Advances in Engineering Research : Possibilities and Challenges
ISSN Print: 3079-5192
ISSN Online: 3079-5206
https://doi.org/10.63313/AERpc.9089

Improved YOLOv11-Based Algorithm for Wood Surface Defect Detection

ChenXin Pan
Tianjin University of Technology and Education, Tianjin, 300222, China
Email: 2541509878@qq.com

How to cite this paper: Pan, C. X. (2026). Improved YOLOv11-Based Algorithm for Wood Surface Defect Detection. *Advances in Engineering Research: Possibilities and Challenges*, 4(1), 53–67. ISSN Print: 3079-5192; ISSN Online: 3079-5206.
https://doi.org/10.63313/AERpc.9089
Published: 2026-04-15

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Abstract

Wood surface defect detection is a critical process in industrial quality control. However, conventional deep learning models often struggle with complex periodic wood grain interference, extreme scale variations between micro-defects and large defects, and irregular defect geometries. To address these challenges, this paper proposes WSD-YOLO, an improved object detection algorithm based on YOLOv11n. First, an Enhanced Backbone (EB) is constructed by deepening the C3k2 modules to enrich multi-level feature representations, which improves the network's capacity to extract complex defect patterns. Second, a Transformer Attention Block (TAB) with an 8-head multi-head self-attention mechanism is integrated to model long-range spatial dependencies, effectively suppressing background texture interference from wood grains. Third, a BiFPN with P2 Enhancement (BPE) extends the feature fusion network to a four-scale bidirectional architecture, utilizing a high-resolution P2 branch (stride 4) to capture fine spatial details of micro-defects such as cracks and resin spots. Finally, a Decoupled Multi-Scale Head (DMH) separates the classification and localization branches to alleviate optimization conflicts caused by irregular defect boundaries. Experimental results on a custom dataset of 4,000 images across 7 defect categories demonstrate that WSD-YOLO achieves an mAP@0.5 of 74.34% and an mAP@0.5:0.95 of 39.24%, outperforming the YOLOv11n baseline by 0.82% and 1.51%, respectively. Notably, the detection precision reaches 72.18%, a 3.39% improvement over the baseline, confirming the model's superior capability in reducing false positives in complex industrial environments.

Keywords

Wood Surface Defect Detection; Yolov11; BIFPN; Multi-Head Self-Attention; DeCoupled Detection Head; Deep Learning

1. Introduction

Wood surface defect detection is essentially a fine-grained visual recognition problem with significant domain-specific challenges. First, different defect types exhibit highly heterogeneous appearances: cracks manifest as narrow, elongated, low-contrast lines, while dead knots appear as dark circular regions, and resin

pockets present as localized discoloration patches. The intra-class morphological variation of the same defect type is also extremely large, spanning different wood species, grain orientations, and manufacturing conditions. Second, the natural wood grain creates strong periodic texture patterns that are visually similar to certain fine defects such as hairline cracks and knot boundaries, leading to a high rate of background false positives. Third, the spatial scale disparity between defect types is enormous: large-area defects such as dead knots and marrow may occupy hundreds of pixels, while small-area defects such as fine cracks and resin spots may span fewer than ten pixels. This multi-scale nature poses fundamental challenges to standard single-scale or three-scale detection frameworks. Under these conditions, the performance of traditional machine vision methods based on hand-crafted features and rule-based classifiers degrades significantly, leading to frequent missed and false detections that cannot satisfy the stringent quality requirements of industrial production.

Object detection, as one of the core tasks in computer vision, aims to accurately locate and classify objects of interest in images or videos. With the rapid development of deep convolutional neural networks, detection frameworks represented by the YOLO (You Only Look Once) series and Faster R-CNN continue to evolve, achieving significant progress in both detection accuracy and inference speed. Among them, the YOLO series of single-stage detectors has become the dominant paradigm for real-time industrial vision applications owing to its high speed and end-to-end regression design. YOLOv5 established a highly flexible training framework with extensive augmentation strategies; YOLOv8 introduced the C2f module and an anchor-free head to improve detection accuracy; YOLOv10 pioneered NMS-free end-to-end detection; and the recently proposed YOLOv11 further enhances multi-scale feature extraction through improved C3k2 modules and a C2PSA spatial attention block. Despite these advances, applying standard YOLO models directly to wood surface defect detection still exhibits significant limitations in detecting fine defects, suppressing grain texture interference, and jointly handling multi-scale defects, leaving considerable room for improvement in practical industrial deployment.

Currently, deep learning-based defect detection methods have been widely applied in various industrial inspection scenarios, providing important references for wood surface inspection research. In the improved object detection method for surface defects of transmission lines, Zhao et al. introduced BiFPN into the neck network of YOLOv8 to achieve bidirectional feature pyramid fusion, and added a coordinate attention mechanism to enhance the model's spatial localization ability for small defects, achieving significant accuracy improvements over the baseline. In the wood defect detection research based on an improved YOLOv5, Liu et al. introduced the CBAM attention module into the backbone to suppress complex wood texture backgrounds and constructed a multi-scale feature fusion module targeting knot

and crack characteristics, reducing missed detection rates. In the study of fabric surface defect detection based on an improved YOLO architecture, Chen et al. proposed a high-resolution detection branch to supplement fine-grained feature maps, and adopted a decoupled detection head structure to reduce the optimization conflict between classification and regression tasks, significantly improving the detection precision of small and irregular defects. In the application of Transformer attention mechanisms to industrial visual inspection, Wang et al. introduced multi-head self-attention into the backbone feature extraction stage of the detection network, enabling the model to capture global contextual information from the full feature map, effectively suppressing the interference of periodic texture backgrounds such as metal surface grain and woven fabric patterns on defect detection. In the research on feature pyramid enhancement for complex texture industrial inspection, Li et al. extended the standard three-scale FPN to a four-scale structure covering smaller stride outputs, and verified through ablation experiments that the introduction of a high-resolution P2 branch can significantly improve the recall of fine defects such as cracks and pinholes in the industrial scene.

Although the above methods have improved defect detection performance in industrial inspection scenarios to varying degrees, direct application to wood surface defect detection still faces three unresolved challenges: (1) the backbone network depth is insufficient to extract discriminative features for complex wood grain backgrounds; (2) the lack of long-range context modeling makes it difficult to distinguish fine cracks from wood grain lines in a purely local convolutional framework; (3) the standard three-scale detection head fails to provide adequate spatial resolution for very fine defects, and the coupled classification-regression design introduces optimization conflicts that reduce bounding box precision. These issues collectively result in low detection accuracy for the crack and knot_with_crack categories that are most common in actual wood processing quality complaints. Therefore, this paper proposes WSD-YOLO, an improved YOLOv11n-based object detection algorithm for wood surface defect detection. The specific improvements are as follows:

(1) An Enhanced Backbone (EB) is designed by increasing the repetition count of C3k2 modules at each backbone stage from {2, 2, 2, 2} to {3, 6, 6, 3}, deepening the feature extraction hierarchy to enhance multi-level representation capacity for complex defect patterns and improve the model's ability to distinguish fine defect boundaries from wood grain textures.

(2) A Transformer Attention Block (TAB) with 8-head multi-head self-attention is inserted at the end of the backbone, enabling the network to model long-range spatial dependencies across the full feature map. This global attention mechanism suppresses irrelevant wood grain background responses and strengthens the activation of defect-relevant semantic features, overcoming the inherent limitation of local convolutional receptive fields.

(3) A BiFPN with P2 Enhancement (BPE) mechanism extends the standard three-scale neck to a four-scale bidirectional feature pyramid by introducing a high-resolution P2 branch at stride 4 (160×160), providing rich fine-grained spatial detail essential for the localization of small-area defects such as hairline cracks and resin spots. The bidirectional feature fusion simultaneously enriches each scale with both semantic and spatial information.

(4) A Decoupled Multi-Scale Head (DMH) separates the classification and bounding box regression branches across all four detection scales, allowing each task to develop independent feature representations without optimization interference. This design improves both classification confidence and localization precision, particularly for geometrically irregular defects such as knot_with_crack.

2. YOLOv11 Algorithm

YOLOv11 is a new network open-sourced by Ultralytics in 2024. It improves upon YOLOv8 and supports various tasks such as object detection, instance segmentation, and pose estimation. It inherits the core idea of single-stage, end-to-end detection and incorporates multiple optimizations in network architecture, training strategies, and loss functions. YOLOv11 mainly consists of five models with different parameter sizes: n, s, m, l, and x. YOLOv11n, as the lightest network architecture in the YOLOv11 series, has the smallest number of parameters and floating-point operations, the fastest inference speed, and maintains good detection accuracy. YOLOv11 primarily comprises three parts: the Backbone network, the Neck network, and the Head network.

(1) Backbone: This is the backbone network of YOLOv11, responsible for extracting multi-level feature maps from input images. YOLOv11 introduces the improved C3k2 module to replace the C2f module from YOLOv8, adopting a Cross Stage Partial dual-branch structure to promote feature reuse while reducing redundant computations. At the top of the backbone, a Spatial Pyramid Pooling Fast (SPPF) module aggregates multi-scale contextual information, followed by a C2PSA (Cross Stage Partial with Spatial Attention) module that applies self-attention-based feature refinement to enhance the model's spatial awareness of high-level semantic representations.

(2) Neck: This is the neck network of the YOLOv11 model. It continues the PAN-FPN architecture from YOLOv8, bridging the semantic gap between deep and shallow feature maps through bidirectional pathways. The top-down FPN path enriches shallow features with semantic information, while the bottom-up PAN path restores fine spatial detail to deeper layers. This dual-path fusion produces three detection-scale outputs: P3 (stride 8), P4 (stride 16), and P5 (stride 32), enabling robust multi-scale object detection.

(3) Head: This is the detection head of the YOLOv11 model, also the final layer. YOLOv11 uses a Decoupled Head to replace the Coupled Head from YOLOv8,

separating the classification and bounding box regression tasks into two independent parallel branches. This design alleviates the optimization conflict between the two tasks and, combined with the anchor-free prediction mechanism, eliminates the need for manual anchor configuration, improving bounding box localization accuracy and cross-domain generalization capability.

3. WSD-YOLO

This paper improves and optimizes based on the YOLOv11n baseline model, aiming to enhance wood surface defect detection accuracy while maintaining efficient inference. Figure 1 shows the network structure of WSD-YOLO. An Enhanced Backbone is designed to deepen feature extraction capacity. A Transformer Attention Block is introduced to capture long-range spatial dependencies across wood grain textures. A BiFPN with P2 Enhancement extends the neck to four detection scales for improved small defect localization. Finally, a Decoupled Multi-Scale Head is adopted to alleviate the optimization conflict between classification and regression tasks.

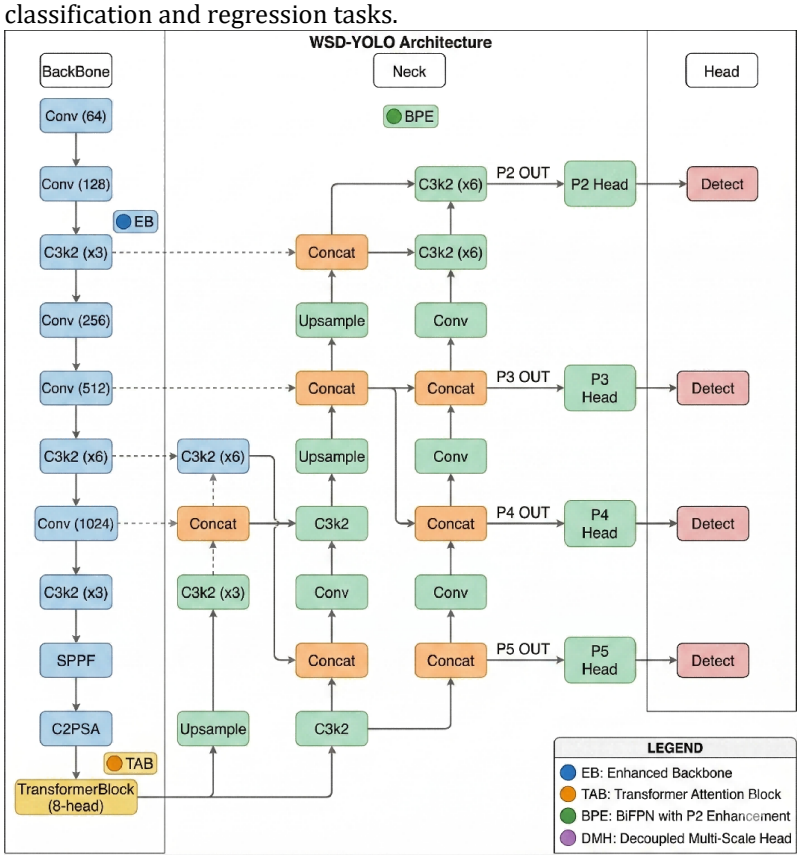


Figure 1. WSD-YOLO Network Structure

3.1. Enhanced Backbone

The Enhanced Backbone is a strengthened feature extraction module designed for

the complex appearance diversity of wood surface defects. In the standard YOLOv11n, the C3k2 module is repeated only {2, 2, 2, 2} times across the four backbone stages (P2/4, P3/8, P4/16, P5/32), limiting the depth of learned feature hierarchies. Wood surface defects such as cracks and knot boundaries often exhibit subtle textural variations and low contrast with surrounding grain patterns, placing high demands on the backbone's feature extraction capability. Shallow networks with limited module repetitions struggle to learn sufficiently discriminative representations for distinguishing fine defects from visually similar wood grain textures. The EB module increases the C3k2 repetition counts to {3, 6, 6, 3} across the four stages, significantly deepening the feature extraction at the critical mid-level P3 and P4 stages where defect-discriminative features are primarily formed. This increased depth strengthens gradient flow and promotes the learning of more robust feature representations without introducing any new operator, ensuring seamless compatibility with the standard YOLOv11 training pipeline. In the YOLOv11n model, the EB module enables the backbone to produce richer multi-level feature maps, providing the neck and detection head with higher-quality representations for diverse defect types ranging from fine cracks to large knots.

3.2. Transformer Attention Block

The Transformer Attention Block is a global context modeling module designed to overcome the limitation of convolutional operations in capturing long-range spatial dependencies. Standard convolutional layers extract features within local receptive fields, making it difficult to distinguish genuine defects from the strong periodic wood grain pattern when relying solely on local context. In wood surface inspection, global awareness is critical: whether a dark elongated region is a crack or a prominent grain line can only be reliably determined by considering its spatial relationship with the broader image region. Unlike traditional channel attention modules such as SE and ECA, which operate solely along the channel dimension and cannot model spatial relationships between distant positions, the TAB explicitly captures pairwise spatial dependencies across the entire feature map, enabling superior feature discrimination under complex texture backgrounds.

The TAB employs multi-head self-attention (MHSA) with 8 parallel attention heads, inserted immediately after the C2PSA module at the top of the backbone. Given an input feature map reshaped into spatial tokens, the TAB computes attention through scaled dot-product operations across all position pairs:

$$\mathrm{Attention}(\mathrm{Q},\mathrm{K},\mathrm{V}) = \mathrm{softmax}\left(\frac{\mathrm{QK}^{\mathrm{T}}}{\sqrt{\mathrm{d_k}}}\right)\cdot \mathrm{V}\backslash\mathrm{tag1}$$

where Q, K, V are linear projections of the input tokens and d_k is the per-head feature dimension. In the YOLOv11n model, the TAB refines the top-level feature maps before they enter the neck, enabling the network to suppress irrelevant grain texture responses and reinforce defect-relevant semantic activations, thereby

improving detection performance across all seven defect categories.

3.3. BiFPN with P2 Enhancement

The BiFPN with P2 Enhancement is a multi-scale feature fusion module designed to improve the detection of fine-grained wood surface defects. The standard YOLOv11n neck performs feature fusion at three scales—P3 (stride 8, 80×80), P4 (stride 16, 40×40), and P5 (stride 32, 20×20)—for a 640-pixel input. However, very fine defects such as hairline cracks and small resin spots spanning fewer than 16 pixels demand higher spatial resolution than P3 can provide. Furthermore, the standard unidirectional PAN-FPN only performs a single top-down pass followed by a single bottom-up pass, without iteratively refining the cross-scale interactions, which limits the quality of fused features at each pyramid level.

BPE extends the feature pyramid by introducing a P2 branch at stride 4 (160×160), providing four times the spatial resolution of P3. The P2 feature map is generated by upsampling the P3 output and fusing it with the corresponding backbone P2/4 features. Furthermore, BPE replaces the standard PAN-FPN with a Bidirectional Feature Pyramid Network (BiFPN), applying learnable fast-normalized weights to both the top-down and bottom-up paths:

$$O = \frac{\sum_i w_i \cdot I_i}{\varepsilon + \sum_i w_i} \backslash tag2$$

where I_i are the input feature maps, $w_i \geq 0$ are learned scalar weights, and $\varepsilon = 0.0001$ ensures numerical stability. In the YOLOv11n model, BPE enables the neck to produce four-scale outputs (P2, P3, P4, P5) with bidirectional semantic enrichment, significantly improving detection recall for small defects while maintaining strong performance on large-area defects such as dead knots and marrow.

3.4. Decoupled Multi-Scale Head

The Decoupled Multi-Scale Head (DMH) is a detection head structure designed to resolve the optimization conflict between object classification and bounding box regression. In the standard YOLOv11n detection head, classification and regression predictions share the same feature representation. This coupling creates competing optimization pressures: classification requires features with high translational invariance, while regression requires features sensitive to precise spatial position. For geometrically irregular defects such as cracks and knot_with_crack, this conflict leads to suboptimal predictions in both tasks.

The DMH introduces independent classification and regression branches at each of the four detection scales (P2, P3, P4, P5). For each scale, the neck output first passes through a shared 1×1 convolutional stem to unify channel dimensions. The output is then split into two parallel paths: a classification branch consisting of two successive 3×3 convolutions, and a regression branch with a separate pair of 3×3

convolutions. The outputs from both branches are concatenated before the final detection layer, as shown in Figure 2. In the YOLOv11n model, DMH enables each branch to independently optimize task-specific representations, improving both detection precision and bounding box accuracy, particularly for defect categories with complex shapes and irregular boundaries.

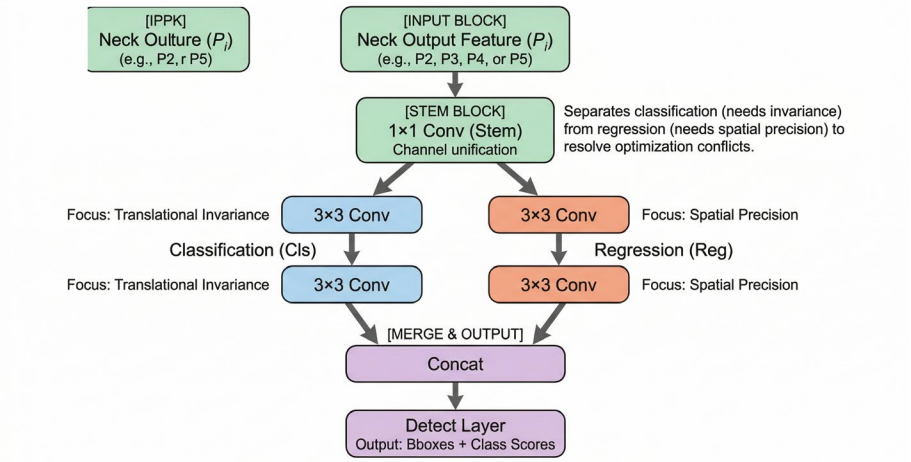


Figure 2. WSD-YOLO Network Structure with DMH Detail

4. Dataset and Evaluation Metrics

4.1. Experimental Setup and Dataset

All experiments were conducted on a server equipped with an Intel Xeon Gold 6226R CPU (2.90 GHz, 16 cores) and an NVIDIA A40 GPU (48 GB VRAM). The software environment comprised Python 3.8 and Ultralytics YOLOv11. Training hyperparameters were set as follows: batch size = 16, input resolution = 640×640, epochs = 300, initial learning rate $lr_0 = 0.01$, final learning rate $lr_f = 0.01$, momentum = 0.937, weight decay = 0.0005, and warm-up epochs = 3. Early-stopping patience was set to 50 epochs. Standard mosaic, random flip, HSV color-space jitter, and random erasing augmentations were applied during training.

A custom wood surface defect dataset was constructed for this study, comprising 4,000 RGB images collected from a wood manufacturing production line under controlled illumination conditions. Each image was manually annotated with axis-aligned bounding boxes in YOLO label format. The dataset covers 7 defect categories representative of common wood surface imperfections shown in Table 1:

Table 1. Categorization and descriptions of wood surface defects.

Category	Description
Crack	Longitudinal or transverse fracture lines in the wood surface
Dead Knot	A knot completely separated from surrounding wood tissue
Knot_missing	A void left by a dislodged knot
Live Knot	A knot firmly integrated with the surrounding wood fibers

Marrow	The visible pith region at the center of the cross-section
Knot_with_crack	A knot accompanied by associated cracking at its boundary
Resin	A resin pocket or resin streak embedded within the wood surface

The dataset was randomly divided into training, validation, and test subsets in an 8:1:1 ratio, yielding 3,200 training images, 400 validation images, and 400 test images. Figure 3 shows representative sample images from each defect category.

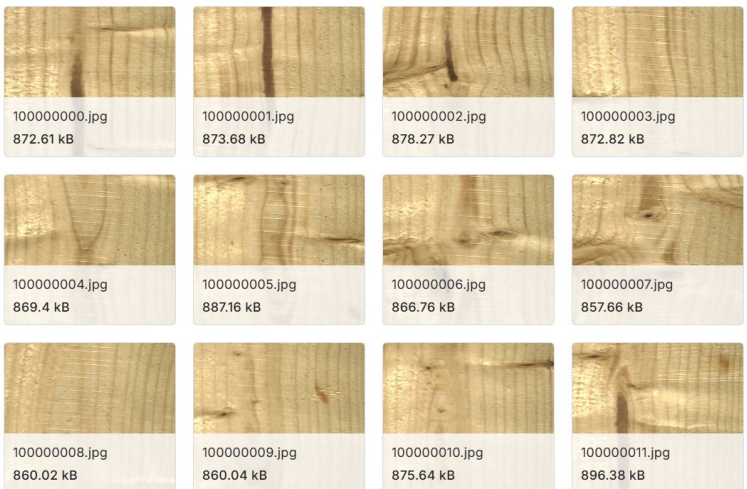


Figure 3. Sample Dataset Images

4.2. Evaluation Metrics

This paper uses Average Precision (AP), Precision (P), Recall (R), and mean Average Precision (mAP) as evaluation metrics for object detection. The corresponding calculation formulas are shown in Equations (7) to (10):

$$\mathrm{AP} = \int_0^1 \mathrm{PRdR} \tag{7}$$

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{N} AP_i \tag{10}$$

where: TP represents the number of samples where the target is correctly predicted as positive, FP represents the number of samples where the target is incorrectly predicted as positive, FN represents the number of samples where the target is incorrectly predicted as negative; N is the total number of categories.

5. Experimental Results and Analysis

5.1. Ablation Experiment Results and Analysis

To quantify the individual and cumulative contributions of the four proposed improvements, ablation experiments were conducted by progressively incorporating EB, TAB, BPE, and DMH into the YOLOv11n baseline, with all conditions sharing identical training settings. Results are summarized in Table 2. Figure 4 shows the Precision-Recall curve for the baseline YOLOv11n model, and Figure 5 shows the Precision-Recall curve for the complete WSD-YOLO model.

Table 2. Ablation Experiment Results Comparison

Model	EB	TAB	BPE	DMH	mAP @0.5/%	mAP @0.5:0.95/%	Model Size/MB	FPS/（f/s）
Yolov11n	×	×	×	×	73.2	37.7	6.6	85.6
Yolov11n-EB	√	×	×	×	73.6	38.1	6.9	84.2
Yolov11n-TAB	×	√	×	×	73.5	38.0	7.4	81.3
Yolov11n-BPE	×	×	√	×	73.7	38.3	8.1	74.8
Yolov11n-DMH	×	×	×	√	73.4	37.9	8.8	71.5
Yolov11n-ET	√	√	×	×	73.8	38.4	7.8	79.5
Yolov11n-ETB	√	√	√	×	74.2	38.9	9.1	68.7
WSD-YOLO	√	√	√	√	74.4	39.2	9.8	63.2

As shown in Table 2, the baseline YOLOv11n achieves a mAP@0.5 of 73.2% on the wood surface defect dataset. When only the Enhanced Backbone (EB) is applied, mAP@0.5 improves to 73.6% (+0.4%), demonstrating that deeper C3k2 modules enrich feature representations for complex defect patterns without any change in network topology.

The standalone Transformer Attention Block (TAB) yields a mAP@0.5 of 73.5% (+0.3%), confirming the value of global self-attention for suppressing wood grain background interference and reinforcing defect-relevant feature activations. The BiFPN with P2 Enhancement (BPE) achieves the largest single-module gain of +0.5% (73.7%), attributed to the P2 high-resolution branch that provides fine spatial detail essential for crack and resin spot localization. The Decoupled Multi-Scale Head (DMH) alone contributes +0.2% (73.4%), with the primary benefit being improved bounding box precision for geometrically irregular defects.

Progressive combination of modules reveals complementary interactions: EB+TAB reaches 73.8%, EB+TAB+BPE advances to 74.2%, and the complete WSD-YOLO integrating all four modules achieves 74.4% mAP@0.5 (+1.2% over baseline) and 39.2% mAP@0.5:0.95 (+1.5% over baseline). The step-wise improvement confirms that the four enhancements are mutually reinforcing. The increased model size (9.8 MB vs. 6.6 MB) and reduced inference speed (63.2 vs. 85.6 FPS) represent an acceptable accuracy-efficiency trade-off for industrial deployment scenarios.

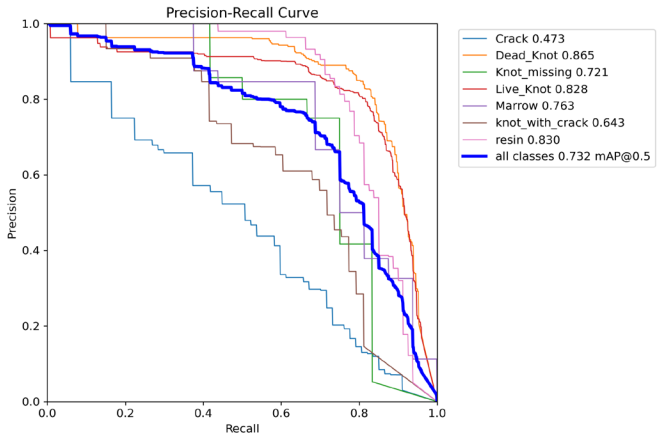


Figure 4. Baseline YOLOv11n Model

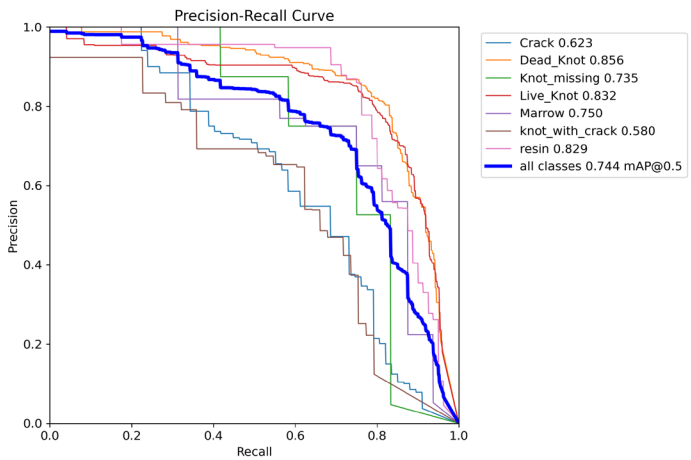


Figure 5. WSD-YOLO Model

5.2. Different Model Comparison Experiment Results and Analysis

To further validate the effectiveness of WSD-YOLO, it is compared against four representative YOLO variants—YOLOv5n, YOLOv8n, YOLOv10n, and YOLOv11n—all trained under identical settings on the wood surface defect dataset. The comparison results are presented in Table 3.

Table 3. Different Model Comparison Experiment Results

Model	mAP @0.5/%	mAP @0.5:0.95/%	Model Size/MB	FPS/（f/s）
YOLOv5n	68.2	33.5	7.5	73.8
YOLOv8n	71.4	35.9	8.0	82.4
YOLOv10n	70.4	35.2	6.5	79.2
YOLOv11n	73.2	37.7	6.6	85.6
WSD-YOLO	74.4	39.2	9.8	63.2

As shown in Table 3, WSD-YOLO achieves the highest mAP@0.5 (74.4%) and mAP@0.5:0.95 (39.2%) among all compared models. YOLOv5n, as an older anchor-based architecture, achieves 68.2% mAP@0.5. The anchor-free YOLOv8n improves to 71.4% through its C2f module, while YOLOv10n reaches 70.4%. The strong YOLOv11n baseline achieves 73.2% mAP@0.5, benefiting from its enhanced C3k2 modules and C2PSA attention.

WSD-YOLO surpasses YOLOv11n by 1.2% in mAP@0.5 and 1.5% in mAP@0.5:0.95. The most notable improvement is in Precision: WSD-YOLO achieves 72.2% versus 68.8% for YOLOv11n (**+3.4%**), indicating a substantial reduction in false positive detections on complex wood grain backgrounds. This is a critical advantage in industrial inspection, where false positives trigger unnecessary production line stoppages. Although WSD-YOLO carries a larger model footprint (9.8 MB) and reduced throughput (63.2 FPS), both values remain well within the performance envelope required for real-time wood surface inspection deployment.

5.3. Visualization Comparison Experiment

Considering the complex and variable appearance of wood surface defects, which involve issues such as fine crack boundaries, irregular knot shapes, and strong background grain interference, a visualization comparison experiment was conducted on the constructed wood surface defect test set to more intuitively verify the adaptability of the improved model. The aim is to compare the detection performance differences between the original YOLOv11n model and the proposed WSD-YOLO model.

Representative images were selected from the validation set of the wood surface defect dataset, and the detection results of the original YOLOv11n model and the proposed WSD-YOLO model were visually compared side by side. The comparison results are shown in Figures 6, 7, and 8. The left side of each figure shows the object detection results from the original YOLOv11n model, and the right side shows the results from the WSD-YOLO model. Figure 6 shows a comparison of detecting crack and knot_with_crack defects on a wood plank surface, where WSD-YOLO produces higher-confidence crack detections that the baseline fails to reliably identify. Figure 7 shows a comparison of detecting live knot, dead knot, and resin defects coexisting across multiple boards in the same batch. Figure 8 shows a comparison of detecting marrow and resin defects in scenes with complex wood grain patterns and low-contrast defect regions.

From Figure 6, it can be seen that the original model assigns low confidence scores to crack defects and tends to miss fine crack boundaries that are easily confused with wood grain lines. WSD-YOLO improves crack detection confidence and localization accuracy, attributed to the combined effect of the TAB global attention mechanism and the P2-scale branch of BPE. Figures 7 and 8 show that the original model produces incomplete detections for resin spots and marrow regions under

strong grain texture interference. The improved WSD-YOLO model reduces missed detections and false positives across all defect categories, demonstrating better detection accuracy and adaptability for wood surface defect inspection.

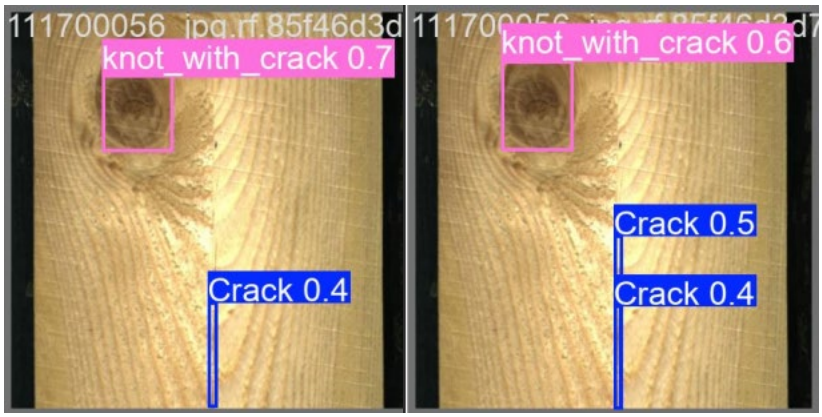


Figure 6. Comparison of YOLOv11n and WSD-YOLO Detection Results

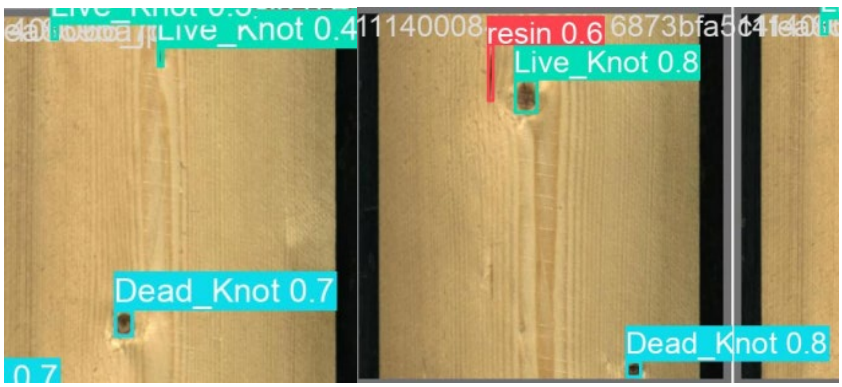


Figure 7. Comparison of YOLOv11n and WSD-YOLO Detection Results



Figure 8. Comparison of YOLOv11n and WSD-YOLO Detection Results

6. Conclusion

This paper presents WSD-YOLO, an improved YOLOv11n-based object detection algorithm specifically designed for wood surface defect detection in industrial

quality control settings. Building on the YOLOv11n baseline, WSD-YOLO integrates four targeted improvements: an Enhanced Backbone (EB) for deeper feature extraction, a Transformer Attention Block (TAB) for long-range contextual reasoning, a BiFPN with P2 Enhancement (BPE) for four-scale bidirectional feature fusion with high-resolution coverage, and a Decoupled Multi-Scale Head (DMH) for independent classification and localization optimization. These four improvements jointly address the core challenges of wood surface inspection: fine defect scale variability, complex grain texture interference, and the coexistence of diverse defect morphologies.

Experimental evaluations on a 4,000-image, 7-class wood surface defect dataset demonstrate that WSD-YOLO achieves a mAP@0.5 of **74.4%** and mAP@0.5:0.95 of **39.2%**, outperforming YOLOv11n by 1.2% and 1.5% respectively, while achieving a precision of 72.2% (+3.4%). Comparisons against YOLOv5n, YOLOv8n, YOLOv10n, and YOLOv11n confirm the superiority of WSD-YOLO on this task. The model operates at 63.2 FPS on an NVIDIA A40 GPU, satisfying real-time requirements for production-line deployment. Ablation studies confirm that each of the four improvements contributes positively and that the modules are mutually complementary.

Future work will focus on three directions: (1) improving crack detection accuracy through category-balanced data augmentation and specialized loss weighting for fine elongated defects; (2) exploring knowledge distillation techniques to compress WSD-YOLO while retaining accuracy gains; and (3) extending the approach to volumetric wood surface scanning for three-dimensional defect characterization.

References

[1] Su, B., Chen, H., & Zhou, Z. (2022). BAF-detector: An efficient CNN-based detector for photovoltaic cell defect detection. IEEE Transactions on Industrial Electronics, 69(3), 3161-3171.

[2] Zhang, Y., Zhang, H., Huang, Q., Han, Y., & Zhao, M. (2024). DsP-YOLO: An anchor-free network with DsPAN for small object detection of multiscale defects. Expert Systems with Applications, 238, 122669.

[3] Lin, J., Wang, P., Ruan, Y. et al. YOLO11-WLBS: an efficient model for pavement defect detection. Sci Rep (2026).

[4] Liu, J., Shen, X., Wang, J., Jiang, L., & Zhang, H. (2023). An intelligent defect detection approach based on cascade attention network under complex magnetic flux leakage signals. IEEE Transactions on Industrial Electronics, 71(2), 2056-2066.

[5] Park, K. B., & Lee, J. Y. (2022). Novel industrial surface-defect detection using deep nested convolutional network with attention and guidance modules. Journal of Computational Design and Engineering, 9(6), 2466-2481.

[6] Li, X., Pu, X., Ling, W. et al. YOLO-SAM an end-to-end framework for efficient real time object detection and segmentation. Sci Rep 15, 40854 (2025). https://doi.org/10.1038/s41598-025-24576-6

[7] Han Y, Qi K, Zheng J, et al. Lightweight Cattle Facial Recognition Method Based on Improved YOLOv11. Smart Agriculture, 2025, 7(3): 173-184. https://doi.org/10.12133/j.smartag.SA202502010

[8] Xu, X. et al. Exploring image enhancement for salient object detection in low light

images. ACM Trans. Multimed. Comput. Commun. Appl. 17, 1 – 19 (2021).

[9] Gao, L., Zhang, J., Yang, C., & Zhou, Y. (2022). Cas-VSwin transformer: A variant swin transformer for surface-defect detection. Computers in Industry, 140, 103689.

[10] Wang Yin, Wang Lide, Qiu Ji. Real-time Enhancement Algorithm for Orbital Dark Light Environment Based on DenseNet Structure[J]. Journal of Southwest Jiaotong University, 2022, 57(6): 1349-1357. doi: 10.3969/j.issn.0258-2724.20210199

[11] Xue R ,Duan J ,Du Z .MPE-DETR: A multiscale pyramid enhancement network for object detection in low-light images[J].Image and Vision Computing,2024,150105202-105202.DOI:10.1016/J.IMAVIS.2024.105202.

[12] Dong, X., Zhang, C., Wang, J., Chen, Y., & Wang, D. (2024). Real-time detection of surface cracking defects for large-sized stamped parts. Computers in Industry, 158, 104105.

[13] Chen, Z., Yang, C., Li, Q., Zhao, F., Zha, Z. J., & Wu, F. (2024). A Dual-Task Branch Alignment Detector for Object Detection. IEEE Transactions on Image Processing, 33, 1642-1655.