APSIPA Transactions on Signal and Information Processing, 2024, 13, e27

Original Paper

Lightweight Plank Grading Based on a Multichannel Spatial Attention Mechanism

Ziqiang Li[1,2], Huixuan Wang[3], Na Lv[1,2], Yihong Guo[4] and Yan Zhang[5]*

[1] *Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan 250022, China*
[2] *School of Information Science and Engineering, University of Jinan, Jinan 250022, China*
[3] *China Mobile Group Shandong Co., Ltd., Jinan 250013, China*
[4] *Mstar Machinery Co., Ltd., Fei County 273400, China*
[5] *School of Science, Shandong Jianzhu University, Jinan 250101, China*

ABSTRACT

The grading and sorting process of the wood planks is a critical stage within the production line. However, in the real world, many factories still rely on humans to perform this task manually. This method is not only inefficient, but also time-consuming and labor-intensive. To solve this problem, a lightweight wood board image classification algorithm based on a multichannel spatial attention mechanism is proposed in this paper, which can be used for the real-time classification of wood planks on the production line. This method is used to classify freshly rotated cut wood planks based on defects such as damage and voids on the production line. Specifically, the received images of the wood planks were processed by a feature extraction module to effectively separate the interfering background from the foreground of the wood planks. After fusing the edge information map with the foreground image of the wood planks, a multichannel convolutional neural network with

*Co-corresponding author: Yan Zhang, Email: zyzx@sdjzu.edu.cn

Received 26 July 2024; revised 28 August 2024; accepted 30 August 2024
ISSN 2048-7703; DOI 10.1561/116.20240050

spatial and channel attention ability was used to learn the features for correctly grading the wood planks. Experimental results show that the proposed method is superior to traditional methods and some existing deep learning algorithms in terms of performance and benefits.

Keywords: Wood plank image grading, real-time classification, feature fusion, lightweight neural network, convolutional neural network

1 Introduction

Defect detection in wood plank production is an essential process. It is usually done manually on the assembly line, which is not only time-consuming and labor-intensive but also subjective. The conveyor speed of the plank rotary cutting production line is up to 120m/min; even the most skilled workers, facing such a fast speed, still need to maintain a high level of mental concentration at all times. This can readily lead to misdetections and missed detections as a result of visual fatigue. The traditional non-destructive automatic wood inspection technology mainly uses ultrasonic [19, 11], laser [6], ray [14, 12] near-infrared spectroscopy [20] and other different means of detection to obtain wood data information (such as images, signals, point clouds, etc.), and these different forms of data, the corresponding method of processing, to obtain the stability of the characteristics of wood. Such as image data preprocessing, feature segmentation, contour detection, and other operations to obtain the image feature parameters; the component characteristics of the signal are obtained by preprocessing, artificial feature extraction, and data reconstruction of the signal band. Finally, the obtained features are evaluated by the classification algorithm, and the results are output. However, the above methods are too expensive and have the disadvantages of slow speed, limited detection capability, poor adaptation to vibration, and inability to scan continuously. In recent years, there have been new research advances in computer vision technology, and one of its basic tasks is image classification. This technique can categorize the input images into different classes, providing new ideas and methods for defect detection. Currently, the use of digital image processing techniques to detect defects in wood has become a popular approach, but these techniques are limited to specific working environments, making them inadequate for the requirements of assembly line production. Therefore, there is an urgent need to develop robust defect detection methods that meet the requirements of the production line.

Classification of wood planks in a plank-rotation assembly line scenario presents the following challenges:

- The production line scene is different from the general detection site, which needs to be cut by the machine and transported by the conveyor belt. Wood chips from rotary cutting and debris in the scene may appear on the inspection screen, which can affect the inspection process.
- When integrity is detected, the color texture in the bark will interfere with classification, and the general algorithm cannot fully learn the location and size of the damaged or empty distribution in the whole bark.
- In a production environment, the system needs to run in real time, and hardware resources may be limited. It is therefore necessary to control the parameters of the network and the computational complexity while maintaining accuracy.

Based on the above challenges, the main contributions of this paper are as follows.

- We propose an image processing module that extracts the foreground of the wood plank and avoids the interference of wood chips or other clutter.
- We propose a channel feature fusion method that directs the network to focus more on the region of interest.
- We use a spatial and channel attention module [22], which improves the network's ability to aggregate long-range information and significantly improves the model's detection performance. Compared to traditional neural networks, our model is relatively flat, has fewer parameters and computational requirements, and is therefore suitable for adapting to high-speed production on assembly lines.

We collected a large number of images of wood planks from the assembly line to create a data set and conducted experiments on it, demonstrating the effectiveness of our method. Compared with renowned models such as ResNet, MobileNet, VGG, EfficientNetV2, and ConvNeXt-Tiny, our algorithm demonstrated the highest precision. In particular, our model, while achieving this level of performance, boasts a competitive number of parameters and computational complexity, ranking second only to MobileNet in terms of efficiency.

2 Related Works

Existing detection methods based on image processing can be divided into two categories according to different technologies, one is to use traditional digital image processing to preprocess the image, extract wood defect features, and use feature induction or machine learning algorithms to detect the image. The results of wood defect segmentation and defect feature extraction in the algorithm directly determine the performance of the algorithm. Another category is to use of deep learning algorithms. In recent years, Convolutional Neural Networks (CNN) have become the mainstream method for image classification.

Traditional Digital Image Processing Approaches. Yu *et al.* [26] used a near-infrared spectrometer to collect spectral information of wood defects after locating the defects using machine vision. The raw spectra were processed and subjected to principal component analysis to obtain classification features. Using Discriminant Partial Least Squares (DPLS) modeling for feature classification, an impressive classification accuracy of more than 92.0% was achieved. Pramunendar *et al.* [13] used the Gray-Level Co-occurrence Matrix (GLCM) [23, 15, 3] to extract texture features from wood images. GLCM is based on statistical analysis of the frequency of occurrence of pixel values, rather than specific pixel values, which makes it robust to image brightness variations. This property is highly advantageous for image recognition tasks. After feature extraction, Pramunendar applied both a multilayer perceptron and a Support Vector Machine (SVM) to classify wood according to quality. The experimental results showed that the self-adaptive multilayer perceptron achieved the highest accuracy, reaching 78.82%. YongHua and Jin-Cong [25] proposed a defect detection method for wood surfaces based on combining Tamura and GLCM features. Tamura features offer advantages in describing the shape and direction of the texture, while fusion with GLCM enables a more comprehensive and accurate texture analysis. A Back Propagation(BP) neural network was then used to classify the wood images, achieving a maximum recognition rate of 90.67%. However, it should be noted that this method requires more computing time and memory for large images, which is a significant limitation in practical applications. Hao *et al.* [4] used an improved genetic algorithm to select feature wavelengths from denoised wood spectral images and then used an improved Bayesian neural network to build a model for defect detection and classification in solid wood planks. Barmpoutis *et al.* [1] proposed a novel spatial descriptor that treats each image as a collection of multidimensional signals. Specifically, this method represents wood images as concatenated histograms of high-order linear dynamical systems generated by vertical and horizontal image blocks. Using an SVM classifier, images were successfully classified with an accuracy rate of 91.47% in the classification of wood cross-section images.

Deep Learning Approaches. Zhu *et al.* [27] constructed an eight-layer CNN to extract features from wood images for classification, which showed improvements in both the efficiency and accuracy of wood defect detection. Jianan *et al.* [7] used Faster Region-based CNN to detect defects in solid wood planks, effectively improving the detection accuracy to a certain extent and reducing the detection time. These methods mainly focus on classifying wood planks by segmenting and extracting defects such as dead knots and live knots for defect detection. However, due to the high computational complexity of these methods, they are more suitable for wood plank screening applications than for defect detection screening tasks on production lines. Jing *et al.* [8] used Local Binary Patterns (LBP) and GLCM to extract the relationships between pixels in wood images. These extracted features were then used as the input layer of a CNN to detect and classify wood textures, achieving an experimentally validated accuracy of 93.94%. Yi *et al.* [24] proposes an improved YOLOx-tine network to improve key feature extraction and reduce the number of computational parameters by introducing a multi-pooled feature fusion module and an integrated feature extraction module instead of the original SPP and bottleneck modules. The model improves the model's performance by extracting key features, but the sensitivity of the model to small objects is lost to some extent. Cui *et al.* [2] proposed the Cascade Center of Gravity YOLOv7 (CCG-YOLO) based on the YOLOv7 model to improve the accuracy of wood defect detection and solve the problem of YOLO network insensitivity to small objects. Based on the existing YOLO network, the CBS feature extraction module of the YOLOv7 backbone network is streamlined to make the network more focused on shallow features and smaller targets. However, although the above feature extraction methods can extract some features of the surface image of wood planks, they often have limited recognition ability for unfamiliar samples, resulting in poor generalization and robustness.

Tan and Le [18] introduces the EfficientNetV2 model, which represents a significant improvement over the EfficientNet network. This model not only further improves accuracy over its predecessor, EfficientNet, but also significantly reduces both training and inference time by incorporating the Fused-MBConv module. As a result, EfficientNetV2 demonstrates excellent performance on the task of grading wooden boards. The advent of Vision Transformer has breathed new life and possibilities into the field of image algorithms. Among these advances, Mehta and Rastegari [10] presents MobileViT, a model that embodies both lightness and precision. By incorporating a lightweight Transformer module and effective dimensionality reduction strategies, MobileViT significantly reduces the number of model parameters and computational complexity, making it suitable for mobile-based image classification applications. Wang *et al.* [21] integrates the components and architecture of the Transformer model with the properties of deformable convolution, allowing it to maintain high efficiency while providing enhanced feature extraction capabilities. This approach demonstrates exceptional performance in image classification tasks; however, its large number of parameters requires significant computational and storage resources, making it impractical for resource-constrained real-time applications.

3 Proposed Methodology

Our proposed algorithm is based on CNN and image processing techniques including grayscale transformation, morphological operations, edge detection, etc. Our algorithm not only removes the interruptions from the image, but also preserves the regions of interest to make the image clearer. In addition, we calculate the edge information of the image and integrate it with the features of the removing background interference image to guide the subsequent network to pay more attention to the region of interest, thereby enhancing detection accuracy. CNN consists of convolutional layers, activation functions, pooling layers, fully connected layers, etc., which can perform feature extraction and classification of input images. However, as the number of network layers increases, the number of parameters and computational complexity increases dramatically, which can consume more time and resources, making it difficult to apply to production pipelines. As the number of layers in the network increases, features may be duplicated as they are passed from layer to layer, resulting in feature redundancy. To solve this problem, we designed a network with a shallow number of layers and introduced the spatial and channel attention operation, as shown in Figure 1. By designing a network structure with a shallow number of layers and combining it with the spatial and channel attention operation, we can avoid feature redundancy and maintain high classification accuracy to achieve more efficient image processing and classification.

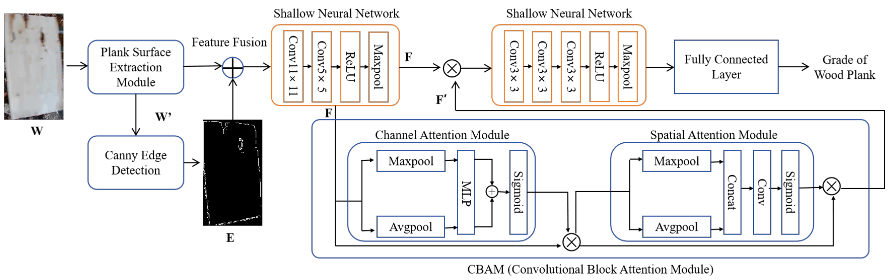

Figure 1: The input image W is fed into the plank surface extraction module to extract the wood plank surface W', W' is routed through the edge information extraction module to generate the edge information E, E and W' is fused into a shallow neural network to extract features F. The attention weights F' are generated by applying a spatial and channel attention module to the input F. After pixel-by-pixel multiplication of F and F', the result is sent into a neural network to extract features and then graded by a fully connected layer. Where $\oplus$ represents the concatenation of features along the channel dimension, while $\otimes$ represents the multiplication of features pixel by pixel.

3.1 Plank Surface Extraction Module

The camera captures the plank being transported on a conveyor belt after rotary cutting, and there may be some interruptions in the image, such as falling chips and debris on the floor, which can affect the detection of plank defects. The feature extraction module well solves this problem in this paper, which first converts the captured image to HSV color space and separates the values in the color space. Different objects in the image have different brightness, so the difference in brightness between them can be used to convert the image into a binary image. The interruptions from the woodchip are later removed using morphological operations. In image processing, morphological operations are often used to remove interruptions and improve the contours of an image. Morphological operations are based on the relative positional relationships between texture elements (also known as convolution kernels) and pixels, which can be achieved by dragging texture elements over the input image and modifying pixel values or pixel selection based on the pixel's relationship to the texture elements. As shown in Equation 1, B is the binary image and S is a structural element. B and S perform the open operation in the morphology, i.e. erosion followed by dilation, thus removing woodchip interruptions from the image.

$$B \odot S = (B \ominus S) \oplus S \tag{1}$$

After the morphological operations, only the plank region remains in the binary image. This binary image is used as a mask and logical operations in the next and manipulations are performed on the original image to obtain a pure wood plank image containing only the wood plank region. This process is illustrated in Figure 2.

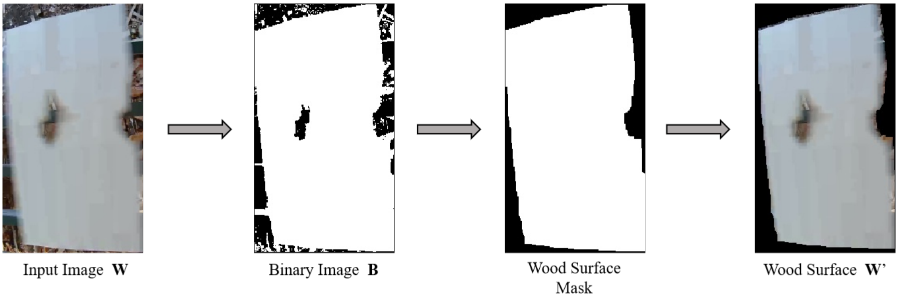

Figure 2: The input image W is converted to binary image B in HSV color space. To generate a wood plank mask using morphological opening operation on the binary image, bitwise AND operation between mask and W to extract wood plank surface W'.

3.2 Area-Guided Information and Feature Fusion

In this work, visual characteristics such as knots, grain and color are not related to board breakage and defects, but they do affect the classification results of the model.

To solve this problem, we use an edge detection algorithm to extract the edge information of the wood plank's VALUE channel in HSV color space. In HSV color space, the VALUE channel represents the luminance information of the image. By performing edge detection on the VALUE image, we can highlight defects such as cracks and voids in the wood plank. This is due to the large difference in luminance between the cracks and void areas and the plank itself. A normal plank will reflect light evenly, whereas cracks and void areas will absorb or scatter light due to damage, resulting in a noticeable difference in brightness. Cracks and void areas in the plank image can be highlighted using the edge information map as a region guide map.

$$G_X = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * B \qquad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * B \tag{2}$$

Detecting edge information using gradient operators. In Equation 2, the gradient components in the vertical and horizontal directions are obtained by convolving the horizontal and vertical operators, respectively, with the original image, where Gx is the gradient component in the horizontal direction, Gy is the gradient component in the vertical direction and B is the binary image of the wood plank.

Fusing an edge information map with a wooden plank image results in the generation of an image with four channels, where the crack and void regions are guided and prominently highlighted. This fused image serves as input data for the model and contains both obvious fracture and defect information, as well as details of the original plank image. The model can learn from the fused image and more accurately identify and locate plank breaks and defects during inference. The edge information map is depicted in Figure 3.

3.3 Spatial and Channel Attention Module

When classifying the plank, the model must evaluate its overall quality to determine the appropriate grade for the plank. In deep learning, the convolutional operation is a local perception method, where each convolutional kernel can only perceive a small region of the input image. Although it is possible to cover the entire input image with a sliding window, each convolutional kernel can still only perceive local information, which may result in the convolutional operation failing to capture global contextual information. Therefore, to better capture global contextual information, we introduce an approach called the

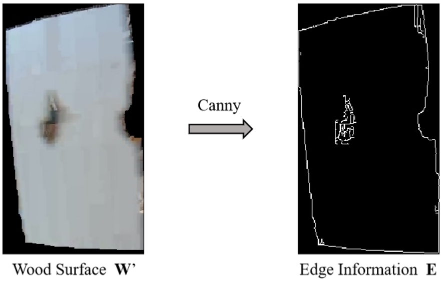

Figure 3: W' is processed by the edge detection algorithm to generate the edge E.

Convolutional Block Attention Module (CBAM) [22] in the spatial and channel attention module. This module performs certain operations on the spatial and channel information in the image so that critical information in the image is given a higher weight and retained, while irrelevant information is given a lower weight and suppressed or removed.

Specifically, the spatial and channel attention module consists of two submodules: the spatial attention module and the channel attention module. The spatial attention module operates on the feature map and calculates the importance of each position. This allows the model to focus on critical regions in the image and ignore irrelevant regions.

The channel attention module learns the correlation between different channels in the input data, automatically learns the importance of each feature channel, and assigns different weight coefficients to each channel, greatly improving the processing of fused features. This module can help neural networks better understand and utilize the information relationships between different channels in input data, thereby improving the performance of the network.

$$F' = M_s\left(M_c\left(F\right) \otimes F\right) \otimes \left(M_c\left(F\right) \otimes F\right) \tag{3}$$

In the context of the spatial and channel attention module, the input features initially undergo a channel attention computation, resulting in the generation of $M_c\left(F\right)$. Subsequently, the spatial attention computation is executed on the features that have been weighted by the combination of $M_c\left(F\right)$ and the original feature map F, resulting in the spatial attention weight map $M_s\left(M_c\left(F\right) \otimes F\right)$. Finally, a pixel-wise multiplication operation is applied to the features weighted by both $M_c\left(F\right)$ and F, using $M_s\left(M_c\left(F\right) \otimes F\right)$ as multiplier, to yield the refined feature map F', as shown in Equation 3.

4 Experiments

In this section, we compare our method with existing methods for plank image data and perform ablation experiments to demonstrate the effectiveness of the method.

4.1 Dataset of Planks

We use the video of the plank conveyor obtained from the factory production line to extract plank data. Due to the limited performance of the factory's filming equipment and the high operating speed of the conveyor belt, the clarity of the image data of the planks we collected is not satisfactory enough. To ensure image classification accuracy, experienced graders screened plank grades in our study and collected 20,000 plank images from Class I, Class II, and Class III planks. The inclusion criteria for Class I planks were intact and clear plank surfaces with no breaks or voids. The inclusion criteria for Class II planks were no obvious defects on the plank surface, but only a few voids in a small area. Class III planks had noticeable defects on the plank surface, a higher number of voids, or a significant area that rendered them unusable. Following an 8:2 ratio, we partitioned the dataset into a training set comprising 16,000 images and a test set containing 4,000 images, with the proportion among the three categories being 3:4:3. Figure 4 presents a comparison of the images of the three types of planks.

4.2 Experimental Settings

We used Pytorch to develop our algorithm and experimented on a NVIDIA GeForce GT 730 with an Intel i5-10500H processor and 16GB of RAM. During model training, wood plank images were uniformly normalized to 300×300 using data enhancement operations such as random scaling and random rotation. The network was trained using the Adam optimizer, the model was constrained using the cross-entropy loss function, and all activation functions used Relu. The initial learning rate was set to 10^{-4} and the batch size to 32. The model was trained for 200 epochs.

4.3 Evaluation Indexes

We use six metrics to evaluate the performance of the model: accuracy rate, precision rate, recall rate, number of parameters, FLOPs, and runtime. The accuracy rate is the ratio of the number of correctly classified samples to the total number of samples. The recall rate is the ratio of the number of Class I plank samples correctly identified by the model to the actual number of all Class I plank samples. The precision rate refers to the proportion of samples

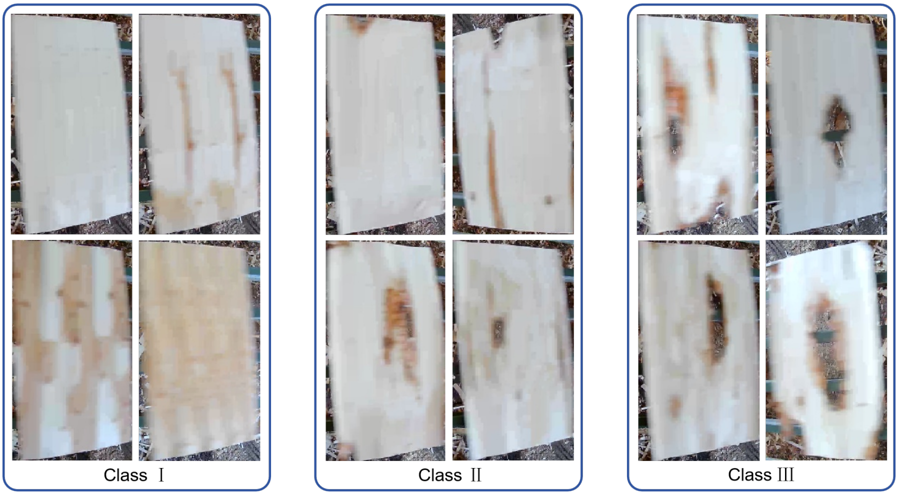

Figure 4: The inclusion criteria for Class I planks were intact and clear plank surfaces without any breaks or voids. The inclusion criteria for Class II planks were no obvious defects on the plank surface, but only a few voids in a small area. Class III planks had noticeable defects on the plank surface, a higher number of voids, or a significant area that rendered them unusable.

classified as Class I planks that are actually Class I planks. The formulae for accuracy rate, precision rate, and recall rate are specifically defined by

$$\begin{aligned} Accuracy &= \frac{TP+TN}{TP+FN+TN+FP} \\ Precision &= \frac{TP}{TP+FP} \\ Recall &= \frac{TP}{TP+FN} \end{aligned} \tag{4}$$

where TP, TN, FN, and FP denote the number of true positives, true negatives, false negatives, and false positives, respectively. The number of parameters is used to assess the complexity of the model. FLOPs refer to the total number of floating-point arithmetic operations required by a model to perform a specific task, which primarily serves as a metric to measure the computational complexity of the model. The runtime is utilized to quantify the duration required by a model to recognize a single image.

4.4 Comparison with State-of-the-Art Methods

Given the current scarcity of specialized classification algorithms for wooden planks, we have conducted an in-depth comparative analysis between the novel algorithm proposed in this paper and various existing image classification algorithms. Within an identical training framework, we retrained these image classification algorithms using a wood plank dataset and subsequently performed real-time classification tests on wooden planks. This process is intended to comprehensively evaluate and compare the performance of the multi-channel spatial attention mechanism introduced in this paper in practical applications.

Baseline Methods. The methods used for comparison include classic classification networks such as ResNet [5], VGG [17], MobileNetV2 [16], and networks that have achieved excellent performance in recent years such as EfficientNetV2 [18] and ConvNeXt [9]. Additionally, as this task requires lightweight algorithms, no ViT-related models were selected for this experiment.

Qualitative Comparisons. The experimental results are shown in Table 1. From the table, we can see that the accuracy of ResNet18 is 85.6%, the precision is 88.2%, and the recall is 75.0%. By adding the extraction method on top of ResNet18, it can achieve an accuracy of 91.1%, precision of 94.9%, and recall of 81.1%. The experiment proves that the feature information extraction module proposed in this chapter is also applicable to other models. Therefore, this experiment is also carried out for other models by adding an extraction module to them. Among them, the number of parameters of MobileNetV2 is 2.2M, and the FLOPs are 0.6G, which is less complex and more lightweight compared to the method proposed in this chapter. However, in terms of classification accuracy, the method in this paper improves the accuracy by 3.6%, the precision by 0.3%, and the recall by 8.2%. In addition, EfficientNet and ConvNeXt-tiny achieved relatively good results with 94.7% and 86.7% accuracy, 95.8% and 89.5% precision, and 88.1% and 71.3% recall, respectively. Although VGG has better results in terms of precision, its number of parameters and FLOPs is extremely large. In comparison, the method proposed in this paper obtained 95.3% accuracy, 96.0% precision, and 89.8% recall with 2.5M parameters and 1.1G FLOPs, which achieved better results in classification precision, boasting significantly fewer parameters and a reduction in FLOPs.

Furthermore, a statistical analysis of the shortest runtime for each method under identical conditions was also conducted in this experiment. The algorithm proposed in this paper, due to its simple structure and relatively shallow network hierarchy, achieved the shortest runtime of 20ms, significantly outperforming MobileNetV2 which required 49ms and ResNet which required 85ms for the same task.

Confusion Matrix. In the present study, we performed a thorough performance evaluation of the constructed classification model. To visually demonstrate the efficacy of classification, we generated a confusion matrix based on the experimental results, as shown in Figure 5. Each row of the confusion matrix corresponds to a true category of planks, each column

Table 1: Qualitative comparison of different methods on wood plank dataset. Our method achieved the best accuracy rate and recall rate, as well as the second-best precision rate. Key: [Best, Second Best, Third Best].

Method	Accuracy	Precision	Recall	Parameter	FLOPs	Runtime
ResNet18	85.6%	88.2%	75.0%			
ResNet18+Extraction	91.1%	94.9%	81.1%	11.2M	3.4G	85ms
MobileNetV2	82.5%	82.9%	72.5%			
MobileNetV2+Extraction	91.7%	95.7%	81.6%	2.2M	0.6G	49ms
VGG11	80.5%	85.0%	73.0%			
VGG11+Extraction	92.3%	96.9%	81.7%	0.1G	13.2G	292ms
EfficientNetV2+Extraction	94.7%	95.8%	88.1%	20.0M	5.4G	216ms
ConvNeXt-tiny+Extraction	86.7%	89.5%	71.3%	27.8M	7.2G	173ms
Ours	95.3%	96.0%	89.8%	2.5M	1.1G	20ms

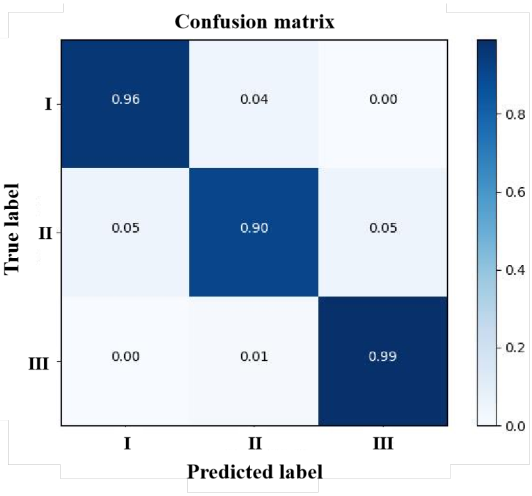

Figure 5: Confusion matrix of the plank classification. In the experiment, 4% of the class I planks were misclassified by the model as class II; 5% of the class II planks were misclassified as class I, and another 5% of the class II planks were misclassified as class III; 1% of the class III planks were misclassified as class II.

corresponds to a predicted category of planks, and the (i, j)-th entry of the matrix represents the rate of the i-th category samples that are classified to the j-th category. By analyzing this confusion matrix, we gained a deep understanding of the performance of the model on various plank categories. As shown by the diagonal entries of the confusion matrix, the model identified all three categories at high accuracies. Nevertheless, we observe that it also misclassified a portion of class I wooden planks as class II, a portion of class II wooden planks as either class I or class III, and a portion of class III wooden planks as class II. Specifically, 4% of the class I planks were misclassified by the model as class II, 5% of the class II planks were misclassified as class I, another 5% of the class II planks were misclassified as class III, and 1% of the class III planks were misclassified as class II.

Reasons of misclassification. Four representative cases of misclassification are presented in Figure 6, where Figure 6 (a) shows a class I plank misclassified to class II, Figure 6 (b) shows a class II plank misclassified to class I, Figure 6 (c) shows a class II plank misclassified to class III, and Figure 6 (d) shows a class III plank misclassified to class II. These cases show that background clutter (seen through plank cavity) and local color change on plank surface are two important factors that affect classification accuracy. As shown in Figures 6 (a) and (c), sharp local color changes, as marked by the red boxes, may have led to misclassification of a plank from a higher quality rank to a lower one. As shown in Figures 6 (b) and (d), background clutters with plank-like appearance seen through cavities, as marked by the red boxes, may have led to misclassification of a plank from a lower quality rank to a higher one. Note that, though misclassifications were made between adjacent quality ranks at minor probabilities, we have found zero severe misclassification between class I and class III in our experiments, which makes our model well aligned with the quality management of production line.

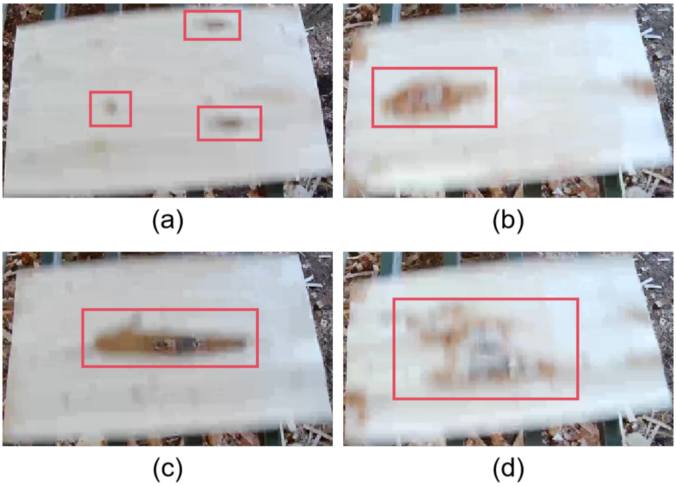

Figure 6: Examples of misclassification. The class I plank shown in (a) is misclassified to class II, the class II plank shown in (b) is misclassified to class I, the class II plank shown in (c) is misclassified to class III, and the class III plank shown in (d) is misclassified to class II. Problematic regions that may have caused the misclassifications are marked with red boxes on the images.

4.5 Ablation Study

To validate the effectiveness of each module in our algorithm, we performed ablation experiments. Specifically, we performed the following experiments on the model using the same training and testing setup on the experimental dataset. First, we used the proposed shallow neural network as a baseline and then added each module proposed in this paper step by step to test the effectiveness of each module. The results of the experiments are shown in Table 2. Where Base represents the shallow neural network, Surface represents the plank surface extraction module, and Fusion represents the feature fusion module.

Table 2: Four independent ablation studies are listed, utilizing the three classical classification evaluation metrics of accuracy, precision, and recall. The best result in each column is in bold font.

Method	Accuracy	Precision	Recall
Base	87.6%	83.8%	83.8%
Base+Surface Extraction	92.7%	95.1%	82.7%
Base+Surface Extraction+Fusion	93.0%	95.7%	84.6%
Base+Surface Extraction+Fusion+Attention	**95.3%**	**96.0%**	**89.8%**

First, we have experimented with the performance of the proposed shallow neural network when no module is added, and the results show that it has an accuracy of 87.6%, and precision and recall of 83.8% , 83.8%. After adding the plank surface extraction module, the accuracy is 92.7% and the precision and recall are: 95.1 % and 82.7%. After using the complete model, the accuracy is 95.3% and the precision and recall are: 96.0% and 89.8%. We found that the model, with the addition of the plank surface extraction module and the area-guided information and feature fusion module, is more accurate than the model without these additions, for the classification of images of conveyor belt-transported planks. This proves that the interruptions in the image have a large impact on the classification results. The performance of the model was highest after using the features after combining the edge information and adding the spatial and channel attention mechanism.

5 Conclusion

For the detection of wood planks in assembly line production, previous research has focused on defect detection and classification without fully considering the balance between performance and efficiency. To fill this gap, we propose a wood plank grading algorithm based on a multichannel spatial attention mechanism. The model uses a plank foreground extraction algorithm to extract the plank foreground from the image and directs the model's attention to important regions by fusing the edge information map of the plank with the foreground. The spatial and channel attention mechanism is then used to assign feature weights to help the algorithm learn. Compared to other benchmark neural networks, our model features fewer parameters, lower computational requirements, and a fast runtime, thereby making it suitable for adapting to high-speed production on assembly lines. The algorithm can more accurately focus on the classification features and ignore irrelevant regions to some extent to ensure classification accuracy. The experimental results conducted on our collected wood plank dataset demonstrate the effectiveness of the proposed model. In the classification experiments in the wood plank dataset, we obtained 95.3% accuracy, 96.0% precision, 89.8% recall, and a runtime of 20ms, which confirms the good performance of our model in wood plank classification. However, our method can identify only cracks and voids in wood planks, but not dead and live knots that should also help the plank grading. In the future, we plan to further investigate the detection of the knots and integrate it into our plank grading model.

Acknowledgements

This work is supported by key R&D Program of Shandong Province, China (No. 2023CXPT003) and Shandong Provincial Natural Science Foundation of China (No. ZR2022MF294).

References

[1] P. Barmpoutis, K. Dimitropoulos, I. Barboutis, N. Grammalidis, and P. Lefakis, "Wood species recognition through multidimensional texture analysis", *Computers and electronics in agriculture*, 144, 2018, 241–8.

[2] W. Cui, Z. Li, A. Duanmu, S. Xue, Y. Guo, C. Ni, T. Zhu, and Y. Zhang, "CCG-YOLOv7: A Wood Defect Detection Model for Small Targets Using Improved YOLOv7", *IEEE Access*, 2024.

[3] A. Fahrurozi, S. Madenda, D. Kerami, *et al.*, "Wood Texture features extraction by using GLCM combined with various edge detection methods", in *Journal of Physics: Conference Series*, Vol. 725, No. 1, IOP Publishing, 2016, 012005.

[4] L. Hao, C. Jun, L. Xue, and Z. Yi-zhuo, "Surface defects detection of solid wood board using near-infrared spectroscopy based on Bayesian neural network", *Spectroscopy and Spectral Analysis*, 37(7), 2017, 2041–5.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, 770–8.

[6] 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*, 152, 2020, 107357.

[7] F. Jianan, L. Ying, H. U. Zhongkang, Z. Qian, S. Luxiang, and Z. Xiaolin, "Solid wood panel defect detection and recognition system based on Faster R-CNN", *Journal of Forestry Engineering*, 2019.

[8] J. Jing, Q. Deng, F. LIP, *et al.*, "Fabric structure classification based on LBP and GLCM fusion", *Electronic Measurement and Instrumentation*, 29(9), 2015, 1406–13.

[9] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s", in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, 11976–86.

[10] S. Mehta and M. Rastegari, "Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer", *arXiv preprint arXiv:2110.02178*, 2021.

[11] M. Mousavi, M. S. Taskhiri, D. Holloway, J. Olivier, and P. Turner, "Feature extraction of wood-hole defects using empirical mode decomposition of ultrasonic signals", *Ndt & E International*, 114, 2020, 102282.

[12] L. Pan, R. Rogulin, and S. Kondrashev, "Artificial neural network for defect detection in CT images of wood.", *Comput. Electron. Agric.*, 187(106312), 2021, 1–7.

[13] R. A. Pramunendar, C. Supriyanto, D. H. Novianto, I. N. Yuwono, G. F. Shidik, and P. N. Andono, "A classification method of coconut wood quality based on Gray Level Co-occurrence matrices", in *2013 International Conference on Robotics, Biomimetics, Intelligent Computational Systems*, IEEE, 2013, 254–7.

[14] Y. Qi, Y. Zhou, J. Xu, and Z. Ge, "Development of a wood computed tomography imaging system using a butterworth filtered back-projection algorithm", *Forest Products Journal*, 68(2), 2018, 147–56.

[15] D. Ramayanti *et al.*, "Feature textures extraction of macroscopic image of jatiwood (Tectona Grandy) based on gray level co-occurence matrix", in *IOP Conference Series: Materials Science and Engineering*, Vol. 453, No. 1, IOP Publishing, 2018, 012046.

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, 4510–20.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *arXiv preprint arXiv:1409.1556*, 2014.

[18] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training", in *International conference on machine learning*, PMLR, 2021, 10096–106.
[19] M. S. Taskhiri, M. H. Hafezi, R. Harle, D. Williams, T. Kundu, and P. Turner, "Ultrasonic and thermal testing to non-destructively identify internal defects in plantation eucalypts", *Computers and electronics in agriculture*, 173, 2020, 105396.
[20] A. Thumm and M. Riddell, "Resin defect detection in appearance lumber using 2D NIR spectroscopy", *European Journal of Wood and Wood Products*, 75, 2017, 995–1002.
[21] W. Wang, J. Dai, Z. Chen, Z. Huang, Z. Li, X. Zhu, X. Hu, T. Lu, L. Lu, H. Li, *et al.*, "Internimage: Exploring large-scale vision foundation models with deformable convolutions", in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, 14408–19.
[22] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module", in *Proceedings of the European conference on computer vision (ECCV)*, 2018, 3–19.
[23] A. R. Yadav, M. Dewal, R. S. Anand, and S. Gupta, "Classification of hardwood species using ANN classifier", in *2013 Fourth national conference on computer vision, pattern recognition, image processing and graphics (NCVPRIPG)*, IEEE, 2013, 1–5.
[24] Z. Yi, L. Luo, Q. Lu, M. Chen, W. Zhu, and Y. Zhang, "An Efficient and Accurate Surface Defect Detection Method for Quality Supervision of Wood Panels", *Measurement Science and Technology*, 2024.
[25] X. YongHua and W. Jin-Cong, "Study on the identification of the wood surface defects based on texture features", *Optik-International Journal for Light and Electron Optics*, 126(19), 2015, 2231–5.
[26] H. Yu, Y. Liang, H. Liang, and Y. Zhang, "Recognition of wood surface defects with near infrared spectroscopy and machine vision", *Journal of Forestry Research*, 30(6), 2019, 2379–86.
[27] Y. Zhu, Q. Ouyang, and Y. Mao, "A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy", *BMC bioinformatics*, 18, 2017, 1–10.