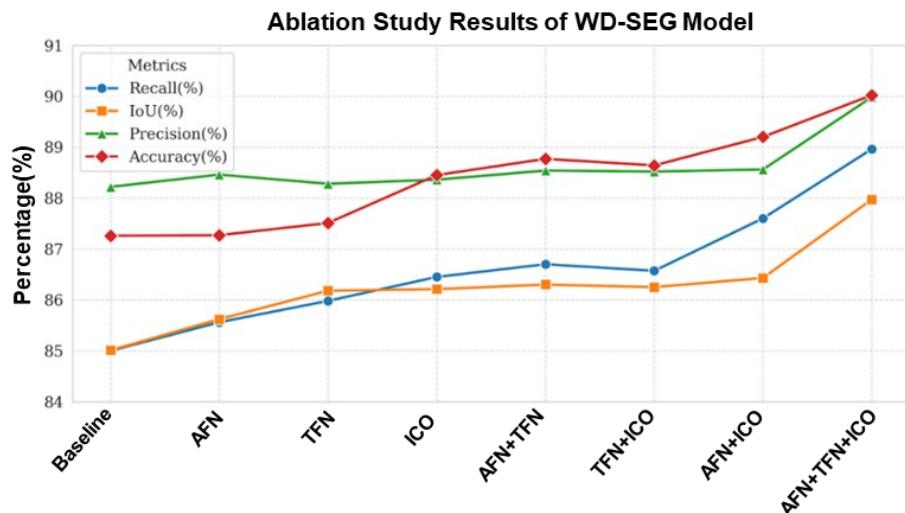
### Ablation Study

To quantify the individual and collective contributions of the proposed modules—namely the AFN, TFN, and ICO—a comprehensive ablation study on the Wood Defect Training dataset were conducted. The results for various component combinations are detailed in Table 3 and visualized in Fig. 7.

The study systematically elucidates the incremental performance gains attributed to each module within the WD-SEG task. The baseline model achieved a Precision of 88.22%. The integration of the AFN increased Precision to 88.46%, validating its efficacy in recovering fine-grained details within shallow feature layers. Conversely, deploying the TFN in isolation yielded a Precision of 88.28%. While superior to the baseline, this was 0.18% lower than the AFN-only configuration, suggesting that applying threshold filtering without prior feature enhancement risks suppressing valid defect information.

**Table 3.** Configuration of Software and Hardware Used in the Experiment

Method	AFN	TFN	ICO	Recall (%)	IoU (%)	Precision (%)
1	-	-	-	85.00	85.01	88.22
2	✓	-	-	85.56	85.62	88.46
3	-	✓	-	85.98	86.18	88.28
4	-	-	✓	86.45	86.21	88.36
5	✓	✓	-	86.70	86.30	88.54
6	-	✓	✓	86.57	86.25	88.52
7	✓	-	✓	87.60	86.43	88.56
8	✓	✓	✓	88.96	87.97	89.98

**Fig. 7.** Comparison of ablation experimental results of various components in WD-SEG model

However, the synergistic combination of AFN and TFN boosted Precision to 88.54% and Accuracy to 88.77%—which represent increases of 0.32% and 1.77% over the baseline, respectively. This substantiates the effectiveness of the “Enhance-then-Filter” strategy, which successfully eliminates redundant artifacts generated by the AFN while preserving high-fidelity defect features. Furthermore, incorporating the ICO algorithm independently raised Precision to 88.36% (+0.14%). This improvement indicates that the algorithm’s multi-trajectory parallel search mechanism smoothed initial loss parameter adjustments, effectively mitigating oscillations caused by random initialization. Consequently, this achieves a simultaneous enhancement of accuracy and stability while reducing training overhead.

Finally, the complete WD-SEG framework achieved optimal performance across all metrics, with Recall, IoU, Precision, and Accuracy reaching 88.96%, 87.97%, 89.98%, and 90.02%, respectively. These results conclusively demonstrate the high degree of complementarity between the proposed architectural enhancements and the intelligent optimization algorithm.

### Comparative Performance Analysis

To rigorously validate the efficacy and competitiveness of WD-SEG in wood defect segmentation, a comparative experiment on the Wood Defect Training dataset against a suite of representative state-of-the-art models were conducted. The selected baselines include U-Net++ (Zhou *et al.* 2019), Matting Anything (Li *et al.* 2024), SNUNet-CD (Fang *et al.* 2021), YOLOv11-seg (He *et al.* 2025), and FovealSeg (Yang *et al.* 2021). These models, widely adopted in industrial inspection, encompass diverse architectures ranging from standard encoder-decoders and attention-enhanced networks to specialized edge-aware designs.

To ensure fair comparison, all models were evaluated using identical data partitioning, preprocessing protocols, and training configurations. Performance was measured using a unified set of metrics. The comparative results are detailed in Table 4, with visual segmentation examples presented in Fig. 8.

**Table 4.** Experimental Results Comparing the Performance of Different Segmentation Models on the Wood Defect Training Dataset

Method	Recall (%)	IoU (%)	Precision (%)	Accuracy (%)
U-net++	86.76 $\pm$ 0.24	86.70 $\pm$ 0.18	86.65 $\pm$ 0.21	87.09 $\pm$ 0.16
Matting Anything	85.34 $\pm$ 0.10	85.59 $\pm$ 0.20	85.84 $\pm$ 0.19	86.10 $\pm$ 0.14
SNUNet-CD	82.17 $\pm$ 0.21	81.53 $\pm$ 0.17	81.09 $\pm$ 0.13	82.25 $\pm$ 0.20
YOLOv11-seg	87.38 $\pm$ 0.14	88.00 $\pm$ 0.22	88.63 $\pm$ 0.16	88.54 $\pm$ 0.15
FovealSeg	85.54 $\pm$ 0.18	85.38 $\pm$ 0.21	85.22 $\pm$ 0.12	85.19 $\pm$ 0.12
WD-SEG	88.96 $\pm$ 0.12	87.97 $\pm$ 0.15	89.98 $\pm$ 0.10	90.02 $\pm$ 0.11

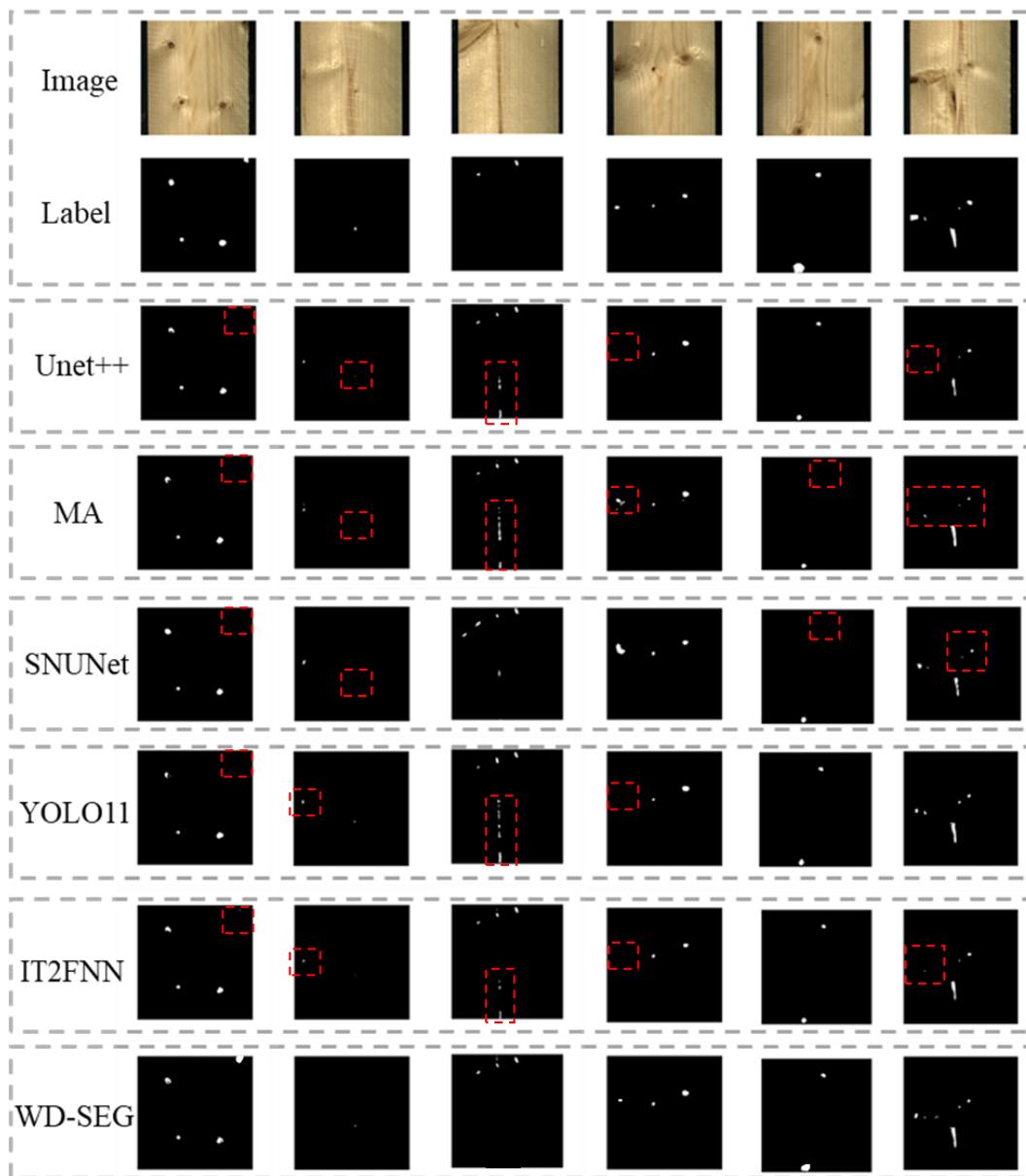
The results demonstrate that WD-SEG consistently outperforms existing classic and advanced models. Achieving a Recall of 88.96%, IoU of 87.97%, Precision of 89.98%, and Accuracy of 90.02%, WD-SEG surpassed all comparison methods across all metrics. These findings validate the effectiveness and universality of the proposed “Enhance-Filter-Accelerate” framework.

UNet++: While its dense nested skip connections yielded an accuracy of 87.09%, its Recall was limited to 86.76%, indicating insufficient sensitivity to faint or blurred defects. In contrast, WD-SEG’s AFN leverages dual-path (spatial and channel) attention to

amplify shallow features, significantly enhancing the detection of subtle defects and mitigating missed detections in low-contrast regions.

**Matting Anything:** Despite leveraging SAM priors for zero-shot capability and lightweight edge refinement, this model struggled with complex texture backgrounds, resulting in a Precision of 85.84% (4.14% lower than WD-SEG). This suggests susceptibility to background noise, leading to spurious contours.

**SNUNet-CD:** Originally designed for bi-temporal change detection, its weight-sharing structure lacks differential guidance in single-image tasks, leading to background overfitting and a low Accuracy of 82.25%. This highlights the limitations of direct architecture transfer.



**Fig. 8.** Examples of different segmentation models performing defect segmentation processing on the Wood defect training dataset. Existing methods struggle to correctly segment wood defect areas (marked in red dashed box).

YOLOv11-seg: Combining an efficient backbone with a lightweight segmentation head, this model achieved a respectable Accuracy of 88.54%. However, its Recall (87.38%) trailed WD-SEG by 1.58%, reflecting insufficient activation in low Signal-to-Noise Ratio (SNR) scenarios. WD-SEG addresses this *via* the AFN's layer-wise amplification, which fundamentally improves SNR and minimizes missed detections.

FovealSeg: While exhibiting balanced performance, its Accuracy was limited to 85.19%, suggesting that high-order fuzzy operations are ineffective at suppressing wood surface artifacts. WD-SEG's synergistic AFN-TFN mechanism effectively discriminates between signal and noise, achieving a superior equilibrium between Precision and Recall.

In conclusion, by integrating feature enhancement (AFN), intelligent filtering (TFN), and accelerated optimization (ICO), WD-SEG not only achieved a benchmark Accuracy of 90.02% but also demonstrated optimal comprehensive performance on the Pareto frontier, offering a novel paradigm for high-precision binary segmentation in complex environments.

Additionally, to evaluate the practical deployment potential of WD-SEG in resource-constrained forestry environments, a quantitative comparison of model complexity and inference efficiency was conducted. As summarized in Table 5, WD-SEG demonstrates a superior balance between segmentation accuracy and computational overhead compared to established benchmarks.

Specifically, WD-SEG achieves the highest accuracy of 90.02% with a significantly compact architecture, requiring only 18.4 M parameters and 42.6 G FLOPs. In contrast, U-Net++ exhibits nearly seven times the parameter count (36.6 M) and four times the FLOPs (150.4 G) due to its dense nested skip connections, which substantially restricts its real-time applicability. While YOLOv11-seg delivers the fastest inference speed (12.5ms) owing to its optimized detection-based backbone, it yields a lower accuracy (88.54%). The efficiency of WD-SEG is primarily attributed to the Threshold Filtering Network (TFN). By transforming continuous feature maps into sparse binary masks, the TFN effectively prunes task-irrelevant background information, thereby concentrating computational resources on salient defect regions without compromising edge fidelity.

**Table 5.** Quantitative Analysis of Model Complexity and Inference Efficiency

Method	Parameters (M)	FLOPs (G)	Inference Time (ms)	Accuracy (%)
U-Net++	36.6	150.4	45.2	87.09
YOLOv11-seg	12.6	31.2	12.5	88.54
WD-SEG	18.4	42.6	20.8	90.02

### Generalization Analysis on the Wood Defect Dataset

To further investigate the generalization capabilities of the proposed WD-SEG model, comparative experiments were conducted on the Wood Defect dataset. The results are detailed in Table 6 and Fig. 9.

The experiments on the Wood Defect dataset provide further validation of WD-SEG's robustness and adaptability. As presented in Table 8, WD-SEG outperformed all comparison models across the four key metrics, achieving a Recall of 86.77%, IoU of 86.50%, Precision of 86.54%, and Accuracy of 85.99%. These results demonstrate the model's consistent ability to identify defects within complex scenarios.

**Table 6.** Generalization Experimental Results of Different Segmentation Networks on the Wood Defect Dataset

Method	Recall (%)	IoU (%)	Precision (%)	Accuracy (%)
U-net++	83.46 $\pm$ 0.24	82.94 $\pm$ 0.28	82.78 $\pm$ 0.25	81.98 $\pm$ 0.20
Matting Anything	84.42 $\pm$ 0.20	83.54 $\pm$ 0.22	83.56 $\pm$ 0.18	82.30 $\pm$ 0.16
SNUNet-CD	85.29 $\pm$ 0.19	84.75 $\pm$ 0.17	84.80 $\pm$ 0.26	83.75 $\pm$ 0.21
YOLOv11-seg	85.99 $\pm$ 0.26	85.03 $\pm$ 0.23	85.12 $\pm$ 0.21	84.94 $\pm$ 0.19
FovealSeg	86.41 $\pm$ 0.17	86.02 $\pm$ 0.20	86.13 $\pm$ 0.13	85.48 $\pm$ 0.14
WD-SEG	86.77 $\pm$ 0.16	86.50 $\pm$ 0.18	86.54 $\pm$ 0.15	85.99 $\pm$ 0.13

**Generalization Experimental Results on Wood Defect Dataset****Fig. 9.** Comparison of generalization experiment results based on Wood defect dataset

WD-SEG's superior performance on this dataset is attributed to the synergistic integration of its core components: the AFN ensures sensitive capture of faint defects; the TFN effectively filters periodic textures and directional noise; and the ICO algorithm provides global optimization during training. The proposed "Enhance-Filter-Accelerate" mechanism demonstrates exceptional generalization when confronting typical interferences such as growth rings, resin spots, and uneven lighting, thereby confirming the model's reliability and potential for practical application.

## CONCLUSIONS

This study addressed critical barriers in automated wood quality inspection, specifically the insufficient segmentation accuracy for subtle, low-contrast defects and the limited generalization capabilities of existing models in complex forestry environments. To overcome these challenges, the WD-SEG model, a novel deep learning framework underpinned by an “Enhance-Filter-Accelerate” paradigm, was developed and evaluated.

1. The Augmented Feature Network (AFN) successfully mitigates the loss of spatial details inherent in deep networks. By synergizing spatial and channel attention mechanisms, AFN amplifies the feature response of faint defects—such as minute cracks and knots—that are typically obscured by blurred edges and minimal grayscale variations.
2. The Threshold Filtering Network (TFN) introduces a directional consistency modeling approach to refine feature maps. By utilizing cosine similarity and adaptive binarization, TFN effectively suppresses task-irrelevant background noise and periodic texture interference without compromising edge fidelity.
3. The integration of the Interstellar Collision Optimization (ICO) algorithm resolves training inefficiencies. By simulating gravitational attraction and collision mechanisms, ICO balances global exploration with local refinement, enabling the model to escape local optima and converge more rapidly than traditional gradient descent methods.
4. Quantitative evaluations confirmed the superior performance of the WD-SEG framework. On the Wood Defect Training dataset, WD-SEG achieved state-of-the-art results, with an IoU of 87.97%, significantly surpassing existing advanced baselines. Furthermore, generalization tests conducted on an independent validation dataset demonstrated the model’s strong robustness, yielding an IoU of 86.50%. These results validated the high precision and reliability of the framework even under challenging conditions, such as varying lighting and complex growth ring textures.

In summary, WD-SEG provides a precise, robust, and computationally efficient solution for wood defect segmentation. By effectively extracting faint defect features while maintaining low computational overhead, this study establishes a solid foundation for the deployment of automated inspection systems in resource-constrained real-world environments. Beyond its robust performance, the modular architecture of WD-SEG provides a technical foundation for selective defect attention. Through the adaptive thresholding mechanism in the TFN module, the system can be tuned to prioritize high-risk structural defects while filtering out task-irrelevant surface variations. Future research will prioritize model compression for deployment on edge-computing hardware and the expansion of the dataset to encompass a broader diversity of rare timber species. Furthermore, the proposed framework will be integrated into a comprehensive wood quality assessment system. By incorporating standardized grading protocols and market-oriented metrics—potentially utilizing the KANO model to categorize defect severity according to industrial requirements—this research aims to bridge the gap between high-precision computer vision and optimized industrial value recovery.

## ACKNOWLEDGMENTS

The authors are grateful for the support of the Opening Project Fund of Hunan Key Laboratory of Intelligent Logistics Technology, Grant No. RRI-KLOF202305.

## Author Contributions

J.Q.: Writing—original draft, Conceptualization, Methodology. Y.P.: Investigation, Validation, Data curation. Z.W.: Writing—review and editing, Funding acquisition, Supervision. All authors reviewed the results and approved the final version of the manuscript.

## Data Availability Statement

Data will be made available on request.

## Conflict of Interest

The authors declare that there is no conflict of interest with any party.

## REFERENCES CITED

Chen, S., Liu, J., Wang, P., Xu, C., Cai, S., and Chu, J. (2024). “Accelerated optimization in deep learning with a proportional-integral-derivative controller,” *Nature Communications* 15(1), article 10263. <https://doi.org/10.1038/s41467-024-54451-3>

Chen, Y., Sun, C., Ren, Z., and Na, B. (2023). “Review of the current state of application of wood defect recognition technology,” *BioResources* 18(1), 2288-2302. <https://doi.org/10.15376/biores.18.1.Chen>

Chen, B., Zhang, Z., Lin, J., Chen, Y., and Lu, G. (2020). “Two-stream collaborative network for multi-label chest X-ray Image classification with lung segmentation,” *Pattern Recognition Letters* 135, 221-227. <https://doi.org/10.1016/j.patrec.2020.04.016>

Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). “DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40(4), 834-848. <https://doi.org/10.1109/tpami.2017.2699184>

Cohen, J. M., Kaur, S., Li, Y., Kolter, J. Z., and Talwalkar, A. (2021). “Gradient descent on neural networks typically occurs at the edge of stability,” *arXiv preprint arXiv:2103.00065*.

Dao, T. (2023). “Flashattention-2: Faster attention with better parallelism and work partitioning,” *arXiv preprint arXiv:2307.08691*.

Dong, Y., Wang, J., Wang, Z., Zhang, X., Gao, Y., Sui, Q., and Jiang, P. (2019). “A Deep-learning-based multiple defect detection method for tunnel lining damages,” *IEEE Access* 7, 182643-182657. <https://doi.org/10.1109/access.2019.2931074>

Dong, X., Bao, J., Chen, D., Zhang, W., Yu, N., Yuan, L., Chen, D., and Guo, B. (2022). “CSWin Transformer: A General vision transformer backbone with cross-shaped windows,” in: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12114-12124. <https://doi.org/10.1109/cvpr52688.2022.01181>

Dong, Y., He, C., Xiang, X., Cui, Y., Kang, Y., Ding, A., Duo, H., and Wang, X. (2025). “IECAU-Net: A wood defects image segmentation network based on improved attention U-Net and attention mechanism,” *BioResources* 20(2), 3545-3556. . <https://doi.org/10.15376/biores.20.2.3545-3556>

Dou, W., and You, J. (2025). “A novel wood surface defect detection model based on improved YOLOv8,” *BioResources* 20(3), 5709-5730. <https://doi.org/10.15376/biores.20.3.5709-5730>

Fang, S., Li, K., Shao, J., and Li, Z. (2021). “SNUNET-CD: A densely connected Siamese network for change detection of VHR images,” *IEEE Geoscience and Remote Sensing Letters* 19, 1-5. <https://doi.org/10.1109/lgrs.2021.3056416>

He, L., Zhou, Y., Liu, L., Zhang, Y., and Ma, J. (2025). “Application of the YOLOv11-seg algorithm for AI-based landslide detection and recognition,” *Scientific Reports* 15(1), 12421. <https://doi.org/10.1038/s41598-025-95959>

Hosseini, M., and Hosseini, P. (2024). “You need to pay better attention: rethinking the mathematics of attention mechanism,” *arXiv preprint arXiv:2403.01643*

Kodytek, P., Bodzas, A., and Bilik, P. (2022). “A large-scale image dataset of wood surface defects for automated vision-based quality control processes,” [Dataset]. in: *F1000 Research* 2022, 10:581. <https://doi.org/https://doi.org/10.12688/f1000research.52903.2>

Lei, D., Ran, G., Zhang, L., and Li, W. (2022). “A spatiotemporal fusion method based on multiscale feature extraction and spatial channel attention mechanism,” *Remote Sensing* 14(3), article 461. <https://doi.org/10.3390/rs14030461>

Li, J., Jain, J., and Shi, H. (2024). “Matting anything,” in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.1775-1785. <https://doi.org/10.1109/cvprw63382.2024.00184>

Liu, H., Li, Z., Hall, D., Liang, P., and Ma, T. (2023). “Sophia: A scalable stochastic second-order optimizer for language model pre-training,” *arXiv preprint arXiv:2305.14342*

Liu, L., Qu, Z., Chen, Z., Tu, F., Ding, Y., and Xie, Y. (2022). “Dynamic sparse attention for scalable transformer acceleration,” *IEEE Transactions on Computers*, 1-14. <https://doi.org/10.1109/tc.2022.3208206>

Loshchilov, I., and Hutter, F. (2017). “Decoupled weight decay regularization,” *arXiv preprint arXiv:1711.05101*

Pavel, K., Alexandra, B., and Petr, B. (2021). “Supporting data for Deep Learning and Machine Vision based approaches for automated wood defect detection and quality control,” [Dataset]. in: *Zenodo* (CERN European Organization for Nuclear Research). <https://doi.org/10.5281/zenodo.4694695>

Ronneberger, O., Fischer, P., and Brox, T. (2015). “U-NET: Convolutional Networks for Biomedical Image Segmentation,” in: *Lecture Notes in Computer Science*, pp. 234-241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)

Yang, X., Li, S., Chen, Z., Chanussot, J., Jia, X., Zhang, B., Li, B., and Chen, P. (2021). “An attention-fused network for semantic segmentation of very-high-resolution remote sensing imagery,” in: *ISPRS Journal of Photogrammetry and Remote Sensing* 177, pp.238-262. <https://doi.org/https://doi.org/10.1016/j.isprsjprs.2021.05.004>

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2019). “UNET++: Redesigning Skip Connections to exploit multiscale features in image segmentation,” *IEEE Transactions on Medical Imaging* 39(6), 1856-1867. <https://doi.org/10.1109/tmi.2019.2959609>

Zhuo, X., Tian, J., and Fraundorfer, F. (2022). "Cross field-based segmentation and learning-based vectorization for rectangular windows," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 16, 431-448.  
<https://doi.org/10.1109/jstars.2022.3218767>

Article submitted: December 30, 2025; Peer review completed: January 24, 2026;  
Revised version received and accepted: January 29, 2026; Published: February 6, 2026.  
DOI: 10.15376/biores.21.2.2925-2947