Surface Defect Detection Method of Wooden Spoon Based on Improved YOLOv5 Algorithm

Siqing Tian, Xiao Li, Xiaolin Fang, Xiaozhong Qi, and Jichao Li *

The available surface defect detection methods for disposable wooden spoons still involve screening with the naked eye. This detection method is not only inefficient but also accompanied by problems such as false detection and missed detection. Therefore, this paper proposes a detection method based on an improved YOLOv5 network model (YOLOv5-TSPP). This method uses the K-Means ++ algorithm to cluster the target samples in the data set to obtain anchor frames that are more in line with different target scales. The Coordinate Attention module is added to the backbone network of the YOLOv5 network model to improve the feature extraction ability of the model. A new SPP module is added to the backbone network to increase the important features in the receptive field extraction network to improve the detection accuracy of small targets. The experimental results show: The YOLOv5-TSPP algorithm has better detection performance and the mAP of defect detection reaches 80.3%, which is 9.2% higher than that of the YOLOv5 algorithm. Among them, the detection accuracy of black knot defect reached 98.6%, the detection accuracy of back crack defect reached 92.1%, and the detection accuracy of mineral line defect reached 92.3%.

DOI: 10.15376/biores.18.4.7713-7730

Keywords: Deep learning; Defect detection; YOLOv5; Wooden spoon

*Contact information: College of Information and Electronics Technology, Jiamusi University, Jimusi 154007, China; *Corresponding author: tian_siqing@126.com*

INTRODUCTION

The disposable wooden spoon is mainly made of birch as raw material, through cutting, soaking, drying, hot pressing, polishing, sorting, and other processes. The surface defects of wooden spoons are mainly divided into two categories; one is the natural defects of wood raw materials and the other is the defects formed during processing. The most common defects are black knots, mineral lines, pollution, and back cracks. These defects affect the appearance and quality of wooden spoons and reduce the export number of wooden spoons. The existing detection methods mainly rely on manual detection. According to the texture, structural characteristics, color of the raw materials of wooden spoons, and surface defects, the wooden spoons are classified and graded by human eyes (Gu *et al.* 2010). This method requires a lot of manual participation and faces problems such as low-quality inspection rates and excessive labor input. With the continuous development of computer vision technology, intelligent methods have become increasingly used for defect detection. YongHua and Jin-Cong (2015) proposed a detection method of mixed surface texture features, which ensured the accuracy and robustness of the model and could detect dead knot and live knot defects. Song *et al.* (2015) proposed a method based on image block percentile color histogram and feature vector texture classification,

which can detect knots and crack defects. Zhang *et al*. (2015) combined principal component analysis with compressed sensing technology and achieved high recognition accuracy in detecting wood defects. Mu *et al*. (2015) combined fuzzy mathematics with a back propagation neural network (BP) to build a fuzzy BP neural network (FBP), which can realize the automatic identification of wood defects. Abdullah *et al*. (2020) used gray dependence matrix (GLDM) for feature extraction and feature analysis to study appropriate displacement and quantitative parameters that can classify wood defects. Yang and Yu (2017) used wavelet-based ultrasonic testing to extract features of wood hole defects. Aleksi *et al*. (2019) detected the defects on wood by calculating the vector difference between the texture without defects and the texture with defects. However, the above detection methods are easily affected by the shape and texture of the wood itself and the surrounding environment (light, angle, *etc*.), such that it can be difficult to meet the needs of defect detection in complex image backgrounds.

With the development of convolutional neural networks, applying convolutional networks to target detection can allow the system to learn higher-level features of images and improve detection efficiency. At present, defect detection based on deep learning is mainly divided into two categories: one is based on region proposal, such as the Faster R-CNN model (Ren *et al.* 2015); the other is object-based regression methods, such as SSD (Liu *et al.* 2016) and YOLO model (Redmon *et al.* 2016). Shi *et al*. (2020) constructed a convolutional neural network and then used multi-channel Mask R-CNN to classify and locate defects, which can identify dead knots, live knots, and cracks in wood. Wang *et al*. (2018) used the fuzzy pattern recognition method to detect the surface defects of particleboard in motion and calculated the number of defects, defect area, and damage degree. Yang *et al*. (2019) used a 3D laser sensor system to classify and identify the surface defects of wood-based panels and obtained a final classification accuracy of 94.7% after applying SVM. Urbonas *et al*. (2019) used a faster region-based convolutional neural network (Faster R-CNN) to identify defects on the surface of wood veneers. He *et al*. (2019) proposed a hybrid fully convolutional neural network (Mix-FCN) to detect the location of wood defects and automatically classify the types of defects from wood surface images. He *et al*. (2020) used deep convolutional neural network (DCNN) to identify and detect defects in wood images collected by laser scanners. Chen *et al*. (2022) used deep learning algorithms to extract image features of the original image and laser alignment to achieve higher accuracy and used AOI to classify the final result defects of WDD-DL. Sun (2022) designed and developed an automatic detection method for wood surface defects based on deep learning algorithm and multi-criteria framework. Based on digital image processing technology, Ye *et al*. (2022) designed a complete set of real-time wood classification detection algorithms. Xia *et al*. (2022) improved the Faster R-CNN algorithm and proposed a surface defect detection algorithm based on the improved Faster R-CNN. Hacıefendioğlu *et al*. (2022) used the deep convolutional neural network (DCNN) model using the K-means clustering algorithm to further improve the detection results in terms of wood defect classification accuracy. The above detection model is complex, which may reduce the detection efficiency and accuracy in complex scenes.

In this paper, a surface defect detection algorithm for wooden spoons based on YOLOv5 is proposed by using deep learning. The K-Means ++ algorithm is used to cluster the target samples in the data set to obtain anchor frames that are more in line with different target scales and improve the accuracy of multi-target positioning and entity segmentation. The Coordinate Attention module is added to the backbone network of the YOLOv5 network model to improve the feature extraction ability of the model. A new SPPnet

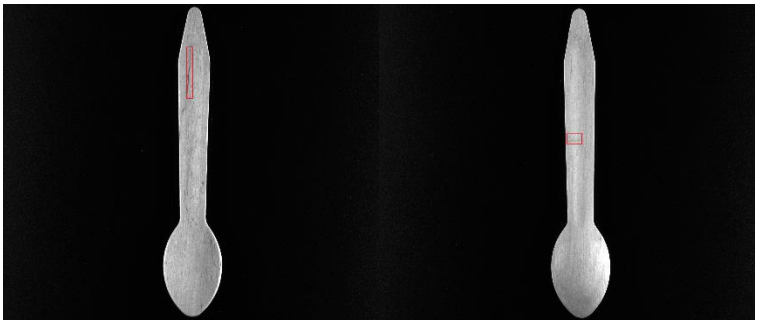
module is added to the backbone network to increase the receptive field to extract important features to improve the detection accuracy of small targets. The improved network can better improve the recognition accuracy of surface defects of wooden spoons and then improve the effective utilization rate of wooden spoons.

EXPERIMENTAL

Materials

The data set of wooden spoons was obtained from image acquisition and data enhancement. The image acquisition imports the image of the wooden spoon through the experimental platform. A representative surface defect image is shown in Fig. 1. The back crack is caused by the splitting of the back of the wooden spoon along the direction of the wood texture during the processing of the wooden spoon, such as the (a) red box mark position in Fig. 1. Black knots are naturally formed during the growth of trees. The wood defect is obvious. The color is deep, and it is patchy, such as the (b) red box mark in Fig. 1; Mineral lines are formed when trees absorb and deposit minerals such as carbonates from the soil, and the defective parts are dark strips, such as the (c) red box mark in Fig. 1. Defects caused by pollution include oil pollution or an unclean area on the production line during the production and processing of the wooden spoon. Defects can also be caused by mildew which shows as a black area on the surface of the wooden spoon. The defect site mainly presents a dark black oil stain state or mildew state, and the shape is not fixed, such as the (d) red box mark in Fig. 1. The training of convolutional neural networks requires a large number of samples. Through the learning of many samples, deep and specific features can be obtained to improve the accuracy of defect detection (Hou *et al.* 2021). To obtain a large number of sample sets and prevent over-fitting during network training, the collected wooden spoon images are enhanced to improve the robustness of the convolutional neural network. Data enhancement includes rotation, translation, cropping, mirroring, and brightness and contrast adjustment of the original image, without changing the pixel value. It only changes the position of the pixel, so that the network model can learn more image invariant features and avoid overfitting (Li *et al.* 2022). A total of 3,178 images were enhanced from the wooden spoon data set. The Labelimg tool was used to label the wooden spoon. Finally, the data set was divided according to the ratio of training set: verification set: test set = 8: 1: 1, 2542 images were used as the training set, 318 images were used as the test set, and the remaining 318 images were used as the validation set.

(a) beilie defect (back crack) (b) heijiezi defect (black knot)



(c) kuangwuxian defect (mineral line) (d) wuran defect (pollution)

Fig. 1. The collected wooden spoon defect samples

Experimental Platform

The image was sourced from the image acquisition experimental platform. The acquisition experimental platform mainly includes an industrial camera, lens, and light source. Considering that the longest side of the wooden spoon is 170 mm, the detection accuracy is 0.1 mm, the field of view is set to 170 mm * 170 mm, the target surface size is 8.8 * 6.6, and the working distance is 31.2 cm. Thus, a focal length of f = 1.6 mm is obtained. The Hikvision MVL-HF1628M-6MPE lens was selected. Finally, combined with the above parameters, the Hikvision MV-CA050-GM camera was selected with a resolution of 2448 * 2048. Because the defects of the wooden spoon mainly exist on the surface, the LED front lighting source was selected. The defects on the surface of the wooden spoon are mainly back crack, black knot, mineral line, and pollution. The defect location was not fixed, the defect size was different, and the mineral line defect was similar to the wood texture shape, which needs to be distinguished by color. In order to reduce the error, the white LED strip light source was finally selected.

The specific configuration of the computer used was Xeon (Skylake, IBRS) processor, Tesla T4 display adapter, and 16 GB memory. The software environment was the Ubantu18.04 operating system, Python3.8 programming language, and the Labelimg annotation tool was used to manually annotate the defect image. The Pytorch deep learning framework was built to train and test the surface defect data set of disposable wooden spoons. The hyperparameter settings in the training phase are shown in Table 1.

Table 1. Hyperparameter Settings

Parameter	Numeric Value
Initial learning rate	0.01
Final decay rate	5×10^{-4}
batch	8
Number of trainings	150
Momentum factor	0.937

YOLOv5 Network Structure

The YOLOv5 target detection network consists of four versions, namely YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x (Jocher *et al.* 2020). The weights of the four models increase in turn, and the detection accuracy increases with the weight. At the same time, the network training and inference time also increase. The detection object of this paper is small target defect detection, and the requirements for detection accuracy are

relatively high. Among the four models, YOLOv5x has the highest detection accuracy. Therefore, YOLOv5x was selected as the detection model. The model structure is shown in Fig. 2. The network model includes four parts: Input, Backbone, Neck, and Head.

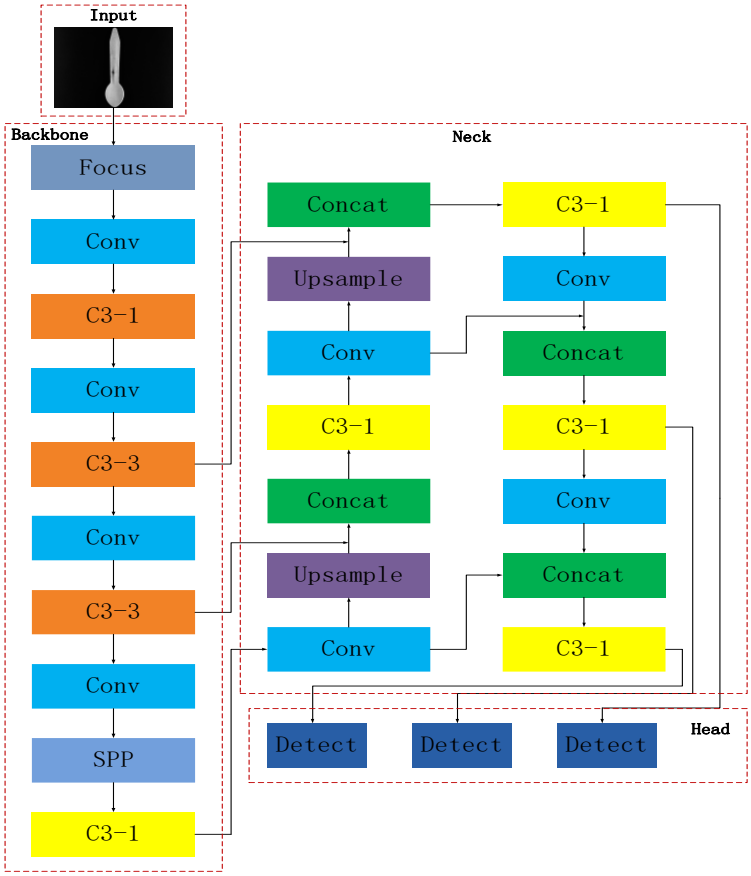


Fig. 2. YOLOv5 network structure

Input mainly includes Mosaic data enhancement, adaptive anchor box calculation, and adaptive image scaling. The data enhancement part can enrich the image background and improve the generalization ability of the network by randomly scaling, cropping, and arranging the images and then splicing them together. The adaptive anchor frame can compare the initial anchor frame with the real frame through training for different data sets, reversely updating, and obtaining the anchor frame parameters that are more suitable for the sample set. Adaptive image scaling scales the image of the input network to a unified standard size and sends it to the network for training. The algorithm can reduce the filling amount of the scaled image to avoid information redundancy and affect the inference speed. The Backbone part consists of a series of convolutional neural networks for extracting image features, mainly including CBS, C3, and SPPnet. Focus slice operation is used to convert the width and height information to the channel dimension, which reduces the information loss caused by feature downsampling. The C3 module is a replacement for the CSP module, which can effectively reduce the amount of calculation and streamline the network structure. The Neck part is a feature fusion network, which uses PANet and FPN. The FPN module performs a top-down multi-scale fusion of the multi-scale feature maps output by the feature extraction network. The PANet module performs a bottom-up multi-scale fusion of the multi-scale feature maps of FPN and finally outputs a feature map

with stronger location information and semantic information. The Head part is the prediction network. By convolution operation on the three outputs of the Neck end, three sets of feature vectors including category prediction box, confidence, and coordinate position are output.

YOLOv5-TSPP Network Structure

Coordinate attention mechanism

Due to the visual bottleneck of human beings, it is necessary to concentrate and ignore other secondary areas when observing a specific area. This behavioral action is called the attention mechanism, which has been helpful for various computer vision tasks. The attention mechanism mainly further extracts features from a given intermediate feature map by adding a simple and effective convolution attention module, aiming to improve the weight of beneficial features and suppress redundant features. Common attention mechanisms include SENet (Hu *et al.* 2020) (Squeeze-and-excitation), CBAM (Woo *et al.* 2018) (Convolutional block attention module) attention mechanism, and CA (Zhao *et al.* 2021) (Coordinate attention) attention mechanism. SENet uses average pooling to extract channel information. SENet captures the weight between channels through two fully connected layers. SENet compresses global spatial information into channel descriptors. It is difficult to retain location information that is critical to capture spatial structure in visual tasks. Selecting SENet channel attention alone will lose location information. CBAM focuses on the relationship between different spaces. By reducing the number of channels and using convolution to extract information, it pays more attention to the information in the spatial direction. However, the convolution method only extracts local relationships and cannot extract long-distance relationships.

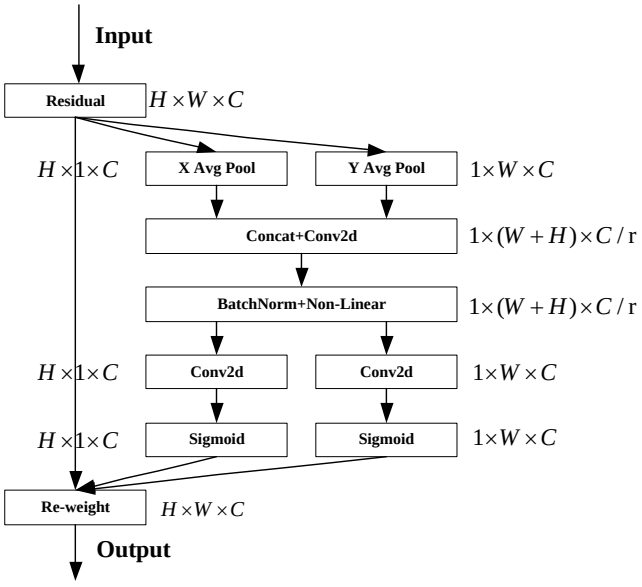


Fig. 3. Coordinate attention mechanism

The CA attention mechanism focuses on the width and height of the image and encodes the accurate position information. Firstly, the input feature map is divided into two directions of width and height for global average pooling, and the feature maps in the two directions of width and height are obtained respectively. Then, the feature maps in the two directions of width and height of the global receptive field, which are spliced together and

then they are sent to the shared convolution module. Finally, the weights on the width and height are obtained through the activation function. Location information is an essential factor for generating spatial selective attention maps in the detection of wooden spoon defects. Therefore, a method of introducing the C attention mechanism is proposed, as shown in Fig. 3, which considers the relationship and location information between channels.

To enable the attention module to capture the feature information with accurate position, the traditional global pooling method is decomposed into two one-dimensional feature codes. Specifically, given the input X, each channel is encoded along the horizontal and vertical coordinates using average pooling with sizes (H, 1) and (1, W), respectively. Therefore, the output of the c-th channel with height (h) and width (w) can be expressed as the following formula respectively. Where Z_c^h represents the output of the c-channel at height h; Z_c^w represents the output of channel c at the width w; the input X comes directly from the convolutional layer with a fixed kernel size.

$$Z_c^h(h)=\frac{1}{W}\sum_{0\le i<\mathrm{W}} x_c(h,i) \tag{1}$$

$$Z_c^w(w)=\frac{1}{H}\sum_{0\le j<\mathrm{H}} x_c(j,w) \tag{2}$$

The above two transformations aggregate features along two spatial directions respectively, and a pair of direction-aware feature maps are obtained. The process also allows the attention module to capture long term dependencies along one spatial direction and save accurate location information along another spatial direction. This helps the network exclude the interference of the image background and more accurately locate the target of interest. After the transformation in the information embedding, the height and width are spliced, and the feature map of the spatial information in the vertical and horizontal directions is generated through the convolution operation. The following formula is shown.

$$f=\delta(F_1([z^h,z^w])) \tag{3}$$

Then it is decomposed into tensor $f^h \in R^{C/r\times H}$ and tensor $f^w \in R^{C/r\times W}$ along the spatial information. Among them, r is used to control the sampling size reduction rate. Then 1×1 convolution transform F_h and F_w are performed on f^h and f^w respectively. Two tensors with the same number of channels are obtained as inputs, and the Sigmoid function transformation is performed as shown in the following formula,

$$g^h=\varphi(F_h(f^h)) \tag{4}$$

$$g^w=\varphi(F_w(f^w)) \tag{5}$$

where φ is the sigmoid activation function. The number of channels of f is reduced by the appropriate reduction ratio r, which reduces the calculation amount and complexity of the model. Then, g^h and g^w are extended as attention weights, respectively, and the following formula is used as an output.

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \quad (6)$$

The network structure before and after the introduction of the CA attention mechanism in the backbone network is shown in Fig. 4.

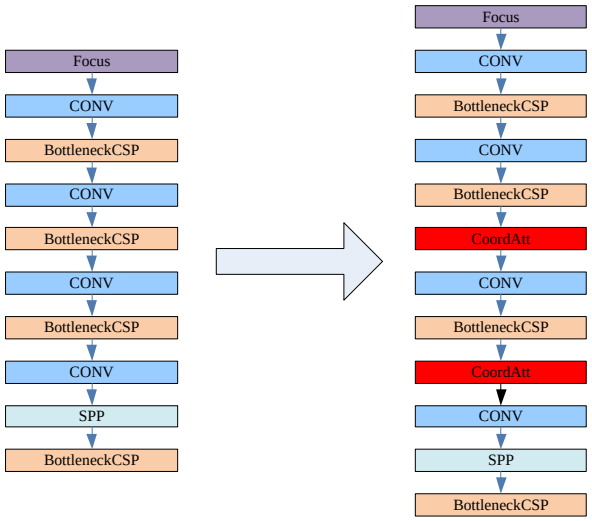


Fig. 4. Adds the CA attention mechanism

K-Means++ clustering algorithm

The initial anchor box in YOLOv5 is based on the data sets such as COCO (Common Objects in Context) or PASCAL VOC (The PASCAL Visual Object Classes), and the initial anchor box is finally obtained by using the K-means clustering algorithm. The implementation steps are as follows:

a) Randomly select k samples from all samples as the initial clustering center.

b) Calculate the Euclidean distance of each sample from the cluster center, and then divide the sample into the class closest to it.

c) The center point position of each cluster is recalculated according to the clustering results.

d) Repeat b) to c) until the internal elements in each cluster do not change, and all the final center point coordinates are the trained parameter model.

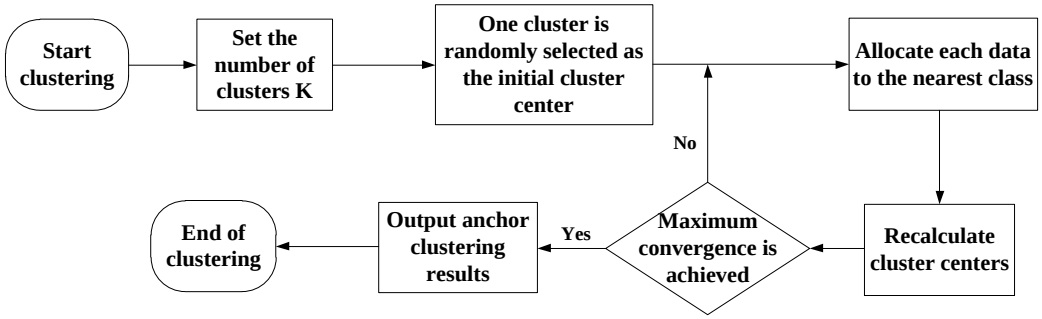


Fig. 5. K-Means++ algorithm flow chart

In YOLOv5, Euclidean distance is used to identify the similarity between the sample marker boxes, which can easily cause the marker boxes in a certain class to be too close to the initial anchor box size, and this will ultimately affect the clustering effect. Given the above problems, this paper proposes a K-Means++ clustering algorithm. The clustered prior box is closer to the target box of the wooden spoon image data set. The specific process of the K-Means++ algorithm is shown in Fig. 5.

K-Means++ is based on the traditional K-Means algorithm (Likas *et al.* 2003) to optimize the selection of the initial clustering center. The implementation steps of K-Means++ to optimize the initialization centroid are as follows:

Step 1: Set the spatial data set $P=\{p_1, p_2, ..., p_n\}$ of the input data point set, and randomly select a point p_i as the first clustering center K_1.

Step 2: For each point in the set P, use Formula (7) to calculate the minimum distance $D(p_i)$ between each object in the set and the current existing cluster center, and use Formula (8) to obtain the Sum of squares of these distances.

$$D(p_i)=\min\{(\sqrt{(p_i-K_n)^2})\} \tag{7}$$

$$Sum=\sum_{i=1}^{n} D^2(P_i) \tag{8}$$

Step 3: Calculate the probability P of each point being selected as the next cluster center using calculation formula (9). Take a random number R_i between the interval [0,1], subtract $\{P_1, P_2, ..., P_i\}$ with R_i in turn, until the result is less than 0. The point corresponding to P_i is the next cluster center.

$$P(i)=\frac{D^2(p_i)}{Sum} \tag{9}$$

Step 4: Repeat steps 2 ~ 3 to find the cluster center that meets the requirements. Through the above steps, the optimized initial clustering center is obtained.

Feature extraction network

SPPnet (He *et al.* 2015) is placed after the last feature layer of CSPDarknet53. After three convolutions of the last feature layer, it is processed with four different sizes of maximum pooling. The sizes of four different sizes of pooling kernels are 13×13, 9×9, 5×5, and 1×1. The structure is shown in Fig. 6.

By adding SPP structure in YOLOv5, the receptive field is increased, the most important contextual features are separated, and the detection speed is not reduced. Through the analysis of the SPPnet structure, it is concluded that the SPPnet used in the YOLOv5 structure cannot effectively extract the feature information of different scale targets. The SPPnet module is used as a variable and added to different positions of the backbone network to increase the receptive field to extract important defect features and improve the accuracy of wood spoon defect detection. The improved network structure is shown in Fig. 7.

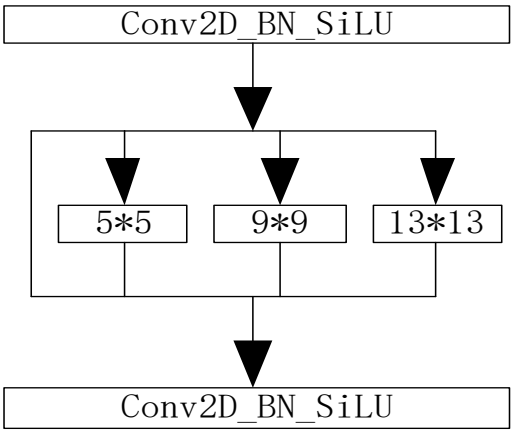


Fig. 6. Structure of SPPnet

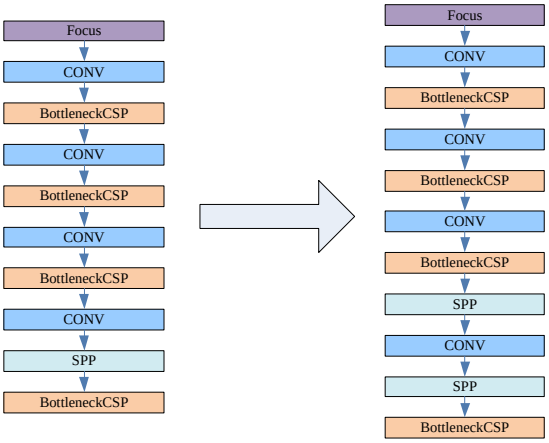


Fig. 7. YOLOv5-TSPP network structure

RESULTS AND DISCUSSION

Evaluating Indicator

According to the combination of the real label and the predicted label, each picture was divided into four categories: true positive (TP), true negative (TN), false positive (FP), and false negative (FN) (Zhao *et al.* 2021). Among them, TP is the object existing in the correctly recognized image, TN is the object existing in the image but not detected, FP is the object existing in the wrongly recognized image, and FN is the object existing in the image but not detected. Precision is the ratio of the number of positive samples correctly predicted to the number of positive samples predicted. Recall represents the proportion of the number of positive samples correctly determined to the total number of positive samples. The PR curve reflects the relationship between precision and recall rate. The higher the precision and recall rate of the model, the larger the area surrounded by the PR curve and the x, y-axis, and the better the overall performance of the model. AP is the area below the PR curve. The larger the AP, the higher the accuracy of the model. The smaller the AP, the worse the performance of the model. To evaluate the detection effect of the model obtained during training, the commonly used neural network performance

evaluation indicators are used in the experiment: Precision (P), Recall (R), and Average Precision (AP). The calculation formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{10}$$

$$R = \frac{TP}{TP + FN} \tag{11}$$

$$AP = \int_0^1 p(r)dr \tag{12}$$

Ablation experiment

To show the performance of the proposed method in the detection and recognition of surface defects of wooden spoons, ablation experiments were carried out for different defects. The average precision (AP), precision (Precision), and recall (Recall) were used as evaluation indicators. The network detection model that introduces improved K-Means++ clustering, CA coordinate attention mechanism, and re-adds a SPPnet structure has different degrees of improvement in accuracy, recall, and average accuracy compared with the original model. The experimental results show that the improved priori box determined by K-Means++ clustering can effectively improve the learning efficiency of the model for the target detection box. Secondly, because CA has a longterm dependence on location information and channel relationship, the introduction of CA effectively improves the efficiency of the model for location information learning and improves the prediction effect. Finally, a SPPnet network structure is added to realize the fusion of local features and global features, improve the learning efficiency of the model for features, and achieve better detection results. The experimental results are shown in Table 2.

Table 2. Ablation Experiment

Model	mAP	Recall				Precision				AP			
		B	H	K	W	B	H	K	W	B	H	K	W
YOLOv5X	71.12	83.75	79.73	74.04	34.21	89.40	89.39	86.41	61.29	87.81	82.70	79.09	34.88
YOLOv5x+K-Means++	71.73	84.42	79.80	74.38	34.53	90.56	92.19	86.63	61.88	88.57	84.12	79.16	35.07
YOLOv5x+CA+K-Means++	75.08	87.13	83.78	80.37	43.21	91.21	96.88	90.79	63.16	89.11	85.65	82.74	42.47
YOLOv5x+CA+K-Means+++SPPnet	80.30	87.73	90.79	83.72	55.83	92.12	98.57	92.31	69.53	89.21	92.24	85.32	54.43

Note: B represents the back crack defect, H represents the black knot defect, K represents the mineral line defect, and W represents the pollution defect.

Analysis of Table 2 shows that after adding K-Means + +, CA attention mechanism, and SPPnet, the overall detection accuracy of the network model for back crack, black knot, mineral line, and pollution is significantly improved. Due to the factors of the formation of

pollution defects and the morphological reasons of the pollution defects themselves, the pollution defects are easily confused with other defects (such as short mineral lines) and the texture features of the wooden spoon, so the network model has a relatively low detection accuracy for pollution.

Through the analysis and improvement of the YOLOv5 detection model, the detection precision images of the original model are shown in Fig. 8, and the detection precision images of the improved model are shown in Fig. 9. The recall images of the original model are shown in Fig. 10, and the recall images of the improved model are shown in Fig. 11. The PR images of the original model are shown in Fig. 12, and the PR images of the improved model are shown in Fig. 13.

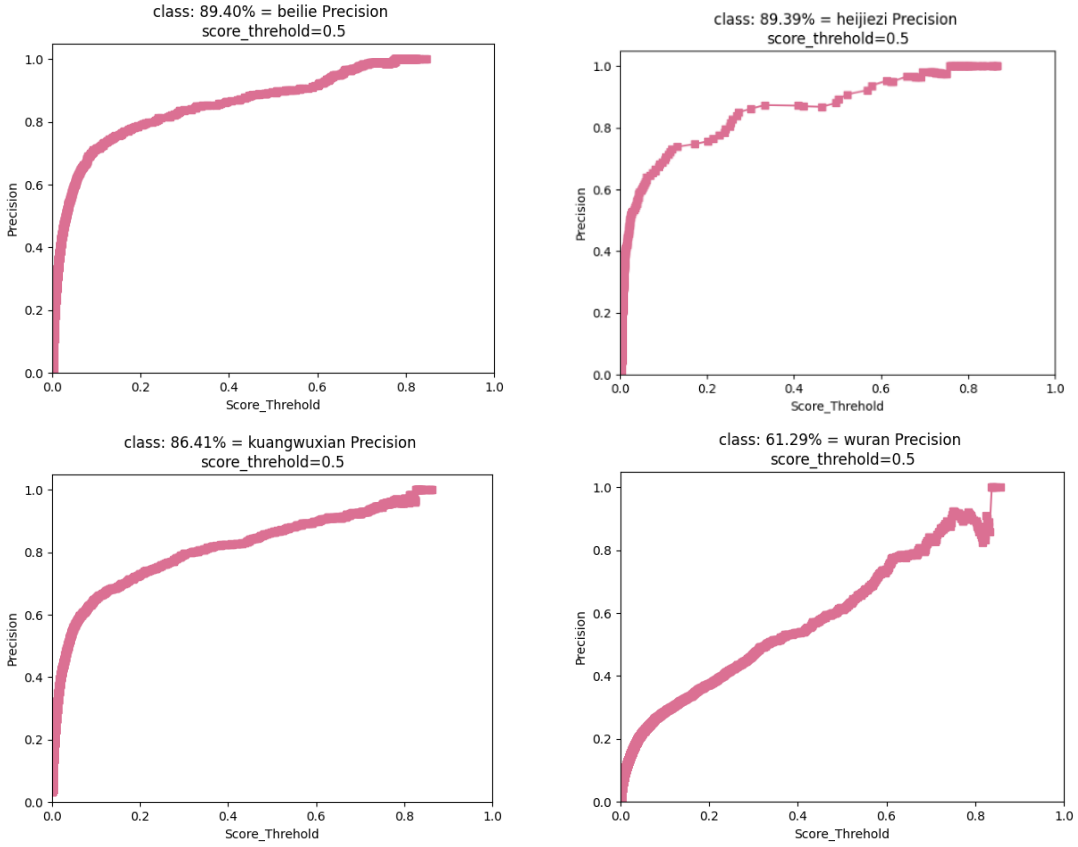


Fig. 8. Detection precision of YOLOv5 for four defects

When the detection confidence is greater than or equal to 0.5, the detection result can be reliable. Figures 8 and 9 show that when the confidence level is 0.5, the detection accuracy of back crack, black knot, mineral line, and pollution defect is 89.40%, 89.39%, 86.41%, and 61.29%. By using the above three improved strategies, the detection accuracy of the YOLOv5-TSPP detection model for back crack, black knot, mineral line, and pollution is 92.1%, 98.6%, 92.3% and 69.5%. The detection accuracy of the model for various defects has improved, indicating that the false detection rate of the model for various defects is decreasing.

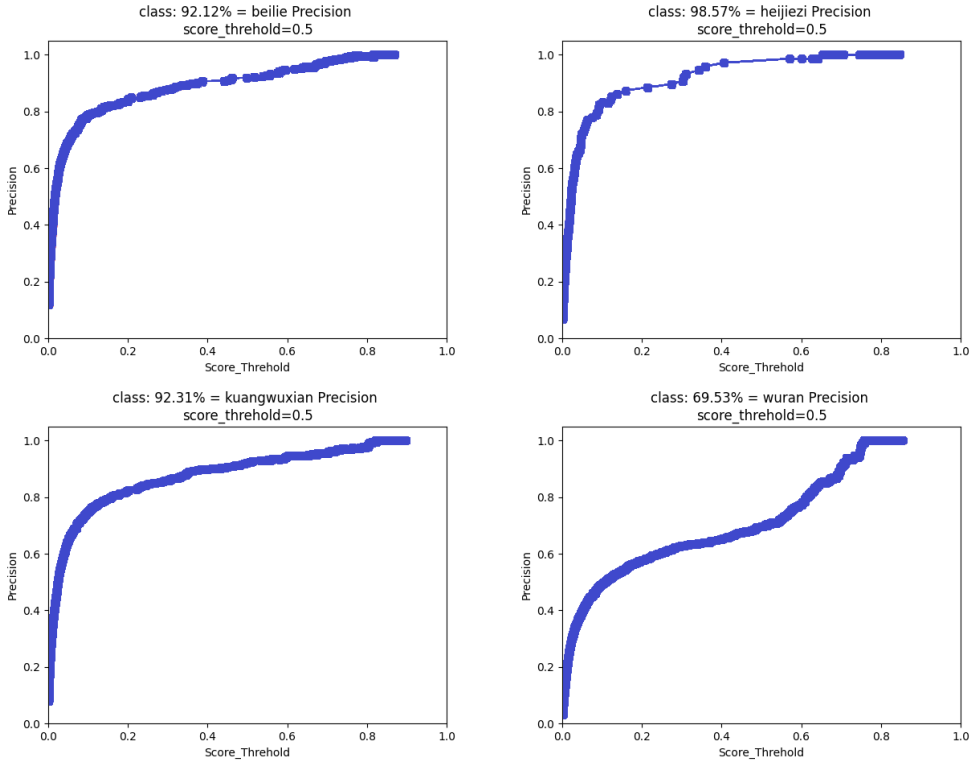


Fig. 9. Detection precision of YOLOv5-TSPP for four defects

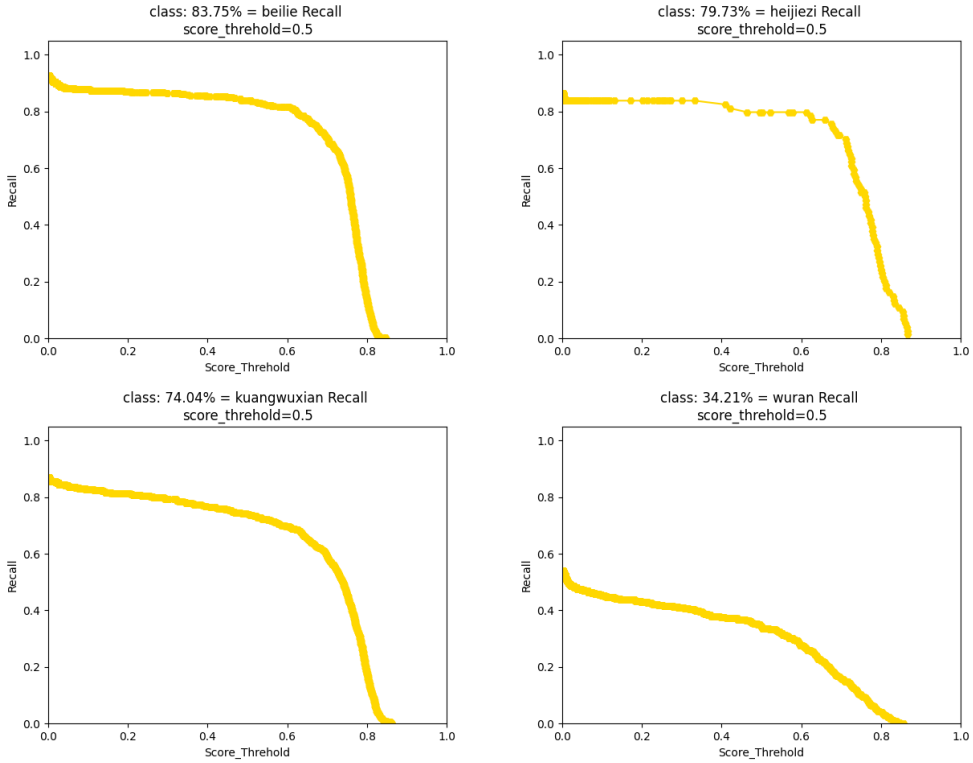


Fig. 10. Recall of YOLOv5 for four defects

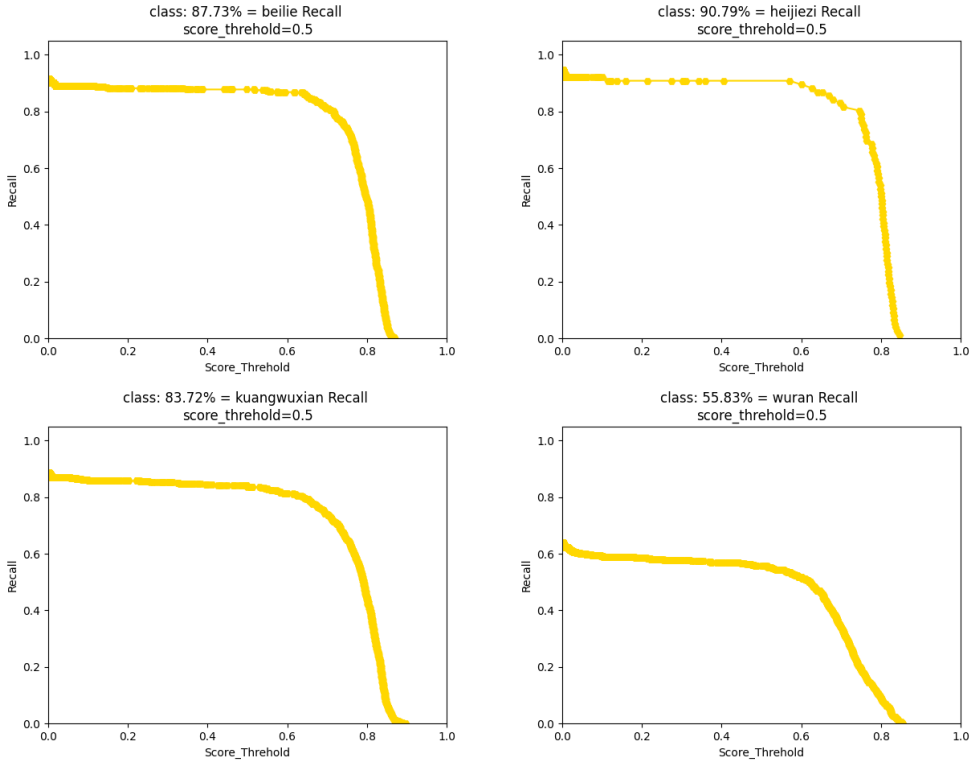


Fig. 11. Recall of YOLOv5-TSPP for four defects

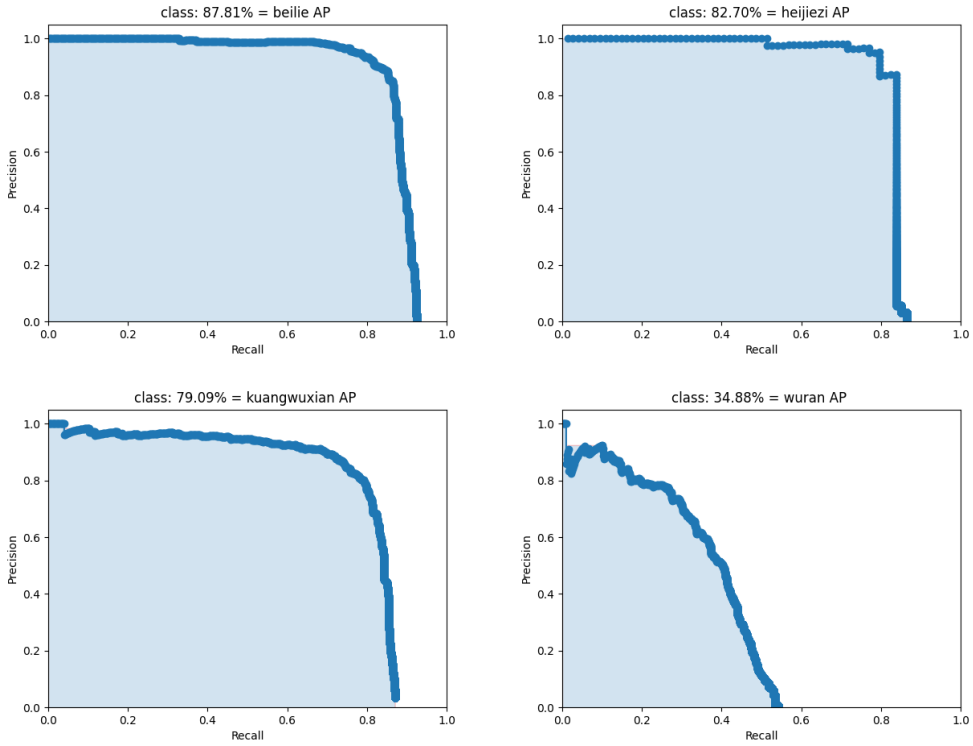


Fig. 12. The average precision of YOLOv5 for four defects

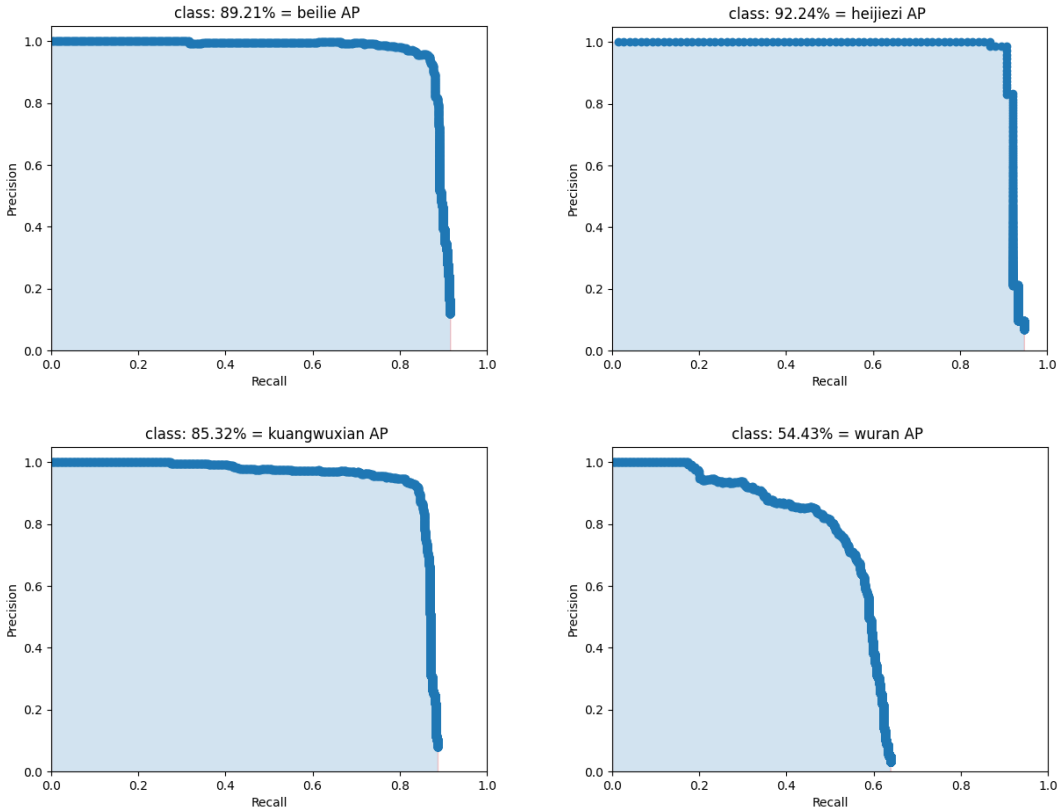


Fig. 13. The average precision of YOLOv5-TSPP for four defects

By comparing the recall rate images of the YOLOv5 model and YOLOv5-TSPP model, when the confidence level is 0.5, the recall rates of back crack, black knot, mineral line, and pollution in YOLOv5 are 83.8%, 79.7%, 74.0%, and 34.2%, respectively. The recall efficiencies of back crack, black knot, mineral line, and contamination in YOLOv5-TSPP were 87.7%, 90.8%, 83.7%, and 55.8%, respectively. YOLOv5-TSPP has a lower missed detection rate than YOLOv5.

The larger the area of the P-R curve (AP), the better the performance of the corresponding model. When the area reaches the maximum value, the precision and recall of the model reaches the maximum value, and the false detection rate and missed detection rate of the model are lower. From Figs. 12 and 13, the APs of back crack, black knot, mineral line, and contamination in YOLOv5 were 87.8%, 82.7%, 79.1%, and 34.9%, respectively. The AP of back crack, black knot, mineral line, and contamination in YOLOv5-TSPP were 89.2%, 92.2%, 85.3% and 54.4%, respectively. Therefore, the performance of YOLOv5-TSPP was better than that of YOLOv5.

Performance comparison of different models

In order to verify the performance of YOLOv5-TSPP, the same data set was trained and tested on SSD and Centernet network models. The mAP, Recall, and Precision of the models were then compared and counted. The results are shown in Table 3, where B represents the back crack defect, H represents the black knot defect, K represents the mineral line defect, and W represents the pollution defect.

Table 3. Performance Comparison of Different Algorithm Network Models

Model	mAP	Recall				Precision			
		B	H	K	W	B	H	K	W
SSD	38.47	17.65	5.41	4.33	0.16	85.71	57.14	72.22	20.00
Centernet	62.85	66.74	72.97	62.23	39.09	90.21	88.52	85.19	56.60
YOLOv5-TSPP	80.30	87.73	90.79	83.72	55.83	92.12	98.57	92.31	69.53

Compared with the other two models, the YOLOv5-TSPP model achieved a better overall detection effect, and the mAP was 41.8% and 17.4% higher than SSD and Centernet. The precision and recall of the four defects were improved compared with SSD and Centernet.

CONCLUSIONS

1. When improving the YOLOv5 model to identify the four main types of defects (back cracks, black knots, mineral lines, and pollution) in wooden spoons, the identification parameters (precision, recall, and average precision) of black knots were the highest, and the recognition effect is better.
2. The average accuracy of the improved YOLOv5 model for back crack, black knot, mineral line, and pollution identification results were 89.2%, 92.2%, 85.3%, and 54.4%, respectively.
3. Compared with the current mainstream detection models (SSD, Centernet), the improved model achieved higher accuracy, recall rate, and average precision rate. This indicates that the improved model exhibited a better recognition effect on the surface defects of wooden spoons than the other two mainstream detection models.

ACKNOWLEDGMENTS

This research was funded by the basic scientific research business fee of Heilongjiang provincial colleges and universities, grant number 2022-KYYWF-0604.

REFERENCES CITED

Abdullah, N. D., Hashim, U. R. a., Ahmad, S., and Salahuddin, L. (2020). "Analysis of texture features for wood defect classificationn," *Bulletin of Electrical Engineering and Informatics* 9, 121-128. DOI: 10.11591/eei.v9i1.1553

Aleksi, I., Sušac, F., and Matić, T. (2019). "Features extraction and texture defect detection of sawn wooden board images," in: *Proceedings of the 2019 27th Telecommunications Forum (TELFOR)*, pp. 1-4.. DOI: 10.1109/telfor48224.2019.8971381

Chen, L.-C., Pardeshi, M.S., Lo, W.-T., Sheu, R.-K., Pai, K.-C., Chen, C.-Y., Tsai, P.-Y., and Tsai, Y.-T. (2022). "Edge-glued wooden panel defect detection using deep learning," *Wood Sc. Technol.* 56, 477-507. DOI: 10.1007/s00226-021-01316-3

Gu, I. Y.-H., Andersson, H., and Vicen, R. (2010). "Wood defect classification based on image analysis and support vector machines," *Wood Science and Technology* 44, 693-704. DOI: 10.1007/s00226-009-0287-9

Hacıefendioğlu, K., Ayas, S., Başağa, H. B., Toğan, V., Mostofi, F., and Can, A. (2022). "Wood construction damage detection and localization using deep convolutional neural network with transfer learning," *European Journal of Wood and Wood Products* 80, 791-804. DOI: 10.1007/s00107-022-01815-5

He, K., Zhang, X., Ren, S., and Sun, J. (2015). "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence* 37, 1904-1916. DOI: 10.1109/tpami.2015.2389824

He, T., Liu, Y., Xu, C., Zhou, X., Hu, Z., and Fan, J. (2019). "A fully convolutional neural network for wood defect location and identification," *IEEE Access* 7, 123453-123462. DOI: 10.1109/access.2019.2937461

He, T., Liu, Y., Yu, Y., Zhao, Q., and Hu, Z. (2020). "Application of deep convolutional neural network on feature extraction and detection of wood defects," *Measurement* 152. DOI: 10.1016/j.measurement.2019.107357

Hou, Q., Zhou, D., and Feng, J. (2021). "Coordinate attention for efficient mobile network design," in: *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13708-13717. DOI: 10.1109/cvpr46437.2021.01350

Hu, J., Shen, L., Albanie, S., Sun, G., and Wu, E. (2020). "Squeeze-and-Excitation Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42, 2011-2023. DOI: 10.1109/tpami.2019.2913372

Jocher, G., Stoken, A., Borovec, J., Changyu, L., Hogan, A., Diaconu, L., and Yu, L.U. (2020). "YOLOv5: v3. 1-Bug Fixes and Performance Improvements," DOI: 10.5281/ZENODO.4154370

Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. (2022). "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Transactions on Neural Networks and Learning Systems* 33, 6999-7019. DOI: 10.1109/tnnls.2021.3084827

Likas, A., Vlassis, N., and Verbeek, J. J. (2003). "The global k-means clustering algorithm," *Pattern Recognition* 36, 451-461. DOI: 10.1016/s0031-3203(02)00060-2

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). "SSD: single shot multibox detector," in: *Proceedings of the 14th European Conf. on Computer Vision (ECCV)*, pp. 21-37. DOI: 10.1007/978-3-319-46448-0_2

Mu, H., Zhang, M., Qi, D., Guan, S., and Ni, H. (2015). "Wood defects recognition based on fuzzy bp neural network," *International Journal Smart Home* 9, 143-152. DOI: 10.14257/ijsh.2015.9.5.14

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788. DOI: 10.1109/cvpr.2016.91

Ren, S., He, K., Girshick, R., and Sun, J. (2015). "Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems* 9199, 2969239-2969250. DOI: 10.1109/tpami.2016.2577031

Shi, J., Li, Z., Zhu, T., Wang, D., and Ni, C. (2020). "Defect detection of industry wood

veneer based on NAS and multi-channel mask R-CNN," *Sensors (Basel)* 20. DOI: 10.3390/s20164398

Song, W., Chen, T., Gu, Z., Gai, W., Huang, W., and Wang, B. (2015). "Wood materials defects detection using image block percentile color histogram and eigenvector texture feature," in: *Proc. First International Conference on Information Sciences, Machinery, Materials and Energy*, pp. 779-783. DOI: 10.2991/icismme-15.2015.163

Sun, P.a. (2022). "Wood quality defect detection based on deep learning and multicriteria framework," *Mathematical Problems in Engineering* 2022, 1-9. DOI: 10.1155/2022/4878090

Urbonas, A., Raudonis, V., Maskeliunas, R., and Damasevicius, R. (2019). "Automated identification of wood veneer surface defects using faster region-based convolutional neural network with data augmentation and transfer learning," *Applied Sciences-Basel* 9. DOI: 10.3390/app9224898

Wang, C., Liu, Y., and Wang, P. (2018). "Extraction and detection of surface defects in particleboards by tracking moving targets," *Algorithms* 12. DOI: 10.3390/a12010006

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). "CBAM: Convolutional Block Attention Module," in: *Proceedings of the 15th European Conference on Computer Vision (ECCV)*, pp. 3-19. DOI: 10.1007/978-3-030-01234-2_1

Xia, B., Luo, H., and Shi, S. (2022). "Improved faster R-CNN based surface defect detection algorithm for plates," *Computational Intelligence and Neuroscience* 2022. DOI: 10.1155/2022/3248722

Yang, H., and Yu, L. (2017). "Feature extraction of wood-hole defects using wavelet-based ultrasonic testing," *Journal of Forestry Research* 28, 395-402. DOI: 10.1007/s11676-016-0297-z

Yang, J., Zheng, X., Yao, J., Xiao, J., and Yan, L. (2019). "3D surface defects recognition of lumber and straw-based panels based on structure laser sensor scanning technology," *Inmateh-Agricultural Engineering* 57, 225-232. DOI: 10.35633/INMATEH_57_25

Ye, R., Pei, Y., Wang, W., and Zhou, H. (2022). "Scientific computational visual analysis of wood internal defects detection in view of tomography image reconstruction algorithm," *Mobile Information Systems* 2022. DOI: 10.1155/2022/6091352

YongHua, X., and Jin-Cong, W. (2015) . "Study on the identification of the wood surface defects based on texture features," *Optik-International Journal for Light and Electron Optics* 126, 2231-2235. DOI: 10.1016/j.ijleo.2015.05.101

Zhang, Y., Xu, C., Li, C., Yu, H., and Cao, J. (2015). "Wood defect detection method with PCA feature fusion and compressed sensing," *Journal of Forestry Research* 26, 745-751. DOI: 10.1007/s11676-015-0066-4

Zhao, J. Q., Zhang, X. H., Yan, J. W., Qiu, X. L., Yao, X., Tian, Y. C., and Cao, W. X. (2021a). "A wheat spike detection method in UAV Images based on improved YOLOv5," *Remote Sensing* 13. DOI: 10.3390/rs13163095

Zhao, Z., Yang, X., Zhou, Y., Sun, Q., Ge, Z., and Liu, D. (2021b). "Real-time detection of particleboard surface defects based on improved YOLOV5 target detection," *Scientific Reports* 11. DOI: 10.1038/s41598-021-01084-x

Article submitted: August 5, 2023; Peer review completed: September 9, 2023; Revised version received: September 14, 2023; Accepted: September 17, 2023; Published: September 28, 2023.
DOI: 10.15376/biores.18.4.7713-7730