Letter

Wood Crack Detection Based on Data-Driven Semantic Segmentation Network

Ye Lin, Zhezhuang Xu, Dan Chen, Zhijie Ai, Yang Qiu, and Yazhou Yuan

Dear Editor,

This letter is concerned with wood crack detection which is important to guarantee the quality of wooden products. In the wood industry, the crack detection is one of the most challenging tasks in the wood defects detection, since the detection accuracy may be reduced due to the stains on the boards, the tiny cracks, and some cracks that are similar to the sound region. To overcome these challenges, we propose a data-driven semantic segmentation network based on U-Net, which is called WCU-Net, for wood crack detection. Specifically, a position attention mechanism is firstly proposed to exaggerate the wood crack positions, then the feature enhancement mechanism is designed to selectively derive more diluted information of tiny crack. Moreover, a residual block is adopted to obtain and fuse multi-scale receptive fields for finding more crack areas that are similar to the sound region. The experimental results show that WCU-Net can improve the accuracy of wood crack detection.

The crack on the wooden board is a typical defect which affects the quality of wooden products. Thus, the wood cracks have to be detected and then cut during the wooden furniture production. Generally, the wood cracks are marked by experienced workers with a fluorescent pen for wood cutting decision-making. The manual process is inefficient and human subjectivity often leads to low accuracy. Therefore, it is important to develop an automatic wood crack detection approach for wood industry.

With the support of machine vision [1], some research works have been proposed to separate the wood defects from background by growing boxes from a set of pixels in defective regions [2] or using a local threshold segmentation algorithm [3]. However, the performance of these algorithms is easily affected by the noise whose color is similar to the wood defects. On the other hand, machine learning algorithms have been used to classify wood defects. The extracted geometric and intensity features of wood defects were used to construct a regression tree (CART) classifier [4]. As the artificial neural networks improve the robustness in the industrial monitoring [5], [6], data-driven deep learning approaches have been designed to detect wood defects. In [7], the defects are firstly located by using watershed algorithm and opening operation, and then identified by deep learning neural network. To increase efficiency, the Gaussian function and the complete intersection over union (CIoU) loss function are adopted to improve YOLOv3 for wood defect detection [8].

Although these works have been proposed to detect wood defects, the crack detection is still one of the most challenging tasks in the wood defects detection. At first, in the industrial process, the wooden boards often contain stains, and the captured wood images vary in brightness due to illumination instability. These dynamic conditions make it difficult to segment wood cracks through traditional segmentation algorithms. Secondly, the wooden board exists lots of tiny cracks that are hard for convolutional neural network to derive their representative features. In this case, enhancing the diluted information of tiny crack is important. Moreover, some cracks are similar to the sound region on the wooden board, which makes the detection task more challenging.

To overcome these challenges, we propose a data-driven semantic segmentation network based on U-Net, called WCU-Net, for wood crack detection. The main contributions of WCU-Net are as follows: 1) We propose a position attention mechanism (PAM), where the attention map is firstly acquired from feature map and then used to exaggerate the wood crack positions. 2) The feature enhancement mechanism (FEM) is designed to selectively derive details based on the attention map. It enhances more detailed information of tiny wood crack that is diluted through convolutional operations. 3) A residual block with atrous convolution is adopted to obtain and fuse multi-scale receptive fields. It can recognize wood crack through larger receptive field, which helps find more crack areas that are similar to the sound region. The experimental results show that WCU-Net can improve the accuracy of wood crack detection.

Data collection: In order to capture wood images for wood crack detection, a machine vision testbed is deployed, as shown in Fig. 1. The wooden board is transported through a conveyor to the image capture area. The line scan cameras installed at the bottom and top of the conveyor start scanning the wooden board for capturing wood images. The obtained images are transmitted to the computer through single twisted pair Ethernet. Then, the captured images are processed to remove the background [9] and cropped into pieces for detection.

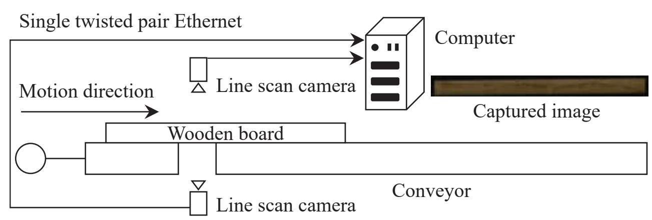

Fig. 1. The framework of machine vision testbed for wood crack detection.

The cropped wood images are illustrated in Fig. 2. It is clear to see that the wood contains stains, and the wood images vary in brightness. They make the traditional segmentation methods difficult for wood crack detection. On the other hand, the scales and texture of wood cracks are various. In particular, the span of wood crack varies, some areas in large-scale cracks are similar to the non-defective areas, and the wood contains many tiny cracks, some of which are very close to the natural wood texture. These characteristics bring enormous challenges to wood crack detection.

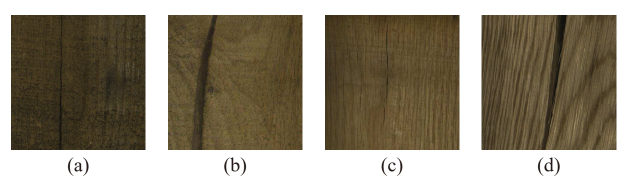

Fig. 2. Characteristics of wood cracks in the cropped wood images.

WCU-Net architecture: To achieve wood crack detection, this letter proposes a data-driven semantic segmentation network based on U-Net [10], which is called WCU-Net. The architecture of the WCU-Net is illustrated in Fig. 3 . WCU-Net includes four parts: a baseline network, a PAM, a FEM and a residual block. The baseline is constructed as encoder-decoder network structure. The PAM is

Corresponding author: Zhezhuang Xu.

Citation: Y. Lin, Z. Z. Xu, D. Chen, Z. J. Ai, Y. Qiu, and Y. Z. Yuan, "Wood crack detection based on data-driven semantic segmentation network," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 6, pp. 1510–1512, Jun. 2023.

Y. Lin, Z. Z. Xu, D. Chen, Z. J. Ai, and Y. Qiu are with College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China (e-mail: 210110010@fzu.edu.cn; zzxu@fzu.edu.cn; t03019@fzu.edu.cn; 200120084@fzu.edu.cn; n190127117@fzu.edu.cn).

Y. Z. Yuan is with the School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China (e-mail: yzyuan@ysu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2023.123357

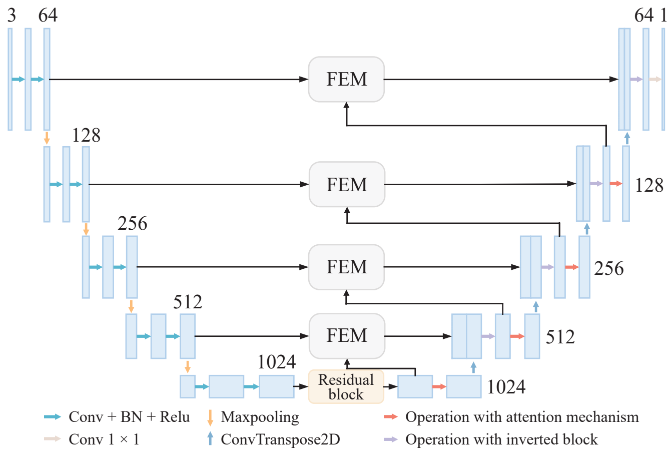

Fig. 3. The architecture of WCU-Net.

designed to focus on the wood crack areas. The FEM enhances the detailed information that is diluted in the high-level feature. The residual block is added to acquire larger receptive field.

1) Baseline network: Due to the excellent segmentation performance in different applications [11], U-Net is used as the baseline network of the proposed WCU-Net. For the encoder part, two convolution layers with a kernel size of 3×3 are used to derive features in different scales and a max-pooling layer with a stride of 2 for down-sampling. Each convolution layer is followed by batch normalization (BN) and a rectified linear unit (ReLU).

To improve the performance of feature fusion, inverted block [12], which consists of depthwise convolution and pointwise convolution, is adopted after concatenate operation in the decoding process. In this research, the number of channels is reduced by half through the operation of first pointwise convolution, and the remained operation in the inverted block keeps the number of channels same. For feature up-sampling, the size of the image is restored by the transposed convolution gradually. Sigmoid is used in the last layer.

2) Position attention mechanism: The stains on the wooden board may hinder neural network from deriving representative information of wood crack. To enhance the wood crack representation, we introduce a position attention mechanism, as shown in Fig. 4. The local feature $\mathbf{X}\in\mathbb{R}^{C\times H\times W}$ is fed into two convolution layers with $1\times W$ and $H\times 1$ filters to produce two vectors $\mathbf{A}\in\mathbb{R}^{H\times 1}$ and $\mathbf{B}\in\mathbb{R}^{1\times W}$, respectively. Then, a matrix multiplication between $\mathbf{A}$ and $\mathbf{B}$ is performed, and a softmax layer is applied to acquire the position attention map $\mathbf{S}\in\mathbb{R}^{H\times W}$. Finally, the feature $\mathbf{X}$ is multiplied by the map $\mathbf{S}$, and the element-wise sum of the output and the feature $\mathbf{X}$ is performed to generate the final output $\mathbf{Y}\in\mathbb{R}^{C\times H\times W}$.

3) Feature enhancement mechanism: Although concatenate operation in the U-Net recovers detailed local information, some noise is also introduced to the features, which affects prediction accuracy. Moreover, the tiny crack information is more easily diluted after several convolutional operations. In order to enhance the tiny wood crack areas, while eliminating the noise influence, a feature enhancement mechanism block is proposed, as shown in Fig. 5. Both high- and low-level features are fed into position attention mechanism to generate two position attention maps, α and β. Then, they are used to enhance the low-level features. The enhanced output is

$$F_e = F_{low} + F_{low}\otimes\beta\otimes U(1-\alpha) \tag{1}$$

where F_e is the enhanced output and U denotes the upsampling operation with the bilinear interpolation. F_{low} represents the low-level feature map. Symbols $+$ and $\otimes$ represent element-wise sum and multiplication, respectively.

4) The residual block: The high-level features in the U-Net have provided a large receptive field. Nevertheless, the span of wood crack is long and some areas in the wide cracks are similar to the sound region, which requires a larger receptive field. Thus, a residual block with atrous convolution is applied to acquire a larger receptive field. The residual block is constructed by connecting 3 residual units. The residual unit is composed of two 1×1 filters and a 3×3 filter. The atrous rates of 3×3 filter are set as 4.

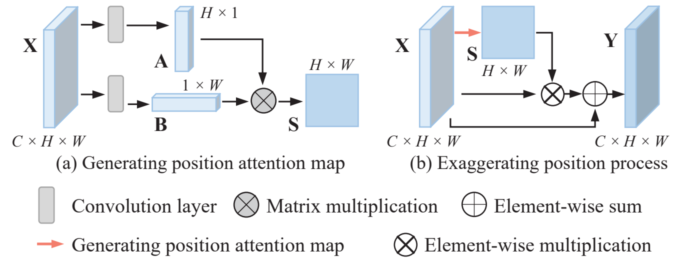

Fig. 4. The position attention mechanism.

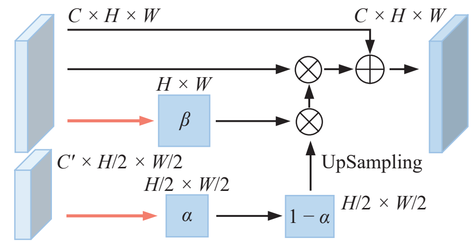

Fig. 5. The feature enhancement mechanism.

5) Loss function: Dice coefficient (DC) represents the similarity between the predicted result and the ground truth. Considering the Dice loss is not easily influenced by the class imbalance, the binary dice loss is chosen as the loss function, which is expressed as follows:

$$L_{\text{Dice}} = 1-\frac{2\sum_{n=1}^{N}p_n g_n+\epsilon}{\sum_{n=1}^{N}p_n^2+\sum_{n=1}^{N}g_n^2+\epsilon} \tag{2}$$

where p_n and g_n denote the prediction and ground truth, respectively. The minimum value ϵ is used to avoid occurring zero denominators.

Experiments: We collect the wood crack data set which is composed of 729 wood crack images with manually annotated segmentations. 549 images of them are randomly selected for network training and 180 images for testing. In the training process, all the images are resized to 416×416. The number of epoch, the initial learning rate, and the batch size are set as 300, 5E–4 and 4, respectively. The Cosine Annealing Learning Rate Scheduling is used to adjust the learning rate and Adam with a weight decay of 5E–4 is chosen as the optimizer. Considering that conditions in the industrial process are dynamic and the orientation of wood crack is different, the image augmentations, such as random crop, random lightness, random rotation and random distortion, are randomly implemented during the training phase to guarantee the robustness of the proposed WCU-Net.

1) Evaluation metrics: To measure the performance of segmentation models, the metrics, precision (Pr), recall (Re), F1-score (F1), and intersection of union (IoU), are introduced for evaluation [12].

2) Ablation study: An ablation study is performed for evaluating various designed modules. As presented in Table 1, the results demonstrate that the network combined with PAM obtains higher Re, F1-score and IoU than the baseline network, indicating that PAM focuses more on the wood crack areas. When fusing both the PAM and FEM to the proposed network, it acquires much higher Re, F1-score and IoU, which proves that FEM derives more detailed wood crack information. Meanwhile, the integrated residual block also yields incremental improvement.

Table 1. Ablation Study on Wood Crack Data Set

Settings	Pr	Re	F1	IoU
U-Net	0.951	0.791	0.864	0.760
U-Net + PAM	0.951	0.817	0.879	0.784
U-Net + PAM + FEM	0.947	0.833	0.886	0.795
U-Net + PAM + FEM + Residual block	0.943	**0.858**	**0.898**	**0.815**

3) Performance comparison: To study the performance of the proposed method for wood crack detection, WCU-Net is compared with some previous approaches, including U-Net [10], PSPNet [13], DeepLabv3+ [14], DeepCrack [15], improved U-Net [16], Mobile CNN [12]. For fair comparison, all the methods use the same settings to train the network. The results of the methods are given in Table 2. It demonstrates that the proposed WCU-Net model achieves a recall of 85.8% and a IoU of 81.5%, outperforming all the previous methods. From the perspective of both the precision and recall, the proposed method obtained the highest F1-score of 89.8%. The results prove the effectiveness of our proposed approach.

Table 2. Results of Various Methods for Wood Crack Detection

Method	Pr	Re	F1	IoU
U-Net [10]	**0.951**	0.791	0.864	0.760
PSPNet [13]	0.841	0.757	0.797	0.662
DeepLabV3+ [14]	0.938	0.832	0.882	0.788
DeepCrack [15]	0.944	0.809	0.871	0.772
Improved U-Net [16]	0.944	0.680	0.790	0.653
Mobile CNN [12]	0.944	0.825	0.880	0.786
WCU-Net	0.943	**0.858**	**0.898**	**0.815**

To visually compare the proposed WCU-Net with previous methods, the predictions of U-Net, DeepLabv3+, Mobile CNN and WCU-Net are conducted on the wood crack dataset. Fig. 6 presents samples of predictions with corresponding original images and ground truth.

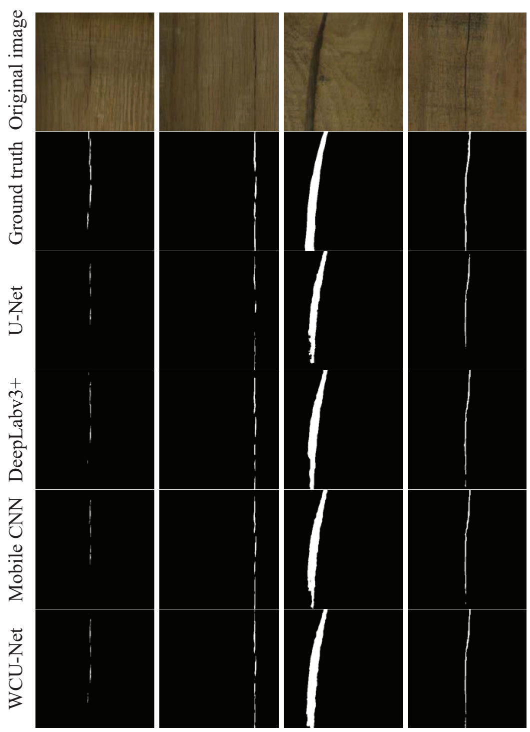

Fig. 6. Comparison of the results of the proposed WCU-Net with the three methods on the wood crack dataset.

It demonstrates that our WCU-Net realizes more accurate segmentation results compared with other methods. Particularly, our proposed method can detect the cracks with more details, which further indicates that the PAM can focus on the wood crack areas and the FEM enhances more detailed wood crack information. Meanwhile, the cracks detected by the WCU-Net are more continuous. This is because the adopted residual block provides a larger receptive field.

Conclusion: In this letter, the WCU-Net is proposed for wood crack detection. Specifically, a position attention mechanism is firstly proposed to exaggerate the wood crack positions, then the feature enhancement mechanism is designed to selectively derive more diluted information of tiny crack. Moreover, a residual block is adopted to obtain and fuse multi-scale receptive fields for finding more crack areas that are similar to the sound region. The experimental results show that WCU-Net can improve the detection accuracy.

Acknowledgments: This work was supported by the National Key Research and Development Program of China (2018YFB1702100), and the open project of Key Laboratory of System Control and Information Processing, Ministry of Education, China.

References

[1] K. Gu, Y. Zhang, and J. Qiao, "Vision-based monitoring of flare soot," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 7136–7145, 2020.

[2] G. A. Ruz, P. A. Estevez, and C. A. Perez, "A neurofuzzy color image segmentation method for wood surface defect detection," *For. Prod. J.*, vol. 55, no. 4, pp. 52–58, 2005.

[3] W. Luo and L. Sun, " An improved binarization algorithm of wood image defect segmentation based on non-uniform background," *J. For. Res.*, vol. 30, no. 4, pp. 1527–1533, 2019.

[4] Z. Chang, J. Cao, and Y. Zhang, "A novel image segmentation approach for wood plate surface defect classification through convex optimization," *J. For. Res.*, vol. 29, no. 6, pp. 1789–1795, 2018.

[5] Q. Chen, Z. Liu, X. Ma, and Y. Wang, " Artificial neural correlation analysis for performance-indicator-related nonlinear process monitoring," *IEEE Trans. Ind. Informat.*, vol. 18, no. 2, pp. 1039–1049, 2021.

[6] F. Li, T. Zheng, N. He, and Q. Cao, "Data-driven hybrid neural fuzzy network and ARX modeling approach to practical industrial process identification," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 9, pp. 1702–1705, 2022.

[7] 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, p. 107357, 2020.

[8] 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, 2020.

[9] Y. Lin, D. Chen, S. Liang, Z. Xu, Y. Qiu, J. Zhang, and X. Liu, "Color classification of wooden boards based on machine vision and the clustering algorithm," *Appl. Sci.*, vol. 10, no. 19, p. 6816, 2020.

[10] O. Ronneberger, P. Fischer, and T. Brox, " U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*, Springer, 2015, pp. 234–241.

[11] K. Liu, Z. Ye, H. Guo, D. Cao, L. Chen, and F.-Y. Wang, "Fiss GAN: A generative adversarial network for foggy image semantic segmentation," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 8, pp. 1428–1439, 2021.

[12] G. Doğan and B. Ergen, "A new mobile convolutional neural network-based approach for pixel-wise road surface crack detection," *Measurement*, vol. 195, p. 111119, 2022.

[13] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, " Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 2881–2890.

[14] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 801–818.

[15] Y. Liu, J. Yao, X. Lu, R. Xie, and L. Li, " Deepcrack: A deep hierarchical feature learning architecture for crack segmentation," *Neurocomputing*, vol. 338, pp. 139–153, 2019.

[16] L. Yang, H. Wang, B. Huo, F. Li, and Y. Liu, "An automatic welding defect location algorithm based on deep learning," *NDT E Int.*, vol. 120, p. 102435, 2021.