

## RESEARCH ARTICLE

# WLS-D-YOLO: A Model for Detecting Surface Defects in Wood Lumber

QIYU ZHANG<sup>1</sup>, LIPING LIU<sup>2</sup> , ZIYI YANG<sup>3</sup>, JINGTAO YIN<sup>4</sup>, AND ZHIZHONG JING<sup>5</sup><sup>1</sup>Department of Artificial Intelligence, North China University of Science and Technology, Tangshan 063000, China<sup>2</sup>College of Mechanical and Energy Engineering, Shanghai Technical Institute of Electronics and Information, Shanghai 201411, China<sup>3</sup>College of Mechanical Engineering, Tianjin University of Technology, Tianjin 300000, China<sup>4</sup>College of Electronic Technology and Engineering, Shanghai Technical Institute of Electronics and Information, Shanghai 201411, China<sup>5</sup>Hebei Sheng Aosong Wood Industry Company Ltd., Tangshan 063000, China

Corresponding author: Liping Liu (20220149@stiei.edu.cn)


**ABSTRACT** As an important raw material supporting the development of society, wood lumber is widely used in the construction and furniture industries. However, traditional methods for detecting surface defects in wood face the challenges such as poor recognition, low efficiency and narrow applicability. To tackle these challenges, this paper proposes a Wood Lumber Surface-defect Detection-YOLO (WLS-D-YOLO) model for the detection of surface defects in wood lumber. Firstly, this model introduces the squeeze-and-excitation (SE) attention mechanism in the backbone feature network component, which enhances the ability of capturing defects. Furthermore, a new GVC-neck layer structure is proposed to reduce the number of parameters and improve the accuracy of detection. Lastly, the combination of Normalized Weighted Distance Loss (NWD) small target detection algorithm and the Wise Intersection Over Union (WIOU) loss function is used to replace the original loss function to enhance the small target detection capability. The experimental results show that WLS-D-YOLO achieves an average recognition accuracy of 76.5% for wood lumber defects. Compared with the original model YOLOv8, the mean average precision (mAP) is improved by 2.9% and the frames per second (FPS) is improved by 3.8. Meanwhile, WLS-D-YOLO reduces the number of parameters to better detect several specific defects that are difficult to identify, which provides high application value for wood lumber processing and manufacturing industry.

**INDEX TERMS** Deep learning, object detection, wood lumber surface defect detection, attention mechanism, YOLO.

## I. INTRODUCTION

Wood, a natural and renewable resource, offers benefits such as easy processing, stability, and a high strength-to-weight ratio, [1]. Furthermore, it possesses unique material properties and excellent environmental characteristics, making it extensively utilized in both daily life and various production processes, [2]. The processing of logs into wood products involves several major steps, such as wood sorting, wood cutting, wood drying, wood gluing, and wood processing. After gluing the wood to form a man-made board, the randomness of the raw material as well as the complexity of the manufacturing process often lead to wood lumber producing

unavoidable defects. These lumber defects will greatly affect subsequent processing, [3], [4] so the detection and classification of lumber defects has become a major problem in the timber industry. Initially, the detection of lumber defects relied on manual visual inspection, which was not only inefficient but also susceptible to subjective factors. In the early 21st century, automated processing systems developed rapidly, and intelligent technologies began to spread to the lumber industry. Researchers have begun to utilize a variety of technologies, such as vibration detection, [5], [6] ultrasonic detection technology, [7], [8] stress wave technology [9], [10] and X-ray technology [11], [12] to detect lumber defects. However, these methods, which are based on physical equipment to inspect the interior of wood lumber, are costly and susceptible to environmental factors.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhongyi Guo .

With the development of computer technology, methods utilizing deep learning for defect detection are beginning to come into focus. Gu et al. [13] proposed a support vector machine algorithm using a tree structure, which produces an average pseudo-colored feature in each region through an order statistical filter and extracts simple and effective features to achieve accuracy improvement. The test was performed on 400 images containing decayed knots, needle knots, dead knots, live knots, and black knot defects, and the recognition of these five defect types was significantly improved compared with the original model. Zhang et al. [14] used a back propagation (BP) neural network in conjunction with Local Binary Patterns (LBP), which is a simple but useful theory for texture algorithms. The algorithm after combination can improve the extraction accuracy of the BP neural network for wood defect features. Huang et al. [15] proposed a wavelet transform-based data fusion algorithm for the previous wood defect edge detection algorithm for feature extraction of one-sided and other defects. Edge detection algorithms are used to solve the problem of segmentation and identification of wood surface defects based on the texture characteristics of wood surface defect images. Compared with early methods of detecting internal problems in wood through physical equipment, wood surface feature extraction based on computer vision is more cost effective, and the detection rate is higher too. However, traditional or early surface defect detection methods still suffer from problems of low robustness, poor accuracy, and lack of universality, which greatly limit their application in production under realistic conditions.

Recently, there has been rapid advancement in deep learning technology. An increasing number of researchers are utilizing deep learning in machine vision. Two notable methods include vision transformers (ViT) and target detection algorithms based on convolutional neural networks (CNN). The introduction of ViT provides another method for global feature learning, and representative vision transformers, such as Adavit [16], Crossvit [17], DETR [18], have achieved good results in traffic sign recognition and medical image classification [19], [20]. But such methods have long training time and poor real-time performance, which are difficult to be applied in the field of defect detection with real-time requirements. The target detection algorithm based on CNN convolutional neural network has become the mainstream algorithm in the field of defect detection with continuous progress. There are two most representative target detection algorithms, one is a single-stage detection algorithm based on regression, such as the YOLO series [21], SSD [22], EfficientDet [23], and so on. The other is a two-stage detection algorithm based on candidate bounding boxes, such as Fast R-CNN [24], Faster R-CNN [25], Cascade R-CNN [26], Trident-Net [27], etc. Yang et al. [28] employed TensorFlow to construct the network, substituting the visual geometry group (VGG) component of the initial single shot multi-box detector (SSD) network with a deep residual network.

This modification aims to enhance input feature optimization for both regression and classification tasks associated with predicting bounding boxes. The input features for the classification task were optimized to detect defects in more than 5000 solid wood lumbers, which improved both the detection accuracy and detection speed, but it is still difficult to meet the real-time requirements in industry. Xia et al. [29] proposed a detection algorithm for surface defects on wood lumber, leveraging an improved Faster R-CNN. The algorithm integrates a feature pyramid network with a variable convolutional Res-Net50 network to generate semantic feature maps for detecting defects. Additionally, it incorporates a cost benefit analysis method (CBAM) attention mechanism to enhance detection accuracy, despite achieving a high level of accuracy, the issue of slow detection persists. Wang et al. [30] To enhance the ability to recognize small target defects in wood, content-based attention (CA) is integrated with omni-dimensional dynamic convolution (ODConv) to produce the dynamic attention mechanism ODCA incorporated into YOLOv7, and an efficient feature extraction network block called S-HorBlock is introduced to achieve good detection results. Nonetheless, research on quartzity, a prevalent defect, is lacking. Additionally, studies on the real-time capabilities of the suggested model are scarce. Zheng et al. [31] proposed a high-precision, lightweight method for real-time detection of wood surface defects based on the YOLO model. This approach enhances the accuracy and speed of identifying surface defects across four typical wood lumbers. Yet, problems remain due to the limited dataset size for training and inadequate coverage of defect types. Haonan et al. [32] constructed a target detection framework for YOLOv5 wood lumber surface defects using a focus, feature pyramid networks (FPN), and path aggregation network (PAN) structure by introducing the squeeze-and-excitation (SE) attention mechanism in the backbone feature network, which provides reference ideas for the detection of four common defects: live knots, dead knots, cracks, and holes, although the detection speed meets the real-time requirements, there is a decline in overall accuracy and a lack of research on the remaining lumber surface defects.

In conclusion, researchers have made significant progress utilizing target detection algorithms for detecting surface defects in wood lumbers. However, several problems remain: 1. Most of the existing datasets openly used for the detection of surface defects in wood lumbers are small in size and focus on a single or common types of defects, such as live knots, dead knots, cracks, holes and so on, which are unable to cover all the defects in the industrial production, and are inevitably limited in practical applications. 2. In practical applications, high real-time performance and high detection rate are equally important, and most of the existing research focuses on one of them, which makes it difficult to achieve a balance between speed and accuracy. 3. Existing target detection algorithms suffer from poor results in detecting

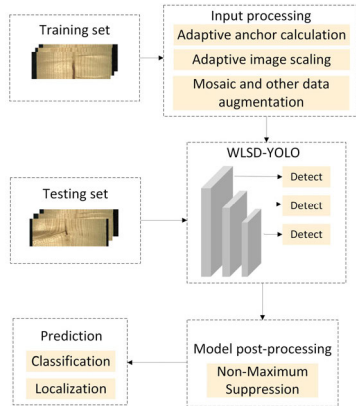


FIGURE 1. WLSO-YOLO workflow.

small targets, and a low rate of identifying targets similar to the background [33].

To solve the above problems and better meet the needs of the industrial production of wood lumber, this study selects a data set of wood lumber surface defects collected from the assembly line of sawmills, from which 4000 images are selected and 8 most common types of wood lumber surface defects are covered. In addition, this paper proposes a WLSO-YOLO model for the recognition of surface defects of wood lumber. The workflow of WLSO-YOLO is shown in Fig. 1. Firstly, adaptive image scaling, adaptive anchor frame computation and data enhancement are performed at the input. Then the extraction and enhancement of defective features are performed using the WLSO-YOLO model. Finally, the detection of surface defects on wood lumber is completed. The contributions of this paper are as follows:

(1) Due to the presence of some defects in wood lumber that are not well differentiated from the background, the SE Attention Mechanism module is introduced in the backbone feature extraction layer, which performs adaptive feature refinement and reduces the background feature training weights by multiplying the attention feature maps obtained from the channel and spatial dimensions into the input feature maps.

(2) For the purpose of improving the network’s ability to identify different defects in wood lumbars, and at the same time to improve the detection speed of the model, this paper proposes a new kind of neck structure-GVC-neck, which can achieve feature extraction at different scales, not only improve the accuracy, but also reduce the number of parameters and the size of the model to a certain extent, so as to achieve a balance between the accuracy and the detection speed.

(3) With the aim of reducing the negative impact of low-quality images on the model and improve the recognition ability of small target defects, the WIOU (Wise-IOU) loss function is introduced and combined with the NWD (Normalised Wasserstein Distance) small target detection mechanism to further improve the overall

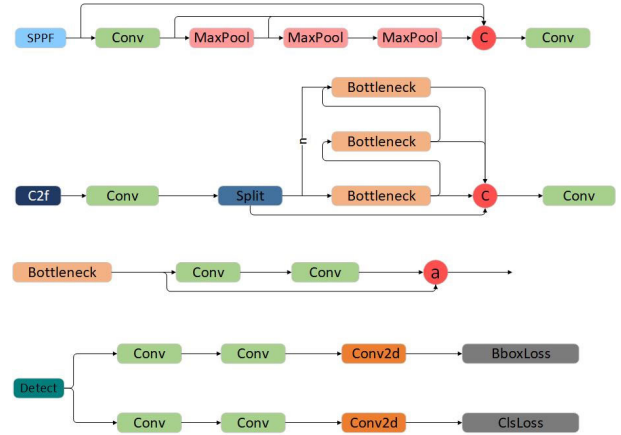


FIGURE 2. YOLOv8 specific structure diagram.

recognition ability of the model for surface defects of wood lumber.

The subsequent contents of this paper are organised as follows: the original model YOLOv8 and the WLSO-YOLO model proposed in this paper are introduced in Section II. Section III describes the experimental preparations and analyses the experimental results to demonstrate the sophistication of the WLSO-YOLO model. Finally, Section IV draws the main conclusions as well as future directions.

## II. PRINCIPLE AND METHOD IMPROVEMENT

### A. YOLOV8 MODEL

YOLOv8 has achieved good results in the field of defect detection, with higher accuracy as well as faster speed compared to previous models, and is very easy to deploy, which is very suitable for industrial application scenarios that require online real-time processing. Therefore, YOLOv8 is chosen as the original model in this paper, and the specific modules of YOLOv8 are explained in this subsection. Similar to its predecessor, YOLOv5, by adjusting the network scaling factor, models of different sizes (N/S/M/L/X) can be chosen —to satisfy diverse scene requirements [34].

As shown in Figure 2, YOLOv8 replaces the C3 module with the C2f module on the basis of YOLOv5 and adjusts the number of channels, and the C2f module was designed with reference to the C3 module as well as the concept of ELAN. The C2f module is designed with reference to the C3 module and the idea of ELAN. Since the C2f structure has more residual connections, it allows YOLOv8 to obtain richer gradient flow information while keeping light weight. In the backbone part, the SPP in v5 is replaced by Spatial Pyramid Pooling-Faster (SPPF), which reduces the computation amount to a certain extent and increases the receptive field, which can effectively improve the detection accuracy. The head separates the CIs and Boxes for prediction and uses the decoupling head, which is changed from anchor-based to anchor-free. The loss aspect uses positive and negative matching of samples and introduces distribution focal loss instead

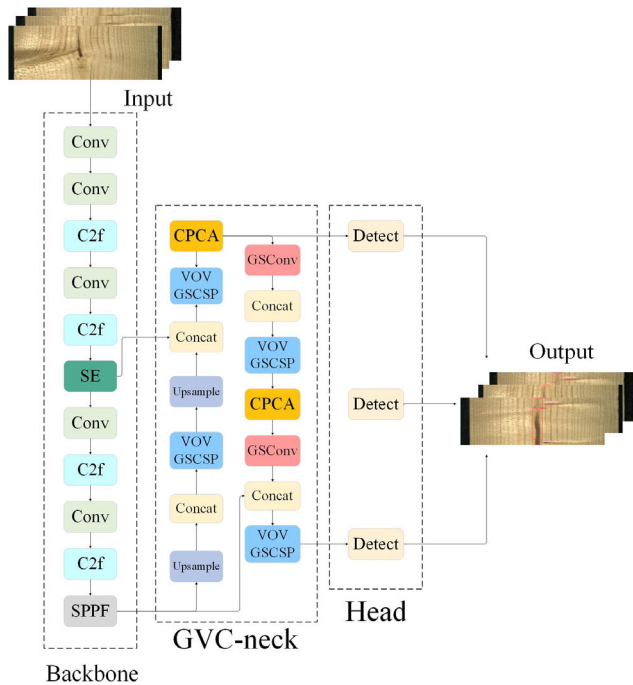


FIGURE 3. WLS-YOLO model.

of simple IOU matching. The YOLOv8 network structure was more streamlined, with a faster detection speed and higher detection accuracy.

**B. THE PROPOSED WLS-YOLO**

To solve the problem of poor detection effect of the existing models for the detection of surface defects on some wood lumbers, this paper proposes a YOLO model (WLS-YOLO) for detecting surface defects on wood lumbers using YOLOv8 as the original model. The SE attention mechanism is incorporated into the backbone to enhance the model’s feature fusion capability. Additionally, the GVC-neck structure is introduced to decrease the parameter count and size of the original model, leading to a dual enhancement in both detection accuracy and speed. Finally, the NWD small target detection mechanism is used in combination with the WIOU loss function to improve the model’s detection ability for small target defects. The structure of WLS-YOLO is shown in Fig. 3. Among them, Conv is a convolution module and Concat layer is used to merge the number of channels. Upsample is an upsampling module and detection module is used to localise and classify the detected targets. The parts of SE, VOV-GSCSP, GSconv and CPCA will be explained in detail later.

**1) SE-BACKBONE**

While pursuing accuracy as the primary goal and at the same time not increasing the computational burden of the model as much as possible, this paper adopts the mechanism of adding Squeeze-and-Excitation (SE) [35] channel attention

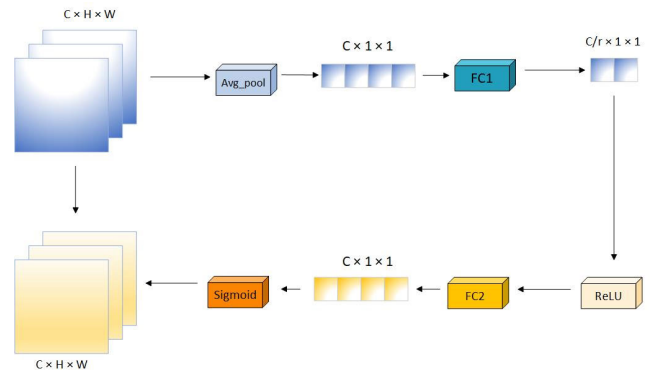


FIGURE 4. SE-Attention module structure.

to the main part to improve model performance. The core function of the SE module is to assign different weights to the image according to the different positions of the channel through the weighting matrix, so as to better capture key feature information. As shown in Figure 4, the idea of SE attention is to first perform global average pooling (Avg\_pool) on the input feature maps in order to compute the average value of each channel, which will be used as the initial weights reflecting the importance of the channel. Then, passing through two fully connected layers sequentially, the first fully connected layer (FC1) is used for dimensionality reduction to reduce the number of input channels to smaller dimensions in order to reduce the number of parameters and computational complexity of the model. Then, after the Rectified Linear Activation Function (ReLU) activation layer, the dimensionality was kept constant. Next, a second fully connected layer (FC2) is used for dimensionality upgrading to map the downgraded vectors back to the dimensions of the original number of channels in order to learn the importance of the weights of each channel. The structure and parameters of both fully connected layers can be trained and optimized by backpropagation to maximize the model performance and detection accuracy. Finally, a sigmoid function is used to limit the output range to between (0 and 1), which is used as the weight of the channel attention mechanism and multiplied with the original feature maps to obtain the final feature maps incorporating the attention mechanism. By establishing feature mapping relationships between channels, SE makes full use of global information to assign higher weights to the channel feature information of small targets. This in turn helps the model to better adapt to the relevant feature information between the channels of small targets while suppressing and ignoring irrelevant information, so that the model is more focused on training the class-specific task targeting the detection of defects in wood lumber.

**2) GVC-NECK**

The original neck layer of YOLOv8 adopts the structure of feature fusion, which is used to fuse features from different layers; however, there are problems such as redundant

information mixing and large computation amounts. In order to solve these problems, this paper proposes a new type of neck layer network, the GVC-neck, which introduces Group Shuffle Convolution (GSCConv) convolution and VoV-GSCSP modules into the original neck layer structure, incorporates the CPCA attention mechanism to optimize the number of parameters and the model, and also greatly improves the detection effect for the surface defects of wood lumber.

GSCConv [36] reduces the computational cost of the model by using a combination of grouped convolution and depth-separable convolution, as shown in Figure 5 (A). First, a normal convolutional (Conv) downsampling is performed to change the number of channels in the input to half of the original number, then a deep convolution (DWConv) is used on it, and the outputs of the normal convolution (SC) and the outputs of the deep convolution (DSC) are spliced together by a concat operation. Finally, the SC-generated information was infiltrated into each part of the DSC-generated information using shuffle, a method that allows the information from the SC to be completely blended into the output of the DSC. The neck layer network using GSCConv convolution not only minimizes the negative impact of deep convolution on the model but also makes effective use of its advantages of small size and high speed.

The VoV-GSCSP module, which can simplify the network structure and reduce the model computation, is introduced in the neck layer, and the structure of the VoV-GSCSP [36] module is shown in Figure 5 (B). The C2f module aggregates intermediate features with different receptive field sizes by means of dense connections, but it shows the disadvantages of slow operation speed and inefficiency. The VoV-GSCSP module uses a one-time The VoV-GSCSP module that uses a one-time aggregation method to design the cross-level partial network (GSCSP), which only aggregates the features of all the previous layers in the last layer of the module, which not only inherits the advantages of the dense connection with multiple receptive fields to represent a variety of features but also solves the inefficiency of the module. The VoV-GSCSP also adopts the idea of GSCConv convolution, the output of ordinary convolution, and after two GSCConv convolution outputs through concatenation. Finally, in the case of the number of unchanged channels, ordinary convolution is used to further extract image features. The VoV-GSCSP module not only simplifies the computation and network structure but also preserves ample accuracy.

In order to further enhance the network’s ability to extract wood defect features and to solve the impact of useless information on the network’s detection performance, a lightweight attention mechanism, Channel Prior Convolutional Attention (CPCA), is added to the neck layer. CPCA [37] combines spatial and channel attention mechanism modules, the main purpose of which is to dynamically allocate attention weights over channel and spatial dimensions to maximize the extraction of important feature information while avoiding the number of parameters raised due to complex computation.

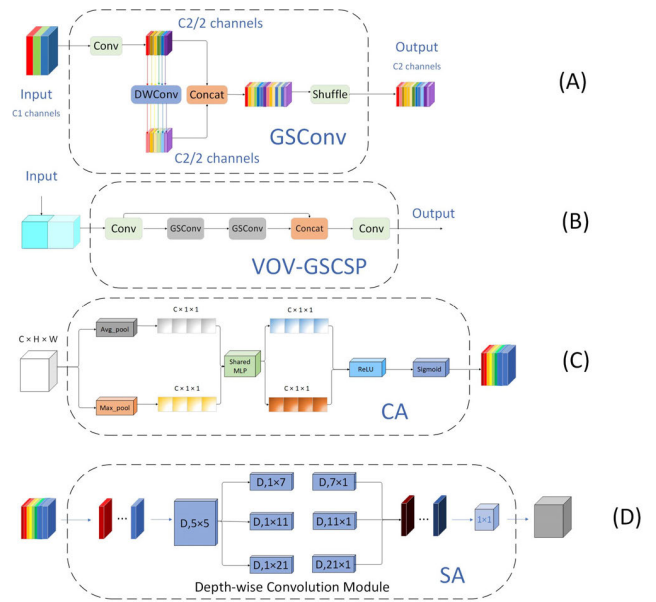


FIGURE 5. Structure of GVC-neck specific modules.

As shown in Figure 5 (C), Channel Attention (CA) mechanism, the input feature maps are passed through two parallel Max\_pool and Avg\_pool layers to aggregate spatial information from feature mapping and then fed into the shared MLP. Similar to the SE Attention, the number of input channels is also compressed before expanding here, aiming at improving the network’s ability to express features and to be more specific to the current task. Subsequently, two activated results are obtained after processing through the ReLU activation function. Following this, the two outputs are element-wise added, and then, after passing through the sigmoid activation function, the output undergoes channel attention processing. Ultimately, the output is restored to the original size of the image. The specific expression for CA is as follows:

$$CA(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (1)$$

As shown in Figure 5 (D), the spatial attention mechanism (SA) first adopts the idea of multi-scale structure to separate the input feature map into multiple parts so that the local features of the image can be described in a simple form on different scales, and then utilizes depth separable convolution for each part to capture the spatial relationship between the features, which ensures the preservation of the inter-channel relationship while simultaneously reducing the computational complexity. Finally, a  $1 \times 1$  convolution is used for channel blending to fuse the results of each part and realize information interaction between channels. The specific expression for the SA is as follows:

$$SA(F) = Conv_{1 \times 1} \left( \sum_{i=0}^3 Branch_i(DWConv(F)) \right) \quad (2)$$

The GVC-neck layer network achieves more efficient and representative feature extraction and classification of wood defects by combining GSConv convolution, VoV-GSCSP module, and CPCA attention mechanism. The combination of the three can reduce the computational complexity and number of parameters of the model, while improving its performance and generalization ability.

### 3) WIOU-NWD

The original loss function CIU lacks consideration for balancing difficult and easy samples. This oversight results in geometric metrics like aspect ratio and distance amplifying negative gradients from low-quality samples, thus compromising the model's generalization capabilities. For this reason, the WIOU loss function is used in the WLS-D-YOLO model as a replacement for the original loss function CIU, assuming that  $w, h$  denote the width and height of the prediction frame, respectively, and that  $w^{gt}, h^{gt}$  are the width and height of the real frame,  $b, b^{gt}$  represent the center points of the boundary between the predicted frame and the actual frame,  $\rho$  represents the Euclidean distance between  $b$  and  $b^{gt}$ .  $w^c$  and  $h^c$  represent the width and height of the smallest outer rectangle of both the prediction box and the real box, while IOU denotes the intersection-over-union. The CIU expression can be obtained as follows:

$$L_{CIU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{(w^c)^2 + (h^c)^2} + \alpha v \quad (3)$$

Among them:

$$\alpha = \frac{v}{(1 - IOU) + v} \quad (4)$$

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (5)$$

Wise-IOU(WIOU) [38] evaluated the quality of anchor frames by using a dynamic non-monotonic focusing mechanism and used a gradient gain that does not intervene in the training too much to ensure high-quality anchor frames and at the same time reduces the influence of harmful gradients, which can improve the overall performance of the algorithm. Overall performance of the algorithm, the WIOU constructs a two-layer attention mechanism to accelerate the convergence speed, improve the convergence accuracy, and enhance the model generalization ability. Assuming that the corresponding position of  $(x, y)$  in the target frame is  $(x^{gt}, y^{gt})$ ,  $R_{WIOU}$  denotes the loss of high-quality anchor frames, and WIOUv1 is expressed as

$$L_{WIOUv1} = R_{WIOU} L_{IOU} \quad (6)$$

Among them:

$$R_{WIOU} = \exp\left(\frac{(x - x^{gt})^2 - (y - y^{gt})^2}{((w^c)^2 + (h^c)^2)^*}\right) \quad (7)$$

In order to prevent the low-quality samples from generating large and harmful gradients, WIOUv3 was constructed

using WIOUv1 to construct the WIOUv3 with the following equation:

$$L_{WIOUv3} = r L_{WIOUv1} \quad (8)$$

$$r = \frac{\beta}{\delta \alpha^{\beta - \delta}} \quad (9)$$

$$\beta = \frac{L_{IOU}^*}{L_{IOU}} \quad (10)$$

Due to the existence of some small target defects in lumber defects, the WLS-D-YOLO model applies the NWD small target detection algorithm [39] in the loss function and uses it together with WIOUv3 to enhance the recognition accuracy of small target defects in lumbars. IOU-based metrics are highly sensitive to positional deviations in small objects, posing a risk of substantial deterioration in detection performance when applied in anchor-based detectors. In response to this challenge, the NWD employs a novel metric termed the normalized Wasserstein distance, defined by the following expression:

$$NWD(N_a, N_b) = \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{C}\right) \quad (11)$$

Application of NWD to the loss function:

$$L_{NWD} = 1 - NWD(N_p, N_g) \quad (12)$$

NWD has more advantages in detecting tiny objects than IOUs. Among them are the Gaussian distribution of the prediction frame and the Gaussian distribution of the GT frame. In the algorithm of this paper, the ratio coefficient of the use of WIOUv3 and NWD is 1:1.

## III. EXPERIMENTAL DETAILS AND COMPARISONS

### A. TRAINING ENVIRONMENT AND PARAMETER SETTINGS

The training and testing hardware platform is a cloud server GPU platform, and the main hardware configuration is shown in Table 1.

TABLE 1. Experimental environment.

serial number	name (of a thing)
GPUs	GeForce RTX 4090
Video Memory/GB	24g
Training environment	CUDA 11.7 CUDNN 11.1
operating system	Linux
development environment (computer)	Python 3.8.10 Pytorch 2.0.1

### B. PERFORMANCE METRICS

The main evaluation indexes of target detection algorithms fall into two categories: detection precision and model complexity. The detection precision is mainly reflected by the model's accuracy precision (P), recall (R), and mean average precision (mAP). Assuming that the number of positive and negative samples are (True Positive) TP and (False Positive) FP, respectively, and the number of positive samples is (False

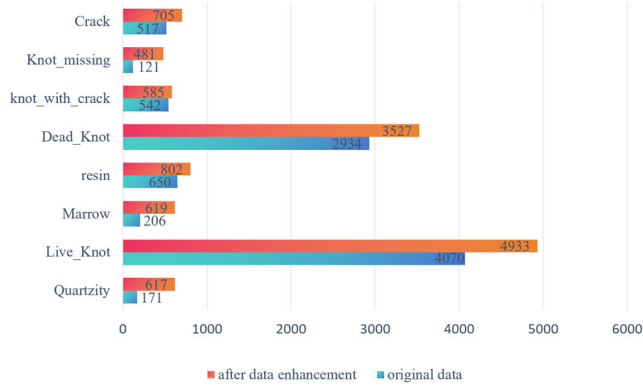


FIGURE 6. Comparison of before and after data enhancement.

Negative) FN, the formulas for precision P, recall R, and mean average precision mAP are as follows:

$$precision = \frac{TP}{TP + FP} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

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

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

C. DATA SETS AND PRE-PROCESSING

In this study, we used a large image dataset of surface defects in wood lumber collected from sawmill assembly lines after gluing and splicing, which contained 10 defects. The image size was 2800 × 1024 pixels. This paper excludes two rare defects, blue stain and overgrown, and retains 4000 images of wood with eight representative defects, including quartzity, live\_knot, marrow, resin, dead\_knot, knot\_with\_crack, knot\_missing, and crack.

In order to solve the problem of an uneven number of types in the original dataset and to ensure that the model has better robustness and generalization ability, as well as to prevent the generation of overfitting phenomena, this paper employs a variety of methods, such as cropping, mirroring, panning, rotating, adding noise, and altering the brightness, to carry out data enhancement operations on the images with fewer types of defects. The results of the comparison of the number of defects of each type before and after data enhancement are shown in Figure 6.

The dataset is divided into training set, validation set and test set in the ratio of 8:1:1 and the number of samples in these three sets are 3971, 497, 497.

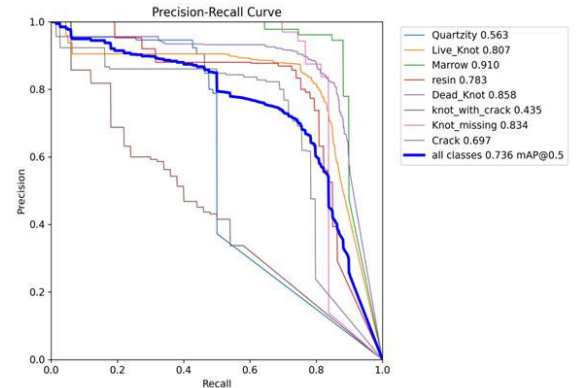
D. DISCUSS

1) ABLATION EXPERIMENTS

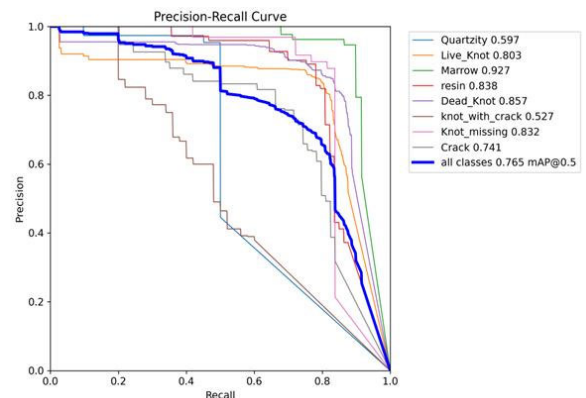
To validate the effectiveness of each improvement in the WLS-D-YOLO model, this paper employed the YOLOv8

TABLE 2. Ablation experiments.

serial number	SE	GVC-neck	WIOU-NWD	FLOPs	Para/M	mAP /%	FPS
1	×	×	×	258.5	68.22	73.6	48.07
2	√	×	×	258.5	68.24	74.3	46.35
3	√	√	×	218.6	55.83	75.8	51.28
4	√	√	√	218.6	55.83	76.5	51.81



(A) YOLOv8



(B) WLS-D-YOLO

FIGURE 7. Comparison of PR curves.

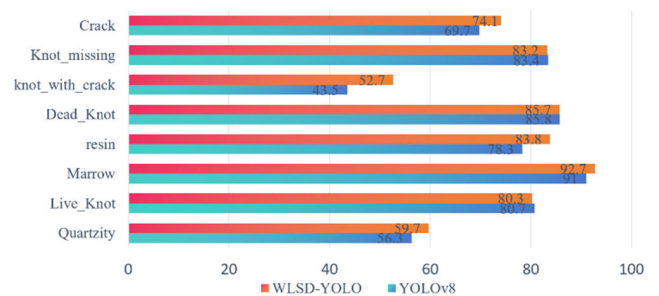
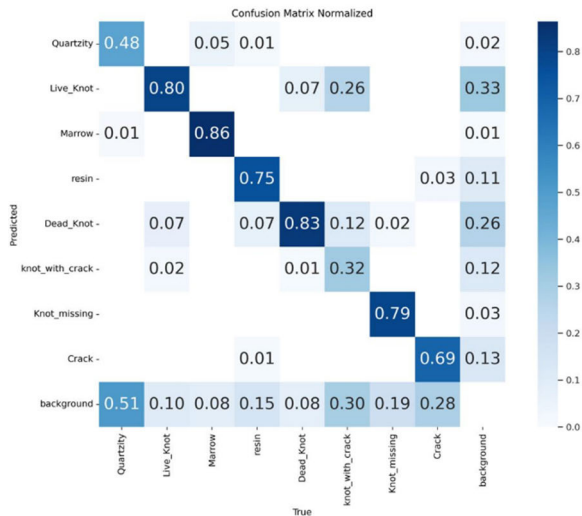
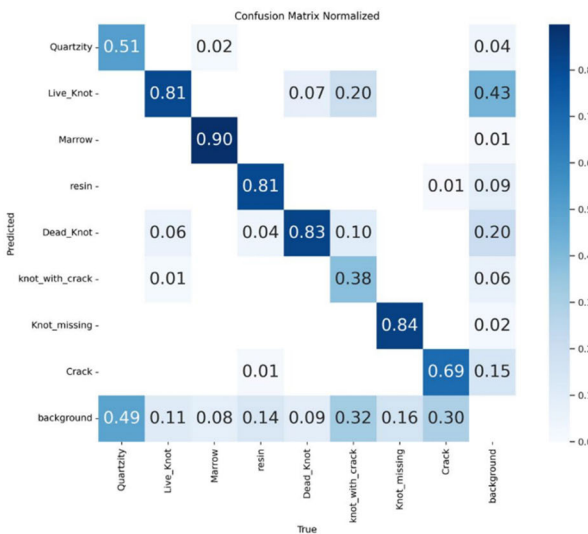


FIGURE 8. Comparison of category accuracy.

model as a reference benchmark, and experiments were conducted incrementally by introducing each improvement point step by step, as shown in Table 2. First, adding the SE attention mechanism to the backbone network improved the accuracy by 0.7 percentage points, although the FPS



(A) YOLOv8



(B) WLSO-YOLO

FIGURE 9. Confidence level comparison chart.

was slightly reduced. Then, the original neck layer in YOLOv8 was optimized using the GVC-neck. GVC-neck greatly reduces the number of parameters in the original model, which improves the accuracy while also increasing the detection speed. Finally, by modifying the loss function, the original CIUO was replaced by WIOU-NWD, which improved the accuracy by 0.7 percentage points. Compared to the original model, the WLSO-YOLO model achieved an overall 2.9% accuracy improvement, a 19% decrease in the number of parameters, a 3.8 improvement in the FPS, and a 39.9 reduction in the FLOPs. A comparison of the PR graphs is shown in Figure 7. The accuracy of knot\_with\_crack, crack, and quartzity, which are the categories of defects with poor recognition results, was also substantially improved, and the category accuracy comparison graph is shown in Figure 8.

The category confidence comparison plot is shown in Figure 9, the confusion matrix serves as a concise summary

TABLE 3. Ablation experiments.

serial number	methodologies	FLOPs	Para/M	mAP/%	FPS
1	YOLOv5	205.7	86.7	73.0	86.95
2	YOLOv7	188.1	71.3	72.6	30.86
3	YOLOX	281.9	99.1	70.2	35.75
4	Faster-Cnn	-	41.3	65.2	15.23
5	YOLOv8	258.5	68.2	73.6	48.07
6	WLSO-YOLO (proposed)	215.9	55.8	76.5	51.81

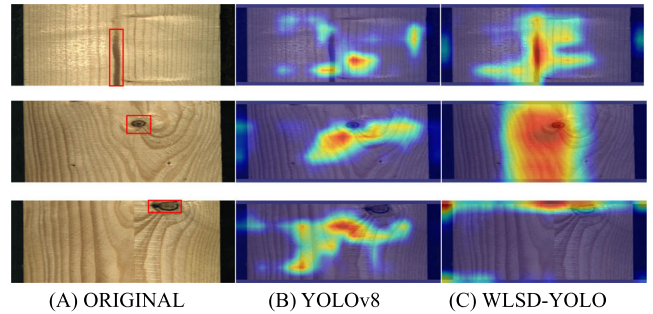


FIGURE 10. Grad-CAM Comparison Chart.

of the classification problem’s prediction outcomes. It encapsulates both correct and incorrect predictions through count values, segmenting them by class. This matrix illustrates how effectively the network model identifies each defect type during the prediction process. As can be seen from the figure, compared with the original model YOLOv8, the WLSO-YOLO model has higher recognition and accuracy for several surface defect categories of wood lumber such as Quartzity, resin, knot\_missing, knot\_with\_crack, Marrow, etc., which significantly improves the classification ability of the original model for different defects.

2) COMPARISON WITH OTHER MODELS

To further validate and evaluate the detection performance of the proposed method, comparative tests are conducted on YOLOv5, YOLOv7, YOLOX, and the traditional network model Faster-Cnn, using the same hardware resources and datasets. The results are shown in Table 3. It can be seen that the traditional network model has a greater disadvantage in wood lumber defect detection; both accuracy and speed are far behind those of the newer YOLO series algorithms. Although the WLSO-YOLO model lags behind YOLOv5 in terms of speed, its accuracy has been greatly improved, and it is more suitable for the wood lumber manufacturing industry, where accuracy is the first priority. For YOLOv7, YOLOX, and YOLOv8, the WLSO-YOLO model has significant advantages in terms of accuracy, speed, and parameter size, reflecting its excellent defect detection performance.

3) GRAD-CAM CHART EVALUATION

Grad-CAM is an attention heatmap technique for visualizing deep neural networks that helps us analyze the image regions



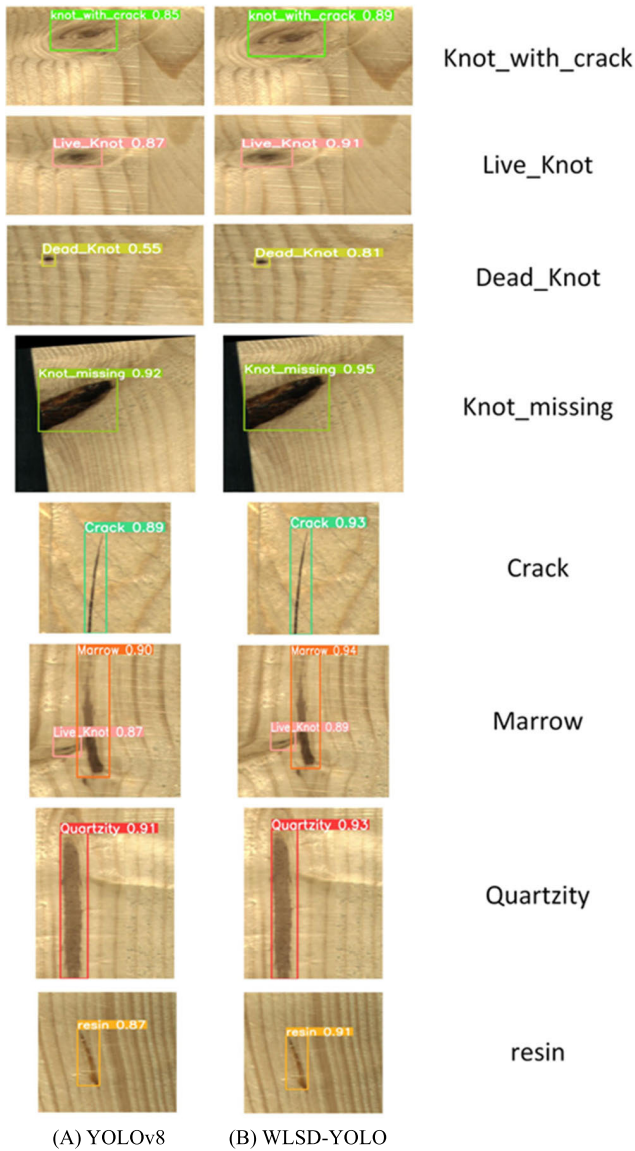


FIGURE 11. Visualization Comparison Chart.

to which the network pays attention for a given category, depicting the distribution of the input image’s contribution to the output prediction. As shown in Figure 10, the locations circled in red in the original diagram are defects, the original model yolov8 suffers from inaccurate localisation when locating certain defects, whereas the WLS-D-YOLO model can alleviate these problems to a certain extent in terms of locating regions associated with wood defects. It can be seen that WLS-D-YOLO covers the target object region better than YOLOv8. This suggests that adding attention can be a good way to learn and use the information in the target object region and aggregate features from it.

4) COMPARISON OF VISUALIZATION EVALUATION

Figure 11 shows a visualized comparison graph between the above models, where each picture contains at least one

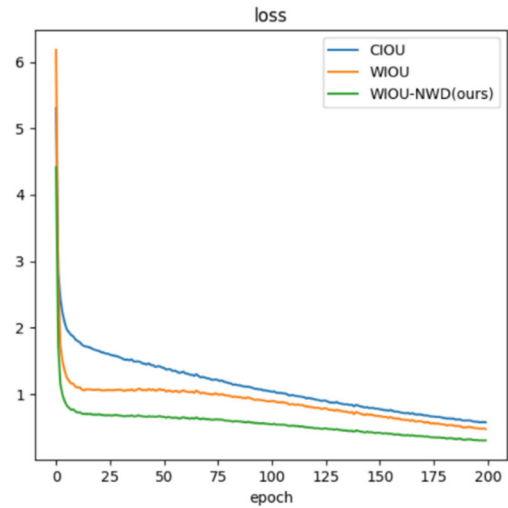


FIGURE 12. Comparison of loss function curves.

surface defect of wood lumber. The visualization result graph can more intuitively show the recognition ability of the models for different defects of wood lumber, and it is easier to reflect the advantages and disadvantages between the models. It can be seen that the advantages of the WLS-D-YOLO model in accurately locating and detecting surface defects on all wood lumbers with higher accuracy are obvious, better results in detecting most wood lumber surface defect images.

5) LOSS FUNCTION PLOTS

The loss function is a measure of the distance between the neural network’s predicted information and the expected information (labels); the closer the predicted information is to the expected information, the smaller the loss function value is. The faster the convergence rate, the easier it is for the model to converge to stability. Therefore, a faster convergence of the loss function and a smaller total loss represent superior performance. A comparison of the loss curves of the original loss function and WIOU as well as the loss function method of this paper on the dataset of this paper is shown in Figure 12, where the horizontal axis epoch represents the number of training rounds of the model, and the vertical axis represents the loss value of the model. It can be seen that the loss function in this paper converges faster and the total loss is lower than the WIOU and the original CIOU loss function, which proves the effectiveness of the loss function in this paper.

IV. CONCLUSION

For the advancement of smart forestry, this paper introduces the WLS-D-YOLO model for detecting defects in wood boards. WLS-D-YOLO enhances feature extraction and small-target defect detection capabilities by incorporating the GVC-neck layer structure, integrating SE attention mechanisms into the backbone, and modifying the loss

function. The test results on a dataset containing eight types of surface defects on wood lumbers indicate that WLS-D-YOLO achieved a mAP of 76.5%, an improvement of 2.9% compared to the baseline model, the average detection time is 0.019 seconds, fully meets the real-time requirements for detecting surface defects on wood lumbers in industrial scenarios, This can enhance the detection accuracy of various wood lumber defects, showing a certain advantage compared to other algorithms and demonstrating stronger competitiveness in practical applications. Nonetheless, there is scope for enhancing this algorithm. While overall detection accuracy has been enhanced, it remains inefficient in identifying two defect types: “knot\_with\_crack” and “Quartzity.” Consequently, this study will persist in refining the model’s recognition capabilities for specific defects, aligning them more closely with the accuracy demands of industrial scenarios. Going forward, efforts will be made to further diminish the number of model parameters and reduce model size to boost defect detection speed and alleviate the cost demand associated with model deployment.

## REFERENCES

- [1] R. B. Hoadley, *Understanding Wood: A Craftsman’s Guide to Wood Technology*. Newtown, NC, USA: Taunton press, 2000.
- [2] M. C. Barbu and E. M. Tudor, “State of the art of the Chinese forestry, wood industry and its markets,” *Wood Mater. Sci. Eng.*, vol. 17, no. 6, pp. 1030–1039, Nov. 2022.
- [3] M. Altgen, S. Adamopoulos, and H. Militz, “Wood defects during industrial-scale production of thermally modified Norway spruce and scots pine,” *Wood Mater. Sci. Eng.*, vol. 12, no. 1, pp. 14–23, Jan. 2017.
- [4] F. H. Schweingruber, *Wood Structure and Environment*. Cham, Switzerland: Springer, 2007.
- [5] Y. Chen, C. Sun, Z. Ren, and B. Na, “Review of the current state of application of wood defect recognition technology,” *BioResources*, vol. 18, no. 1, pp. 1–14, Dec. 2022.
- [6] G. Defflorio, S. Fink, and F. W. M. R. Schwarze, “Detection of incipient decay in tree stems with sonic tomography after wounding and fungal inoculation,” *Wood Sci. Technol.*, vol. 42, no. 2, pp. 117–132, Feb. 2008.
- [7] Y. Fang, L. Lin, H. Feng, Z. Lu, and G. W. Emms, “Review of the use of air-coupled ultrasonic technologies for nondestructive testing of wood and wood products,” *Comput. Electron. Agricult.*, vol. 137, pp. 79–87, May 2017.
- [8] H. Yang and L. Yu, “Feature extraction of wood-hole defects using wavelet-based ultrasonic testing,” *J. Forestry Res.*, vol. 28, no. 2, pp. 395–402, Mar. 2017.
- [9] X. Li, W. Qian, L. Cheng, and L. Chang, “A coupling model based on grey relational analysis and stepwise discriminant analysis for wood defect area identification by stress wave,” *BioResources*, vol. 15, no. 1, pp. 1171–1186, Jan. 2020.
- [10] X. Du, J. Li, H. Feng, and S. Chen, “Image reconstruction of internal defects in wood based on segmented propagation rays of stress waves,” *Appl. Sci.*, vol. 8, no. 10, p. 1778, Sep. 2018.
- [11] Q. Wang, X. Liu, and S. Yang, “Predicting density and moisture content of populus xiangchengensis and phyllostachys edulis using the X-ray computed tomography technique,” *Forest Products J.*, vol. 70, no. 2, pp. 193–199, Mar. 2020.
- [12] Q. Qiu, “Thermal conductivity assessment of wood using micro computed tomography based finite element analysis ( $\mu$ CT-based FEA),” *NDT E Int.*, vol. 139, Oct. 2023, Art. no. 102921.
- [13] I. Y.-H. Gu, H. Andersson, and R. Vican, “Wood defect classification based on image analysis and support vector machines,” *Wood Sci. Technol.*, vol. 44, no. 4, pp. 693–704, Nov. 2010.
- [14] Y. X. Zhang, Y. Q. Zhao, Y. Liu, L. Q. Jiang, and Z. W. Chen, “Identification of wood defects based on LBP features,” in *Proc. 35th Chin. Control Conf. (CCC)*, Jul. 2016, pp. 4202–4205.
- [15] Y. Huang, M. Yi, W. Yang, and M. Yang, “Wood defect recognition based on wavelet transform data fusion edge detection algorithm,” in *Proc. 7th Int. Conf. Cyber Secur. Inf. Eng.*, Sep. 2022, pp. 236–239.
- [16] L. Meng, H. Li, B.-C. Chen, S. Lan, Z. Wu, Y.-G. Jiang, and S.-N. Lim, “AdaViT: Adaptive vision transformers for efficient image recognition,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 12299–12308.
- [17] C. R. Chen, Q. Fan, and R. Panda, “CrossViT: Cross-attention multi-scale vision transformer for image classification,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 347–356.
- [18] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. J. A. P. A. Dai, “Deformable detr: Deformable transformers for end-to-end object detection,” 2020, *arXiv:2010.04159*.
- [19] Y. Zheng and W. Jiang, “Evaluation of vision transformers for traffic sign classification,” *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–14, Jun. 2022, doi: 10.1155/2022/3041117.
- [20] S. Tummala, S. Kadry, S. A. C. Bukhari, and H. T. Rauf, “Classification of brain tumor from magnetic resonance imaging using vision transformers ensembling,” *Current Oncol.*, vol. 29, no. 10, pp. 7498–7511, Oct. 2022.
- [21] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, “A review of YOLO algorithm developments,” *Proc. Comput. Sci.*, vol. 199, pp. 1066–1073, May 2022.
- [22] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “SSD: Single shot multibox detector,” in *Proc. 14th Eur. Conf. Amsterdam*, The Netherlands: Springer, Oct. 2016, pp. 21–37.
- [23] M. Tan, R. Pang, and Q. V. Le, “EfficientDet: Scalable and efficient object detection,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10778–10787.
- [24] R. Girshick, “Fast R-CNN,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
- [25] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
- [26] Z. Cai and N. Vasconcelos, “Cascade R-CNN: Delving into high quality object detection,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6154–6162.
- [27] Y. Li, Y. Chen, N. Wang, and Z.-X. Zhang, “Scale-aware trident networks for object detection,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6054–6063.
- [28] Y. Yang, H. Wang, D. Jiang, and Z. Hu, “Surface detection of solid wood defects based on SSD improved with ResNet,” *Forests*, vol. 12, no. 10, p. 1419, Oct. 2021.
- [29] B. Xia, H. Luo, and S. Shi, “Improved faster R-CNN based surface defect detection algorithm for plates,” *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–11, May 2022.
- [30] R. Wang, F. Liang, B. Wang, and X. Mou, “ODCA-YOLO: An omni-dynamic convolution coordinate attention-based YOLO for wood defect detection,” *Forests*, vol. 14, no. 9, p. 1885, Sep. 2023.
- [31] Y. Zheng, M. Wang, B. Zhang, X. Shi, and Q. Chang, “GBCD-YOLO: A high-precision and real-time lightweight model for wood defect detection,” *IEEE Access*, vol. 12, pp. 12853–12868, 2024.
- [32] J. Haonan, X. Huadong, W. Lihai, Z. Jinsheng, C. Xiaohui, and T. Xu, “Quantitative identification of surface defects in wood paneling based on improved YOLOv5,” *J. Beijing Forestry Univ.*, vol. 45, no. 4, pp. 147–155, 2023.
- [33] Y. Chen, Y. Ding, F. Zhao, E. Zhang, Z. Wu, and L. Shao, “Surface defect detection methods for industrial products: A review,” *Appl. Sci.*, vol. 11, no. 16, p. 7657, Aug. 2021.
- [34] G. Jocher, A. Chaurasia, and J. Qiu. (2023). *Ultralytics YOLO*. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [35] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

- [36] H. Li, J. Li, H. Wei, Z. Liu, Z. Zhan, and Q. Ren, "Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles," 2022, *arXiv:2206.02424*.
- [37] H. Huang, Z. Chen, Y. Zou, M. Lu, and C. Chen, "Channel prior convolutional attention for medical image segmentation," 2023, *arXiv:2306.05196*.
- [38] Z. Tong, Y. Chen, Z. Xu, and R. Yu, "Wise-IoU: Bounding box regression loss with dynamic focusing mechanism," 2023, *arXiv:2301.10051*.
- [39] J. Wang, C. Xu, W. Yang, and L. Yu, "A normalized Gaussian Wasserstein distance for tiny object detection," 2021, *arXiv:2110.13389*.



**ZIYI YANG** received the bachelor's degree from Tianjin University of Technology, in 2022. Her current research interests include machine vision and pattern recognition and intelligent systems.



**QIYU ZHANG** was born in 1999. He is currently pursuing the master's degree from North China University of Science and Technology. His current research interests include machine vision and algorithm research.



**JINGTAO YIN** is currently pursuing the Ph.D. degree. He is a Professor. His current research interest includes industrial automation.



**LIPING LIU** is currently pursuing the Ph.D. degree. She is a Professor. Her main research interests include machine vision, pattern recognition and intelligent systems, and mining engineering.



**ZHIZHONG JING** is currently with Hebei Sheng Aosong Wood Industry Company Ltd.

...