IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 December 2023, accepted 6 January 2024, date of publication 10 January 2024, date of current version 23 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3352445

METHODS

CCG-YOLOv7: A Wood Defect Detection Model for Small Targets Using Improved YOLOv7

WENQI CUI, ZHENYE LI, ANNING DUANMU, SHENG XUE, YIREN GUO, CHAO NI, TINGTING ZHU, AND YAJUN ZHANG

College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, Jiangsu 210037, China

Corresponding author: Yajun Zhang (zhangyj@njfu.edu.cn)

This work was supported in part by the National Science Foundation of China under Grant 62203223, Grant 62006120, and Grant 31570714.

ABSTRACT The Chinese furniture market has a high demand for wood floors. Manual defect detection in wood floors is inefficient and lacks stability in accuracy. It is necessary to conduct research on automatic defect detection in wood floors. To improve the accuracy of detecting small defects in wood floors, this paper proposed a new network based on YOLOv7. The new network is called the cascade center of gravity YOLOv7 (CCG-YOLOv7). This paper designed cascade efficient layer aggregation networks (C-ELAN), streamlined the CBS, replaced the ELAN with the C-ELAN, introduced the rapid supervised attention module to connect the backbone and head layers, and simplified the head layer of the YOLOv7 network. These methods improved the detection accuracy and speed for detecting small defects on wood floor surfaces. The improved network can effectively detect small defects on the wooden board surfaces, including knots, scratches, and mildew. Compared to the original YOLOv7, CCG-YOLOv7 improves precision, recall, and mean average precision by 2.1%, 1.6%, and 1.2%, respectively.

INDEX TERMS Deep learning, small target, wood floor defect detection, YOLOv7.

I. INTRODUCTION

Wood floor is a premium and environmentally-friendly choice for home decoration. In recent year, the demand for wood floor in furniture market has been steadily growing. Due to the characteristics of the raw materials, wood floor inevitably has defects such as knots, scratches, and mildews during the production process, as shown in Figure 1. These defects affect the aesthetic appeal and product quality of wood floor. For businesses, detecting these defects early can avoid the production of substandard goods, save production costs, and enhance competitiveness. For consumers, it can improve product quality and lifespan, thereby increasing consumer satisfaction. These defects have small and random shapes, and they are distributed irregularly on the surface of the wood floor, without any regular pattern. Therefore, accurate identification of these defects is crucial in the wood floor production process.

FIGURE 1. Images of solid wood floor defects: (a) scratch, (b) live-knot and dead-knot, (c) mildew.

Given the nature of wood floor as a commodity, non-destructive testing methods are commonly used for defect detection. Traditional non-destructive testing methods for wood floor defect detection include manual inspection, electrical testing [1], [2], laser scanning [3], [4], and ultrasonic testing [5], [6]. However, traditional non-destructive testing methods are susceptible to environmental influences, expensive in terms of equipment, and require well-trained operators [7]. More importantly, the results of the detection are often insufficient to meet industrial requirements.

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

With the development of computer technology, more and more researchers are applying computer vision and deep learning methods [8], [9], [10] to wood floor defects detection. Hashim et al. [11] conducted a visual exploratory analysis on the defects of lauan wood species, and the results showed that they could significantly distinguish different defect categories. He et al. [12] designed a hybrid total convolution neural network (Mix-FCN) for wood defect recognition and localization and achieved an overall accuracy of 99.13%. Hu et al. [13] used the Mask R-CNN algorithm in wood defect detection. The accuracy of defect identification for knots, dead knots, and insect holes was 99.05%, 97.05%, and 99.10%, respectively. Shi et al. [14] combined the Mask R-CNN algorithm with the Glance network and achieved a 98.70% overall classification accuracy and a 95.31% mean average precision. Mohammad et al. [15] applied Principal Component Analysis (PCA) technique in wood defect detection, corresponding modeling and development process was conducted for the PCA procedure. Silvén et al. [16] implemented unsupervised clustering methods to detect and identify wood defects. The achieved false detection and error escape rates are low. Yang et al. [17] proposed a method for detecting surface defects of solid wood panels based on a Sigle Shot MultiBox Detector algorithm. The average detection accuracy of the defects was 89.7%, and the average detection time was 90 ms. Wu et al. [18] proposed a defect detection method based on affinity propagation clustering, effectively improving the clustering speed with an accuracy of 87.68%. Qi et al. [19] proposed a fast wood defect detection method based on the BP neural network, achieving an accuracy of 90%. With the rise of YOLO network, some researchers have also applied YOLO network to surface defect detection of wood floor. Tu et al. [20] proposed a surface defects detection method based on improved YOLOv3 for sawn lumbers. Urbonas et al. [21] employ a pre-trained ResNet152 neural network model combined with faster R-CNN to detect surface defects on wooden boards, achieving an accuracy of 96.1%. Lim et al. [22] proposed a lightweight object detection model based on the YOLOv4-Tiny architecture for the detection of four types of wood defects. The model led to better accuracy and inference speed. Yang et al. [23] combined the deep learning feature extraction method and extreme learning machine classification method to establish a depth extreme learning machine model for wood image defect detection. The wood defect recognition accuracy reached 96.72%. These networks all performed excellently on their respective datasets. However, the detection of small target defects on the surface of wood floor did not perform well.

In this paper, we propose Cascade Center of Gravity YOLOv7 (CCG-YOLO) based on the latest YOLOv7 model to improve the accuracy of wood floor defect detection to address the issue of YOLO networks being less sensitive to small objects. Building upon the existing YOLO network, we have streamlined the CBS, the feature extraction module of the YOLOv7 backbone, leading the network to focus more on shallow features and smaller targets. We proposed Center Efficient Layer Aggregation Networks (C-ELAN) to improve the performance of original ELAN module in the backbone by introducing an algorithm named Cascade Center of Gravity Batch Norm (CCG-BN). CCG-BN is designed to normalize the centroid of features within the same batch.

The wood floor images used in this study were captured at Dehua TB New Decoration Material Co., Ltd. The images were captured using an LT-400CL 3CMOS RGB line scan industrial camera. Compared to area scan cameras, line scan cameras are not only cost-effective but also capable of high-speed image capture for moving objects, making them suitable for industrial production lines. Based on the different causes of wood floor defects, we classify the defects into two categories: growth defects and processing defects. Generally, growth defects can be further categorized as dead knots and live knots, while processing defects include scratches and mildews. A live knot is a knot in which the surrounding wood structure has intact grain continuity and it has not separated from the surrounding wood. On the other hand, a dead knot is a knot where the wood tissue has detached from the surrounding wood. A scratch is a defect that occurs on the surface of a wood floor during production or transportation. Unlike cracks, scratches are limited to the surface of the wood floor and do not result in a complete fracture of the board. The size of these defects ranges from 40×30 to 60×130 pixels. The defects occupy a relatively small proportion compared to the entire image. The position of defects in the images appears randomly. The defects exhibit different shapes. Figure 2 shows examples of dead knots, live knots, scratches, and mildews.

FIGURE 2. Common defects of wood floor such as dead-knots, live-knots, scratches and mildews.

II. METHODOLOGY

The proposed Cascade Center of Gravity YOLOv7 (CCG-YOLOv7) is presented in Figure 3. The 3-channel color images of solid wood floor are divided into squares with the original height as the edge length. These squares are then proportionally scaled to form images of size 640×640. It is important to note that this operation does not result in any cutting or cropping of edge defects. Then, 3 pairs of MP1 and C-ELAN modules are sequentially concatenated, forming the backbone of the proposed model. It should be noted that the MP1 module has the same structure as in the original YOLOv7 to half the size and double channel counts of feature maps while extracting information at different depth. The major difference lies on the design of the C-ELAN module combining with the Rapid Supervised Attention Module

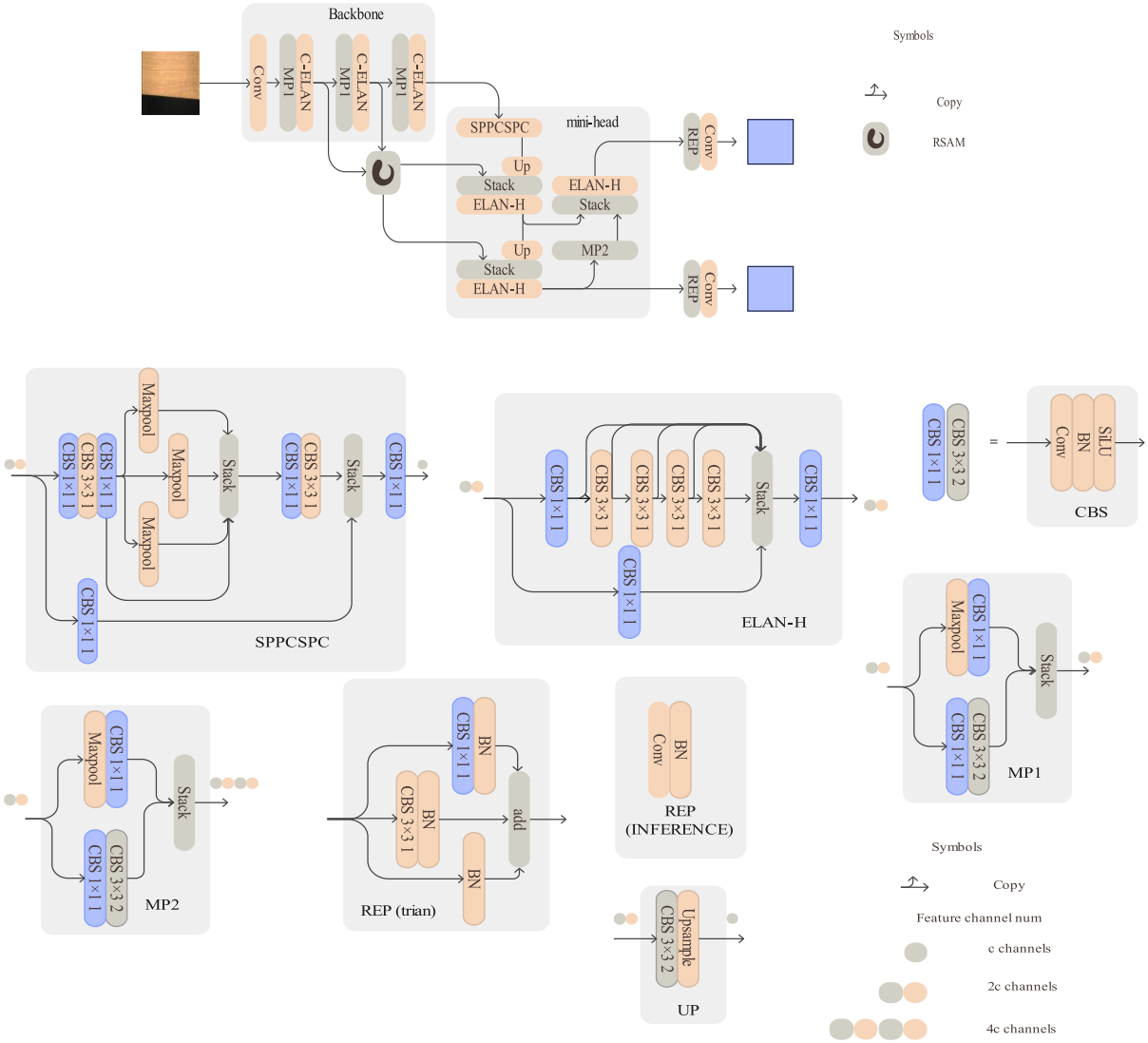


FIGURE 3. The structure of CCG-YOLOv7.

(RSAM) that transmits features from paths of the backbone at multiple scales.

We improved and simplified the head module of the original YOLOv7, replacing it with outputs only containing 40 × 40 and 80 × 80. Therefore, the novel head we designed is named Mini-Head. Coordinating with features from RSAM, the succinct structure in Mini-Head focuses on small object detection including dead-knots, live-knots, scratches, and mildews. Below, we will provide a detailed explanation of the modifications made to the backbone and head layers in YOLOv7 network.

In Figure 3, the Conv denotes the convolution layer, the BN denotes the batch normalization layer, and Silu denotes the activation function. The CBS module consists of a Conv layer, a BN layer, and a Silu layer, identical to the original YOLOv7. In Figure 3, "CBS 1 × 1.1" denotes a 1 × 1 convolution with a stride of 1. Similarly, "CBS 3 × 3.2" indicates a 3 × 3 convolution with a stride of 2. The Rep module is divided into train and inference modules. The inference module consists of a 3 × 3 convolution with a stride of 1.

A. BACKBONE

In the original network, with 4 CBS modules stacked at the beginning of the backbone, the subsequent ELAN and MP1 modules take in deep features that are not suitable for defects in solid wood flooring, whose background is wood grain with a relatively simple texture [24]. Besides, many defects on the solid wood flooring are of small area, such as dead-knots and live-knots, leading to a poor recall score when extracted by a large receptive field from the deep layers. So, we simplified the network architecture of this part and thus the feature paths to head are moved to a shallower position. This enables the network to focus more on shallow features and smaller

targets, thereby improving the efficiency of training and the accuracy of defect detection.

Figure 4 and Figure 5 display the overall distribution of defects in the dataset used in this study. The calculation formulas for y, $width$,and $height$ in the Figure 4 are as follows:

$$x = \frac{x_d}{w} \tag{1}$$

$$y = \frac{y_d}{h} \tag{2}$$

$$width = \frac{width_d}{w} \tag{3}$$

$$height = \frac{height_d}{h} \tag{4}$$

where, w and h represent the width and height of the image. x_d represents the coordinate of the defect in the width direction, while y_d represents the coordinate of the defect in the height direction. $width_d$ represents the width of the defect, and $height_d$ represents the height of the defect. Figure 4 displays the overall spatial distribution of defects in the dataset. The heat map uses color intensity to represent the density of defects in different locations. Darker colors indicate a higher concentration of defects in that area. Figure 5 shows the distribution of defect sizes. It can be observed from Figure 5 that the defects on the surface of the wooden floor are very small in size. As can be seen in Figure 4, in the dataset used in this study, the distribution of wood floor defects is random across the entire image, and the proportion of defective pixels in the entire image is relatively small. It should be noted that the coordinates in the heatmaps of Figure 4 and Figure 5 are relative coordinates.

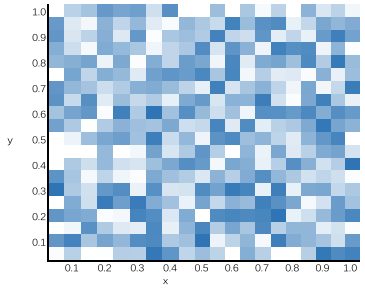


FIGURE 4. Heatmap of defect location distribution.

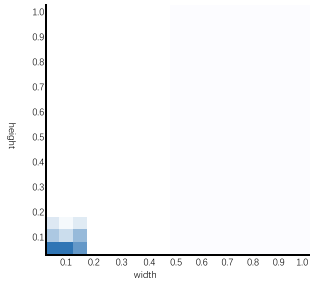


FIGURE 5. Heatmap of defect size statistics.

The defection distribution of solid wood floor is quite different from that of objects in public datasets, lettuce pallets Image Dataset. In contrast, the targets in general datasets like ImageNet and COCO are typically located at the center of the image, showing characteristics of a normal distribution. Secondly, the wooden floor defect dataset used marks targets like bad spots and scratches, which have fairly fixed shape characteristics. While, YOLOv7 network uses predefined anchor boxes to detect objects of different sizes. Statistical analysis reveals that the aspect ratio of the bounding boxes is either close to a 1:1 square or a long and narrow shape greater than 1:5. Therefore, to leverage the YOLOv7 framework designed for general object detection, it is necessary to transform the wooden floor defect data, making it possess adjustable bounding box distribution and shape characteristics.

1) CCG-BN

We proposed Center Efficient Layer Aggregation Networks (C-ELAN) to improve the performance of original ELAN module in the backbone by introducing an algorithm named Cascade Center of Gravity Batch Norm (CCG-BN). CCG-BN is designed to normalize the centroid of features within the same batch. The centroid of the grayscale image is usually defined as follows:

$$x_c = \frac{\sum_{i=1}^{w}\left(i\sum_{j=1}^{h}\mathbf{F}_{in,i,j}\right)}{\sum_{i=1}^{w}\sum_{j=1}^{h}\mathbf{F}_{in,i,j}} \tag{5}$$

$$y_c = \frac{\sum_{j=1}^{h}\left(j\sum_{i=1}^{w}\mathbf{F}_{in,i,j}\right)}{\sum_{j=1}^{h}\sum_{i=1}^{w}\mathbf{F}_{in,i,j}} \tag{6}$$

where, $\mathbf{F}_{in,i,j}$ is the grayscale value of the pixel at coordinates (i, j) in the image, in which i is the coordinate in the width direction, and j is the coordinate in the height. x_c and y_c are the coordinates of the gravity of center along the axes of width and height of the image. For color images, we succinctly define the centroid as a linear combination of the centroid coordinates in the 3 channels. We vectorize the original centroid formula for the image to facilitate parallel acceleration of computations:

$$\mathbf{F}_{flattened}^{s\times(3s)} = \text{flatten}^3(\mathbf{F}_{in}^{s\times s\times 3}) \tag{7}$$

$$\mathbf{Rn}^{2\times 3} = \mathbf{Rm}^{2\times s}\times\mathbf{F}_{flattened}^{s\times(3s)} \odot \text{flatten}^{2,4}\left(\begin{bmatrix}\mathbf{Rm}_1^{1\times s}\\ \mathbf{Rm}_2^{1\times s}\end{bmatrix}\times\mathbf{S}^{1\times 3}\right) \tag{8}$$

$$\begin{bmatrix}x_c\\ y_c\end{bmatrix} = \mathbf{Rn}^{2\times 3}\times\mathbf{w}^{3\times 1} \tag{9}$$

where, $\mathbf{S}^{1\times 3} = \begin{bmatrix}\frac{1}{\Sigma r} & \frac{1}{\Sigma g} & \frac{1}{\Sigma b}\end{bmatrix}$, $\mathbf{Rm}^{2\times s} = \begin{bmatrix}1 & 1 & \cdots & 1\\ 1 & 2 & \cdots & s\end{bmatrix}$, $\mathbf{Rm}_1^{1\times s}$ is a row vector consisting of all elements in the first

row of matrix. $\mathbf{Rm}^{2\times s}$ and $\mathbf{Rm}_2^{1\times s}$ is a row vector consisting of all elements in the second row of matrix $\mathbf{Rm}^{2\times s}$. $\mathbf{w}^{3\times 1}$ is a learnable parameter used for linear transformation of the centroid coordinates in each channel, resulting in the computation of the centroid position of the color image. The matrix $\mathbf{F}_{in}^{s\times s\times 3}$ is a three-channel color image for which the centroid is to be computed. The function flatten[3] represents the flattening of the third channel dimension, resulting in a two-dimensional matrix that contains only the first and second channels. The function flatten[2,4] represents the flattening of the second channel and the fourth channel dimension, resulting in a two-dimensional matrix where the second channel is merged into the first channel of the matrix, and the fourth channel is merged into the third channel. Based on the above analysis, we have designed a method to shift the centroid of the image by transforming pixel values. This method includes learnable parameters that can specify the position of the image centroid after the shift, allowing the image features to automatically adjust during the training process as needed by the network. The algorithm for centroid shift is as follows:

$$\mathbf{F}_{out,i} = \mathbf{F}_{in,i,j} + \frac{\left(x_c - x_c'\right)\sum_{i=1}^{w}\sum_{j=1}^{h}\mathbf{F}_{in,i,j}}{hwx_c' - \frac{1}{2}h} \tag{10}$$

$$\mathbf{F}_{out,j} = \mathbf{F}_{in,i,j} + \frac{\left(y_c - y_c'\right)\sum_{i=1}^{w}\sum_{j=1}^{h}\mathbf{F}_{in,i,j}}{hwy_c' - \frac{1}{2}w} \tag{11}$$

$$\mathbf{F}_{out} = \mathbf{F}_{out,i} + \mathbf{F}_{out,j} \tag{12}$$

where, $\mathbf{F}_{out,i}$ represents the feature after centroid normalization in the width direction, $\mathbf{F}_{out,j}$ is the feature after centroid normalization in the height direction, and $\mathbf{F}_{out}$ is the output of the module. x_c' and y_c' are learnable parameter vectors, each containing scalars equal to the number of input feature channels.

2) C-ELAN

In the YOLOv7 network, the Enhanced Local Attention (ELAN) layer is a special attention mechanism used to enhance the network's perception of local regions. The ELAN layer adjusts the weights of feature maps adaptively, increasing the focus on regions of interest and improving the accuracy of object detection. To improve the detection accuracy of small target defects in the YOLO network, it is necessary to enhance the depth of the ELAN layer. Therefore, we propose the Center Efficient Layer Aggregation Networks (C-ELAN) layer.

The C-ELAN is designed as a feature exactor that bring the output to larger receptive field and deeper-level features. The C-ELAN module is presented in Figure 6. As shown in Figure 6, the C-ELAN module consists of 13 convolutional layers, 2 SiLU activation layers, 2 ReLU activation layers, 2 CCG-BN layers, and 2 stacking computation along the axis

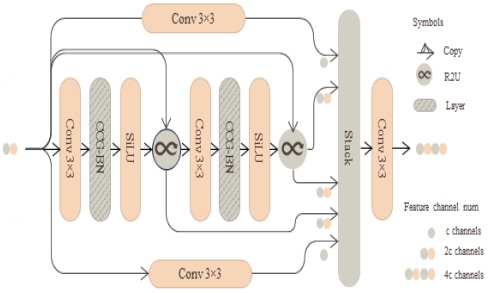


FIGURE 6. Structure of C-ELAN.

of the channels. The number of feature channels is doubled when it flows to the output layer.

The input 3-dimensional feature flows through a convolutional layer with $3 \times 3 \times c$ kernel size, which keeps the size and channels unchanged. Then, it is normalized by CCG-BN and SiLU activation function. In the C-ELAN module, R2Us are used to route gradients to 2 different paths: the forward path and the residual path. Adjusting the gradient propagation path enhances the effective utilization of network parameters, and enables different computational units to learn diverse information, thereby achieving higher parameter utilization efficiency. Then, the feature map is fed to an identical structure, and the new output is aggregated with the residual output of the previous R2U module to form a residual feature with twice the number of channels.

Simultaneously, another 2 convolution layers with kernel size of $3 \times 3 \times (c/2)$ and paddings of 1 perform channel compression on the input feature. This operation generates 2 independent outputs with a half channel number. The presence of the above-mentioned strategy ensures that the learning capability of C-ELAN is stable enough to avoid degradation during training. The reason is that the gradient directly determines and propagates information to update the weights in each branch. Finally, the features from the 3 branches and the residual-aggregated features from each R2U module are stacked together to form a feature map with four times the number of channels. Then, a 3×3 convolutional layer with a stride of 2 is applied to obtain an output with a reduced spatial size compared to the input features, while doubling the number of channels.

B. HEAD

Concatenation and Bottleneck Structure (CBS) layer plays a role in feature fusion and dimension adaptation between the backbone and head layers. The CBS layer first concatenates the feature maps from the backbone layer and the head layer along the channel dimension to achieve feature fusion across different scales. Then, it reduces the number of channels in the feature maps using a bottleneck structure, which helps to reduce the computational complexity. To improve the detection speed of YOLOv7 network for small target defects, it is necessary to introduce channel attention mechanism to adaptively adjust the weights of feature map channels. Therefore,

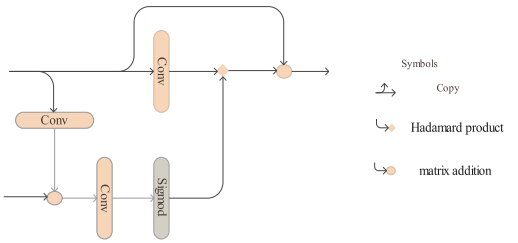


FIGURE 7. Structure of SAM.

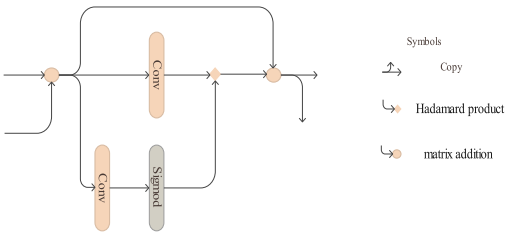


FIGURE 8. Structure of RSAM.

we replaced the CBS module with the Rapid Supervised Attention Module (RSAM) to achieve this.

Rapid Supervised Attention Module is designed to transfer features from the backbone layer to the mini-head layer at multiple scales. Rapid Supervised Attention Module (RSAM) improves upon the Multi-Stage Progressive Image Restoration (MRPNet), proposed by Zamir et al. [25]. The Supervised Attention Module (SAM) used in MRPNet to pass features to the next stage includes the Locally Supervised Predictions (LSP) structure for generating attention maps, appropriately suppressing features with smaller amounts of information in the current stage. The SAM structure is shown in Figure 7 The input is the image input feature $\mathbf{F}_{SAM,in}^{h\times w\times c}$ from the previous stage, where h and w are its height and width, and c is the number of channels. At the same time, SAM takes a low-quality (Degraded) image input $\mathbf{F}_{degarded}^{h\times w\times 3}$, which, when added to the convoluted $\mathbf{F}_{SAM,in}^{h\times w\times c}$, computes the repaired image and calculates the loss function. To introduce features into the next stage, a 1×1 pointwise 2D convolution is used to transform the repaired image, and an image attention map is calculated using the Sigmoid function. This is then multiplied by $\mathbf{F}_{SAM,in}^{h\times w\times c}$ to apply the attention map, and finally, the residual structure of $\mathbf{F}_{SAM,in}^{h\times w\times c}$ is adopted to form the LSP structure, outputting the image output feature $\mathbf{F}_{SAM,out}^{h\times w\times c}$ required for the next stage. The structure of RSAM is shown in Figure 8. Compared to SAM, RSAM replaces the image feature $\mathbf{F}_{RSAM,in,1}^{h\times w\times c}$, therefore, it directly adds the input feature matrix $\mathbf{F}_{RSAM,in,2}^{h\times w\times 3}$ to reduce one convolution that reduces the number of channels. The size of the convolution kernel on the path from $\mathbf{F}_{RSAM,in}^{h\times w\times 3}$ to $\mathbf{F}_{SAM,out}^{h\times w\times c}$ is 1, with both stride and padding set to 1, so the size and channel count of the output from this convolution layer are consistent with the input. Another convolution layer on the lower path in the figure has the same parameters as the aforementioned convolution, and the size of its output feature also satisfies $h\times w\times 3$. Since the input feature $\mathbf{F}_{RSAM,in}^{h\times w\times 3}$ undergoes one less convolution, if the output features $\mathbf{F}_{RSAM,out}^{h\times w\times 3}$ is output directly after the first sum, there would be no learnable parameters in the RSAM structure. Hence, $\mathbf{F}_{RSAM,out}^{h\times w\times 3}$ is used directly as the output features. The parameter count of RSAM is approximately 2.5% of the SAM with an 80-channel input using the same convolution kernel size, which is why it is called the Rapid Supervised Attention Module.

In the YOLOv7 network, the role of the head layer is to process the feature maps extracted from the backbone layer and output the final detection results. The head layer consists of a series of convolutional layers and fully connected layers, responsible for transforming the feature maps into predicted bounding boxes, class probabilities, and confidences. To make it focus more on shallow-level features and smaller objects, we need to decrease the depth of the head layer.

We simplified the head layer to the mini-head layer, allowing it to focus more on shallow features and smaller targets, thereby improving the efficiency of training and the accuracy of defect detection. As shown in Figure 9, the grid of 20×20 is deleted from the original head, making it less complex and more effective. The output of the head is refactored to 2 different features with size of $80\times 80\times 9\times 3$ and $40\times 40\times 9\times 3$, to speed up the inference speed of the model. Each of the features consists of information of the detected defects, including the grid shape, 3 bounding boxes (height, width, and center coordinates), defect type, and confidence score. With the elimination of the 20×20 output and its associated output structure from the original YOLOv7, the parameter count in the head is reduced by 72% compared to its original value. This optimization has significantly improved the efficiency of our modified YOLOv7 network for the task of defect detection in solid wood floor.

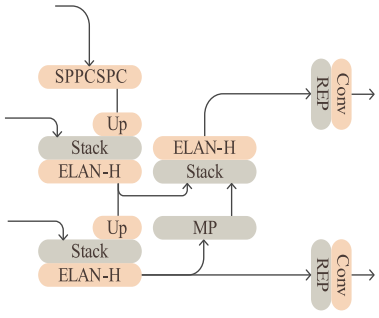


FIGURE 9. Structure of mini-head.

III. RESULT & DISCUSSION

A. TRAINING STRATEGIES

During the training process of the CCG-YOLOv7 network, we utilized the data augmentation technique of copy-pasting a small object into various positions in an image to create new annotated samples, which helps improve the model's ability

to detect small defects and does not result in an increase in the number of training samples. The pasted small object can undergo scaling, flipping, rotation, and other random transformations. This method increases the contribution of small objects to the loss calculation during training by increasing the number of small objects in each image and matching anchor boxes. To enhance the robustness and generalization ability of the model, we adopt this data augmentation technique during the model training process. The training hyperparameters in the experiment are shown in Table1.

TABLE 1. The training hyperparameters in the experiment.

Hyperparameters	Value
Epoch	300
Initial learning rate	0.01
Batch size	8
Momentum	0.913
Wight decay	0.0005
Loss function	BCE, CIoU

The dataset used in this study consists of 3473 images. Among these, there are 698 images that contain only the dead-knots defect, 752 images that contain only the live-knots defect, 623 images that contain only the scratches defect, and 522 images that contain only the mildews defect, and 694 images do not contain any defects as negative samples. Additionally, there are 184 images that contain multiple types of defects.

To build and test the proposed defect detection model, the collected dataset of 3473 wood floor images were divided into three parts: training set, testing set, and validation set, in a ratio of 3:1:1, as shown in Table 2. The training set is used to train the constructed network, the testing set is used to evaluate the network, and the validation set is used for final assessment.

TABLE 2. Dataset split for training, validation, and testing.

Defects Classes	Training Set	Validation Set	Test Set
Dead-knots	418	140	140
Live-knots	450	151	151
Scratches	373	125	125
Mildews	314	104	104
Multiple defects	110	37	37
Negative samples	416	139	139

B. EXPERIMENT OF SOME METHODS APPLYING TO YOLOV7

To build and test the proposed defect detection model, the collected dataset of 3473 wood floor images were divided into three parts: training set, testing set, and validation set, in a ratio of 3:1:1. The training set is used to train the constructed network, the testing set is used to evaluate the network, and the validation set is used for final assessment. The training platform specifications are presented in Table 3.

TABLE 3. Hardware and software parameters of the experimental environment.

Name	Parameter
Memory	32.00 GB
CPU	Intel Core i7-8700 CPU @3.2GHz
Graphics card	NVIDLA GeForce RTX 3080 Ti
System	Windows 10
Environment Configuration	Python3.6, TensorFlow-GPU 1.14.0, Pytorch

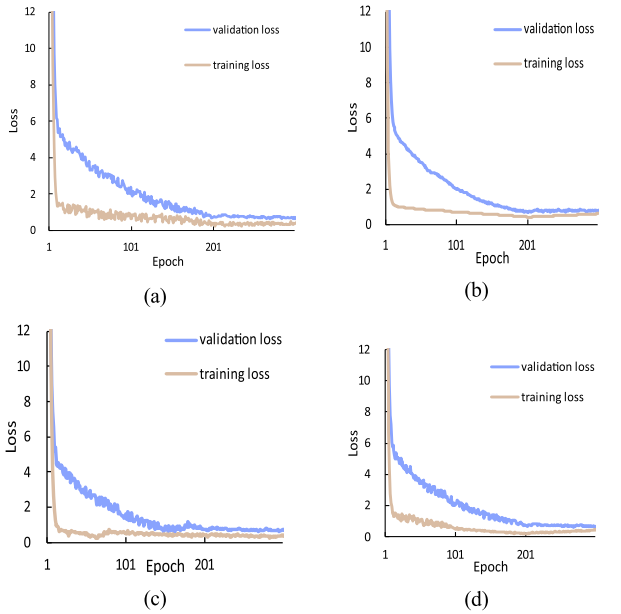


FIGURE 10. The training process of the different methods introduced into YOLOv7(a) YOLO-RSAM result, (b) YOLO-C-ELAN result, (c) YOLO-NC result, (d) YOLO-mini-head.

In order to improve the detection performance of the YOLOv7 model, several methods that may affect the detection performance were employed. A comparison was made between these methods and the original YOLOv7 model to observe whether these methods had a positive effect on enhancing the detection performance of YOLOv7. Figure 10 displays the variation curves of the network training loss and validation loss after training and testing under the same conditions.

In Figure. 10, YOLO-RSAM represents the independent integration of the RSAM module into the YOLO network. YOLO-C-ELAN indicates the replacement of all ELAN modules in the original YOLO network with C-ELAN modules. YOLO-NC signifies the simplification of the backbone layer by removing four CBS modules. YOLO-mini-head denotes

TABLE 4. The comparison of classical algorithms.

Methods	$P(\%)$	$R(\%)$	$mAP(\%)$
SSD	75.6	76.4	77.5
Faster R-CNN	87.6	87.7	87.9
YOLOv5	89.3	90.7	90.9
YOLOv6	89.7	91.4	91.5
YOLOv7	90.6	93.2	93.6

the replacement of the head layer in the original YOLO network with the mini-head. Also, Precision(P), Recall(R), and mean average precision (mAP) are selected as the main evaluation indexes which are defined as follows:

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{14}$$

$$mean\ Average\ Precision = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{15}$$

where, TP refers to the number of samples where positive samples are correctly classified; FP means the number of negative samples which are correctly classified as negative, and FN indicates the number of positive samples which are incorrectly classified as negative. AP is the area abounded by the $Precision - \mathrm{Re}\,call$ curve which represents the detection accuracy of one kind of wood floor defects. *mean Average Precision*(mAP) refers to the overall detection accuracy of all defects categories. As shown in Figure 8, during the training of 200 epochs, the loss curves of all the aforementioned models on the validation set and training set become relatively flat, indicating that the model training has converged.

We conducted comparative experiments with some classic detection networks. From Table 4, it can be observed that compared to other classical algorithms, the YOLOv7 demonstrates superior performance in wood floor defect detection.

C. ABLATION EXPERIMENT

We conducted an ablation experiment to investigate whether the fusion of these modules could further enhance the performance of the YOLOv7 network. Using the original YOLOv7 network as the baseline, we introduced RSAM, mini-head, and C-ELAN in separate experiments. The results of this experiment are presented in Table 5.

In the table 5, T represents the average detection time for 100 images, measured in milliseconds. From the table, it can be observed that the detection accuracy improves when NC and C-ELAN modules are individually integrated into the network. Additionally, the detection speed significantly increases when the RSAM module is integrated with the mini-head into the network. However, the introduction of mini-head and NC results in a decrease of 0.1% in P, 3.3% in R, and 1.4% in mAP. The introduction of mini-head and C-ELAN results in a decrease of 1.3% in P, 1.5% in R, and 1.2% in mAP. The mini-head is composed of a series of convolutional and fully connected layers, responsible for transforming the feature map into predicted bounding boxes, class probabilities, and confidence scores. In comparison to the original head layer, the mini-head layer eliminates the 20×20 grid, which includes detected defect information. Therefore, when the backbone focuses too much on either shallow or deep image features, the network's detection accuracy decreases. With the integration of C-ELAN, NC, and the mini-head into the network, both the detection accuracy and speed have been improved. Further introducing RSAM resulted an improvement of 2.1% in P, 1.6% in R, and 1.2% in mAP. And detection speed of the network improves to 19ms. This is because the RSAM module adaptively adjusts the weights of the feature map channels, connecting the feature maps between the backbone and the mini-head layer, achieving feature fusion across different scales. Therefore, we choose to apply RSAM, C-ELAN, mini-head, and NC to the YOLO network based on these results. The figure 11 shows the curves of YOLOv6, YOLOv7 and CCG-YOLOv7.

D. EXPERIMENT UNDER DIFFERENT LIGHT CONDITION

In industrial production processes, continuous and prolonged operation can cause a decrease in the intensity of the light source. We defined the illuminance range of 4600 lux - 6100 lux as strong lighting, 4400 lux - 4600 lux as moderate lighting, and 3300 lux - 4400 lux as weak lighting. The numbers of images captured under strong, weak, and moderate lighting conditions were 98, 103, and 112, respectively. These images were utilized to evaluate the detection performance of the CCG-YOLOv7 network under diverse lighting conditions. These images, taken under different light intensities, feature defects like knots, scratches, and mildews, with only a few defects presented in each image. The experiments were conducted using the images under different lighting conditions. The experimental results are presented in Table 7. Figure 11 showcases partial visual results of wood floor defect detection under different lighting conditions using the CCG-YOLOv7 model.

As shown in Table 6, when using images captured under strong lighting conditions for detection, the CCG-YOLOv7 model exhibited no decrease in P and mAP, with a slight decrease of 0.1% in R. However, when using images captured under weak lighting conditions for detection, the CCG-YOLOv7 model showed a decrease of 0.1% in P and mAP, with no decrease in R. From the results in Table 6, it can be observed that the proposed CCG-YOLOv7 model achieves high detection accuracy under different lighting conditions. Compared with the Faster R-CNN, the CCG-YOLOv7 model possesses sufficient robustness to detect wood floor defects under varying lighting conditions, enabling prolonged industrial production.

TABLE 5. Ablation experiment of RSAM, mini-head, NC and C-ELAN.

RSAM	C-ELAN	mini-head	NC	$P(\%)$	$R(\%)$	$mAP(\%)$	T(ms)
				90.6	93.2	93.6	38
			√	91.8	93.4	93.9	37
		√		90.7	93.5	93.7	25
	√			91.5	93.7	94.2	41
√				90.7	93.3	93.8	23
		√	√	90.5	89.9	91.2	25
	√	√		89.3	91.7	92.4	29
	√	√	√	90.8	93.7	94.1	29
√	√	√	√	**92.7**	**94.8**	**94.8**	**19**

TABLE 6. The detection results of CCG-YOLOv7 and Faster R-CNN under different tilt angle conditions.

Base Model	Light Condition	$P(\%)$	$R(\%)$	$mAP(\%)$
CCG-YOLOv7	Weak	92.0	94.3	94.4
	Normal	92.1	94.3	94.5
	Strong	92.1	94.2	94.5
Faster R-CNN	Weak	86.9	86.9	87.1
	Normal	87.6	87.7	87.9
	Strong	87.1	87.3	87.4

TABLE 7. The detection results of CCG-YOLOv7 under different tilt angle conditions.

Base Model	Tilt Angle	$P(\%)$	$R(\%)$	$mAP(\%)$
CCG-YOLOv7	Small	92.1	94.3	94.4
	Normal	92.1	94.3	94.5
	Big	92.0	94.1	94.2

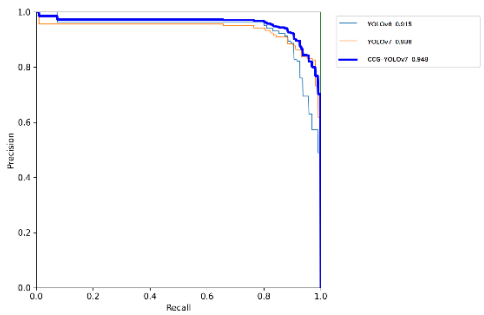
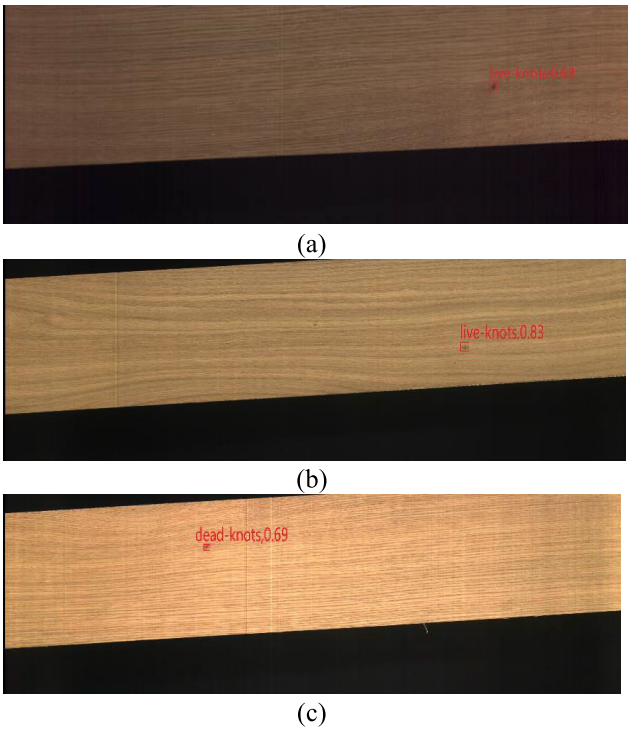


FIGURE 11. Curve diagram of the evaluation index.



(a) (b) (c)

FIGURE 12. The visualization results of detecting wood floor defects under different light conditions by using the CCG-YOLOv7 model:(a) weak light, (b) normal light, (c) strong light.

E. EXPERIMENT UNDER DIFFERENT TILT ANGLE

To validate the tolerance of CCG-YOLO to skewed images during the wood floor sorting process, we selected images with different tilt angles from the test set for testing. The numbers of images captured under small, big, and normal tilt angle conditions were 63, 87, and 78, respectively. These images were utilized to evaluate the detection performance of the CCG-YOLO network under different tilt angle conditions. These images, taken under different tilt angle, feature defects like knots, scratches, and mildews, with only a few defects presented in each image. Table 7 presents the experimental results for detecting skewed images.

As shown in Table 7, when using images with smaller tilt angles for detection, the CCG-YOLOv7 model exhibited no decrease in precision and recall, with a slight decrease of 0.1% in mAP. However, when using images with larger tilt angles, the RDEYOLOv7 model showed a decrease in precision, recall, and mAP by 0.1%, 0.2%, and 0.3% respectively. The experimental results indicate that the impact of image tilt angle on the CCG-YOLOv7 model is negligible and can be disregarded.

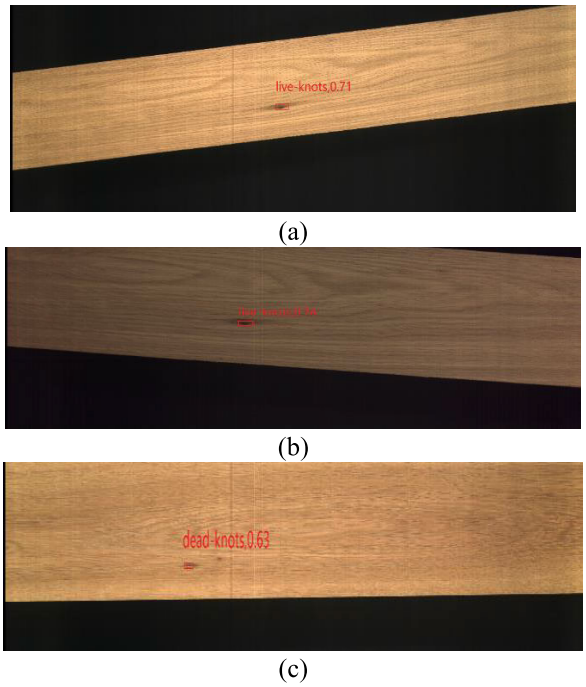


FIGURE 13. The detection results of images with different tilt angle by using the CCG-YOLOv7. Conditions: (a)big tilt angle, (b)small tilt angle, (c) normal tilt angle.

IV. CONCLUSION

This study proposes a wood floor defect detection method called CCG-YOLOv7, which can accurately detect small defects on the surface of wood floors. YOLOv7 introduces RSAM, mini-head, and C-ELAN to replace the CBS modules connecting the backbone and head layers, as well as the head layer and ELAN, resulting in improved detection performance. It has been demonstrated that CCG-YOLOv7 can increase P, R, and mAP by 2.1%, 1.6%, and 1.2%, respectively. CCG-YOLOv7 is constructed by incorporating C-ELAN, RSAM, mini-head, and removing certain CBS modules in YOLOv7. Through the analysis of experimental results, the following findings are obtained: Compared to the original YOLOv7, CCG-YOLOv7 improves P, R, and mAP by 2.1%, 1.6%, and 1.2%, respectively. CCG-YOLOv7 also demonstrates good performance in detecting wood floor surface defects under different lighting conditions, which is of significant importance in industrial production. Furthermore, CCG-YOLOv7 is capable of detecting surface defects on inclined images. The experiments confirm that the proposed CCG-YOLOv7 method is effective for wood floor defect detection.

REFERENCES

[1] A. Mohammadabadi and R. Dugnani, "Detection of wood defects using low acoustic impedance-based PZT transducers," *J. Indian Acad. Wood Sci.*, vol. 17, no. 2, pp. 107–113, Dec. 2020.

[2] M. Radwan, N. Becker, D. V. Thiel, and H. G. Espinosa, "In-line wood defect detection using simple scalar network analyzer," *Sensors*, vol. 22, no. 23, p. 9495, Dec. 2022.

[3] Z. Ai, Y. Lin, Z. Xu, Y. Zhu, and L. Wu, "Wood broken defect detection with 3D laser scanning," in *Proc. China Autom. Congr. (CAC)*, Nov. 2022, pp. 6513–6517.

[4] Y. Yang, Z. Chen, Y. Liu, Y. Li, Z. Hu, and B. Gou, "Detection system for U-shaped bellows convolution pitches based on a laser line scanner," *Sensors*, vol. 20, no. 4, p. 1057, Feb. 2020.

[5] U. Hilbers, J. Neuenschwander, J. Hasener, S. J. Sanabria, P. Niemz, and H. Thoemen, "Observation of interference effects in air-coupled ultrasonic inspection of wood-based panels," *Wood Sci. Technol.*, vol. 46, no. 5, pp. 979–990, Sep. 2012.

[6] M. Tiitta, V. Tiitta, M. Gaal, J. Heikkinen, R. Lappalainen, and L. Tomppo, "Air-coupled ultrasound detection of natural defects in wood using ferroelectret and piezoelectric sensors," *Wood Sci. Technol.*, vol. 54, no. 4, pp. 1051–1064, Jul. 2020.

[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] Y. Quan, Z. Li, C. Zhang, and H. Ma, "Object detection model based on deep dilated convolutional networks by fusing transfer learning," *IEEE Access*, vol. 7, pp. 178699–178709, 2019.

[9] W. Li, Z. Li, X. Yang, and H. Ma, "Causal-ViT: Robust vision transformer by causal intervention," *Eng. Appl. Artif. Intell.*, vol. 126, Nov. 2023, Art. no. 107123.

[10] S. Chen, Z. Li, and Z. Tang, "Relation R-CNN: A graph based relation-aware network for object detection," *IEEE Signal Process. Lett.*, vol. 27, pp. 1680–1684, 2020.

[11] U. R. Hashim, S. Z. M. Hashim, A. K. Muda, K. Kanchymalay, I. E. A. Jalil, A. N. Anuar, and M. H. Othman, "Extraction and exploratory analysis of texture features on images of timber defect," *Adv. Sci. Lett.*, vol. 24, no. 2, pp. 1104–1108, Feb. 2018.

[12] T. He, Y. Liu, Y. Yu, Q. Zhao, and Z. Hu, "Application of deep convolutional neural network on feature extraction and detection of wood defects," *Measurement*, vol. 152, Feb. 2020, Art. no. 107357.

[13] K. Hu, B. Wang, Y. Shen, J. Guan, and Y. Cai, "Defect identification method for poplar veneer based on progressive growing generated adversarial network and MASK R-CNN model," *BioResources*, vol. 15, no. 2, pp. 3041–3052, Mar. 2020.

[14] J. Shi, Z. Li, T. Zhu, D. Wang, and C. Ni, "Defect detection of industry wood veneer based on NAS and multi-channel mask R-CNN," *Sensors*, vol. 20, no. 16, p. 4398, Aug. 2020.

[15] M. M. Tafarroj, H. Kalani, M. Moavenian, and A. Ghanbarzadeh, "An application of principal component analysis method in wood defects identification," *J. Indian Acad. Wood Sci.*, vol. 11, no. 1, pp. 33–38, Jun. 2014.

[16] O. Silvén, M. Niskanen, and H. Kauppinen, "Wood inspection with non-supervised clustering," *Mach. Vis. Appl.*, vol. 13, nos. 5–6, pp. 275–285, Mar. 2003.

[17] 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.

[18] D.-Y. Wu and N. Ye, "Wood defect recognition based on affinity propagation clustering," in *Proc. Chin. Conf. Pattern Recognit. (CCPR)*, Oct. 2010, pp. 1–5.

[19] D. Qi, P. Zhang, and L. Yu, "Study on wood defect detection based on artificial neural network," in *Proc. IEEE Conf. Cybern. Intell. Syst.*, Sep. 2008, pp. 951–956.

[20] Y. Tu, Z. Ling, S. Guo, and H. Wen, "An accurate and real-time surface defects detection method for sawn lumber," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[21] A. Urbonas, V. Raudonis, R. Maskeliūnas, and R. Damaševičius, "Automated identification of wood veneer surface defects using faster region-based convolutional neural network with data augmentation and transfer learning," *Appl. Sci.*, vol. 9, no. 22, p. 4898, Nov. 2019.

[22] W.-H. Lim, M. B. Bonab, and K. H. Chua, "An optimized lightweight model for real-time wood defects detection based on YOLOv4-tiny," in *Proc. IEEE Int. Conf. Autom. Control Intell. Syst. (I2CACIS)*, Jun. 2022, pp. 186–191.

[23] Y. Yang, X. Zhou, Y. Liu, Z. Hu, and F. Ding, "Wood defect detection based on depth extreme learning machine," *Appl. Sci.*, vol. 10, no. 21, p. 7488, Oct. 2020.

[24] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.
[25] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14816–14826.

WENQI CUI received the B.S. degree from the College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, China, in 2020, where he is currently pursuing the M.S. degree. His research interests include the application of intelligent algorithms in industry, image data processing, and time series prediction.

ZHENYE LI was born in Yangzhou, Jiangsu, China, in 1997. He received the bachelor's degree in engineering from Nanjing Forestry University, in 2019, where he is currently pursuing the master's degree.

His research interests include the application of intelligent algorithms in industry, image data processing, and hyperspectral analysis. He was a recipient of the International Association of Geomagnetism and Aeronomy Young Scientist Award for Excellence, in 2008, and the IEEE Electromagnetic Compatibility Society Best Symposium Paper Award, in 2011.

ANNING DUANMU received the bachelor's degree in mechanical and electronic engineering from Nanjing Forestry University, where he is currently pursuing the master's degree in control science and engineering. His research interests include computer vision, deep learning, and image processing. In this study, he was responsible for proofreading the manuscript.

SHENG XUE was born in Huaian, Jiangsu, China, in 2000. He received the bachelor's degree in engineering from Nanjing Forestry University, in 2022, where he is currently pursuing the master's degree. His research interests include the application of intelligent algorithms in industry and image data processing.

YIREN GUO received the B.S. degree from the College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, China, in 2020, where he is currently pursuing the M.S. degree. His research interests include the application of intelligent algorithms in industry, image data processing, and time series prediction.

CHAO NI was born in Nanjing, Jiangsu, China, in 1979. He received the B.S. degree in automation from the Nanjing University of Science and Technology, Nanjing, in 2001, and the Ph.D. degree in control theory and control engineering from Southeast University, Nanjing, in 2008. From October 2017 to November 2018, he was a Visiting Scholar with the University of Maryland, College Park, MD, USA. He is currently an Associate Professor with the Automation Department, Nanjing Forestry University, China. His research interests include artificial intelligence in industrial application, data processing, and spectroscopy analysis.

TINGTING ZHU received the Ph.D. degree in pattern recognition and artificial intelligence from the School of Automation, Southeast University, in 2019. She was a Visiting Student with the Department of Atmospheric and Oceanic Sciences, McGill University, Canada, from 2017 to 2018. She is currently with the College of Mechanical and Electronic Engineering, Nanjing Forestry University, China. Her current research interests include machine learning, data processing and modeling, renewable energy generation forecast, and climate feedback. She achieved a fellowship jointly awarded by Fonds de Recherche du Quebec—Nature et Technologies (FRQNT) and the China Scholarship Council.

YAJUN ZHANG received the B.S. degree in electrical engineering from the Shanghai Institute of Technology, in 2014, the M.S. degree in control science and engineering from Donghua University, in 2017, and the Ph.D. degree in control science and engineering from Southeast University, in 2021. He is currently a Lecturer with the School of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, China. His research interests include production systems control, supply chain management, and deep learning.