An Automatic Detection and Online Quality Inspection Method for Workpiece Surface Cracks based on Machine Vision

Cuili Mao\textsuperscript{1}\textsuperscript{*}, Wen Ma\textsuperscript{2}

\textsuperscript{1}School of Intelligent Manufacturing, Nanyang Institute of Technology, Nanyang 473004, China
\textsuperscript{2}Xuefeng middle school, Nanyang 474250, China

\textsuperscript{*}Email: CuiliMMao@163.com

Received: June 27, 2021. Revised: December 12, 2021. Accepted: January 5, 2022. Published: January 7, 2022.

Abstract—The wide application of intelligent manufacturing technologies imposes higher requirements for the quality inspection of industrial products; however, the existing industrial product quality inspection methods generally have a few shortcomings such as requiring many inspectors, too complicated methods, difficulty in realizing standardized monitoring, and the low inspection efficiency, etc. Targeting at these problems, this paper proposed an automatic detection and online quality inspection method for workpiece surface cracks based on the machine vision technology. At first, it proposed a vision-field environment calibration method, gave the specific method for workpiece shape feature recognition and size measurement based on machine vision, and achieved the on-line monitoring of workpiece quality problems such as feature defects and size deviations. Then, this study integrated the multi-scale attention module and the up-sampling module that can restore the locations of image pixels based on the high-level and low-level hybrid feature maps, built a workpiece crack extraction network, and realized workpiece crack feature extraction, crack type classification, and damage degree division. At last, experimental results verified the effectiveness of the proposed method, and this paper provided a reference for the application of machine vision technology in other fields.

Keywords—Machine vision, Workpiece surface cracks, Automatic detection, Online quality inspection

I. INTRODUCTION

With the implementation and progress of the Made-in-China 2025 action plan, now the advanced intelligent manufacturing technologies are providing the various links of industrial production with special solutions such as robotics, 3D printing, artificial intelligence, Internet of Things, cloud computing, and big data, etc., and there are a lot of successful cases already [1-5]. The fast and stable intelligent manufacturing has imposed higher requirements for the quality inspection of industrial products. The ideas of traditional industrial product quality inspection methods are mostly to use special instruments to obtain the information of the shape and size of the products, and match them with the information of standard products [6-8], however, such ideas would require a lot of inspectors, the inspection methods are extremely complex, it’s difficult to realize standardized monitoring, and the inspection efficiency is very low, and all of these problems need prompt solutions [9-12].

The existing studies are combed and summarized as follows: Aiming at the problems with manual online workpiece quality inspection such as the low inspection efficiency, inconsistent quality evaluation standards, and the inspection results are easily interfered [13-17], Sevillano et al. [18] gave a design of the software and hardware of an online workpiece quality inspection system and done works such the camera model selection, illumination mode design, centerline positioning, and 2D and 3D crack detection algorithm design, etc. Nandakumar and Shankar [19] proposed an image denoising algorithm for the online quality inspection of large-scale and complex structures, in their study, the shape recognition of the structures to be inspected was realized by an edge-based template matching algorithm, and the size measurement of the structures was realized by a bounding rectangle-based fitting method. Prawin and Rama Mohan Rao [20] introduced the machine vision technology that is usually used for non-contact inspections into the PCB manufacturing process of the electronics industry; aiming at the characteristics of PCB boards such as many components, high density, complex circuits, and hard to be detected, they summarized the common defects of PCB boards and chose the double peak method and the eight-direction connection method to realize the separation of
the board images and the detection of short circuit and bumps; as for stains, upside-down installation, and missing soldering, the regional statistics method and the template matching algorithm were adopted for inspection. Rodríguez et al. [21] employed the ultrasonic guided waves that can detect the cross-section of an entire rail from a long distance away to detect railway cracks; based on the analysis of the modal dispersion characteristics and vibration features of the ultrasonic guided waves in the rail, the obtained modal shape information was converted into RGB images, in this way, the complicated analysis problems had been simplified. Nguyen et al. [22] clustered the high-frequency complex modalities of steel rails and optimized the detection modal excitation mode based on image processing; aiming at the problems that the cracks cannot be subdivided and located and it’s difficult to distinguish the echoes of different modalities, they adopted a 3D-array guided wave signal analysis method to extract the single modality of the steel rails; at last, they realized the accurate identification and location of rail cracks based on the real-time group velocity calibration method. For the micro-cracks of concrete that appeared in the early stage of concrete material molding due to the impact of complex internal and external factors, Wang et al. [23] proposed a non-destructive detection method based on nonlinear acoustic field modulation; then, for concrete models with different micro-crack extension directions, they constructed and solved the nonlinear wave equations based on broadband excitation, and used experiments to verify that the defined damage factor had a certain sensitivity to cracks and it’s of certain feasibility to take it as the basis for identifying concrete micro-cracks. Knitter-Piątkowska and Gumiński [24] verified the pros and cons of four common metal surface crack detection methods in signal-to-noise ratio performance through comparative experiments, then, combining with an electromagnetic thermal coupling analysis, they analyzed the influencing factors such as the thickness of the metal surface under anticorrosive paint layer and uneven excitation; at last, based on the Fourier transform results of the temperature signal of the measured area, they performed simulation experiments on the relationship between amplitude and phase when cracks exist on the surface of the concrete. Ayachi Errachdi et al., Hira Lal Gope et al., and Bacha Sawssen et al. [25-27] performed adaptive internal model neural networks control for nonlinear system, carried out classification based on the extreme learning machine (ELM) for Gaussian radial basis, and classified breast images with kernel principal component analysis (KPCA).

It can be inferred from the existing studies that, In recent years, more and higher technical requirements have been imposed on the detection of industrial product shape defects, size deviations, scratches and cracks, and other quality inspection scenarios, however, the traditional quality inspection methods cannot realize the non-contact standardized inspection requirements while maintaining high inspection efficiency. To this end, this study proposed an automatic detection and online quality inspection method for workpiece surface cracks based on the machine vision technology. This paper aims to change the traditional manual quality inspection of large and complex workpieces, and facilitate the application of machine vision in the quality inspection of such workpieces. The proposed machine vision approach can inspect the quality of large and complex workpieces in an automatic and adaptive manner, and meet the needs of the production of such workpieces.

The content of this paper has the following aspects: the second chapter completed the vision-field environment calibration, realized shape feature recognition and size measurement of workpieces based on machine vision, and achieved online monitoring workpiece quality problems such as shape defects and size deviations; the third chapter integrated the multi-scale attention module and the up-sampling module that can restore the locations of image pixels based on the high-level and low-level hybrid feature maps, built a workpiece crack extraction network, and realized workpiece crack feature extraction, crack type classification, and damage degree division. At last, experimental results proved the effectiveness of the proposed method. The research results promote the development of the quality inspection techniques for large and complex workpieces, and enable enterprises to adopt the modern management model of information, which is centralized and highly efficient.

II. WORKPIECE SHAPE RECOGNITION AND SIZE MEASUREMENT BASED ON MACHINE VISION

A. Vision-field environment calibration

The vision-field environment calibration processes involve the transformation of the vision coordinate system, the principle of photo perspective imaging, and the multi-objective optimization strategy, and they are the basis for the automatic surface crack detection and online quality inspection of workpieces based on the technology of machine vision. Figure 1 shows the principle of the transformation of pixel coordinate system \((P_{ox}, x, y)\), image coordinate system \((P_a, a_x, b_x)\), camera coordinate system \((P_{ca}, a_{ca}, b_{ca}, c_{ca})\), and world coordinate system \((P_{wo}, a_{wo}, b_{wo}, c_{wo})\). Suppose \(E_1, E_2, \) and \(FO\) respectively represent the internal parameter matrix, external parameter matrix and focal length of the camera; \(RO\) represents the rotation matrix from \((P_{ca}, a_{ca}, b_{ca}, c_{ca})\) to \((P_{wo}, a_{wo}, b_{wo}, c_{wo})\); and \(DE\) represents the corresponding deviation matrix; for a known pixel coordinate \((x_p, y_p)\) of the image taken by the camera, the corresponding length values are represented by \(SL_a\) and \(SL_b\), then the transformation of \((P_{ox}, x, y)\) and \((P_{wo}, a_{wo}, b_{wo}, c_{wo})\) can be expressed by Formula 1:

\[
\begin{bmatrix}
0 & 0 & 0 & x_p \\
0 & 0 & 0 & y_p \\
0 & 0 & 0 & 1
\end{bmatrix}
= \begin{bmatrix}
0 & 0 & 0 & x_p \\
0 & 0 & 0 & y_p \\
0 & 0 & 0 & 1
\end{bmatrix}
(1)
\]

\[
= E_1 E_2 E_3
\]
Suppose: for the transformation from \((P_{wo}, a_{wo}, b_{wo}, c_{wo})\) to \((P_{ca}, a_{ca}, b_{ca}, c_{ca})\), \(\delta\) represents the Euler angle of the rotation of corresponding points, then \(RO\) can be represented by Formula 2:

\[
RO = \begin{bmatrix}
\delta_1 \\
\delta_2 \\
\delta_3 \\
\delta_4 \\
\delta_5 \\
\delta_6
\end{bmatrix}
\]  

(2)

Suppose: for the transformation from \((P_{wo}, a_{wo}, b_{wo}, c_{wo})\) to \((P_{ca}, a_{ca}, b_{ca}, c_{ca})\), \(DE_a\), \(DE_b\), and \(DE_c\) respectively represent the translations of corresponding points on \(a\), \(b\), and \(c\) axes, then \(DE\) can be represented by Formula 3:

\[
DE = [DE_a \ DE_b \ DE_c]
\]  

(3)

The camera was calibrated as follows: according to the camera’s radial constraints and the relationship between \((P_{wo}, a_{wo}, b_{wo}, c_{wo})\) and \((P_{ca}, a_{ca}, b_{ca}, c_{ca})\), there is:

\[
x_j = \begin{bmatrix}
a_{wo,j} \ y_j \\
b_{wo,j} \ y_j \\
c_{wo,j} \ x_j \\
a_{wo,j} \ y_j \\
b_{wo,j} \ y_j \\
c_{wo,j} \ x_j
\end{bmatrix}
\]  

(4)

Based on Formula 4, the camera was calibrated based on \(E\) non-coplanar points \((a_{wo,j}, b_{wo,j}, c_{wo,j})\) corresponding to pixels \((x_j, y_j)\) on the reference object that had been calibrated, wherein the value range of \(j\) was \([1, E]\). The least square method was adopted to solve Formula 2 with the coordinates of the \(E\) points being substituted into the equation, then the values of unknown parameters \(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, DE_a, DE_b,\) and \(\lambda\) could be obtained.

By solving Formula 5, the focal length \(FO\) of the camera could be obtained:

\[
\begin{align}
&x_j \lambda / DE_b \\
&y_j \lambda / DE_b
\end{align}
\]  

(5)

By solving Formula 6, the \(DE_{ca}\) in \(DE\) could be obtained:

\[
\begin{align}
&G_a = \delta_1 a_{wo,j} + \delta_2 b_{wo,j} + DE_a \\
&G_b = \delta_3 a_{wo,j} + \delta_4 b_{wo,j} + DE_b \\
&Q = \delta_5 a_{wo,j} + \delta_6 b_{wo,j}
\end{align}
\]  

(6)

where, \(l\) is the distortion coefficient of camera imaging. Because \(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, DE_a, DE_b,\) and \(\lambda\) were all known parameters, again the least square method was adopted to solve Formulas 5 and 6 with the coordinates of the \(E\) points being substituted into the equation, then, the values of \(FO, l,\) and \(DE\) could be obtained. Obtaining reliable internal and external parameters of the camera provides a basis for calculating the position information of image pixels in a 3D environment.

B. Workpiece shape feature recognition and size measurement

To better obtain the features of the shape, color, and texture of the workpiece in the image, the edges of the image need to be detected in the first place. At first, the threshold value was determined according to the change range of pixel gray scale, and the judgement criterion of edge pixels was that whether the gray scale change is greater than the threshold value or not. For a 2D image \(F(a, b)\) of a workpiece, the gradient at the pixel coordinate \((a, b)\) can be calculated by Formula 7:

\[
\nabla F(a, b) = \begin{bmatrix} H_a \\ H_b \end{bmatrix} = \begin{bmatrix} \partial F / \partial a \\ \partial F / \partial b \end{bmatrix}
\]  

(7)

The corresponding gradient amplitude can be calculated by Formula 8:

\[
|\nabla F(a, b)| = |H_a| + |H_b|
\]  

(8)

Figure 1 shows the execution process of workpiece crack extraction and quality inspection method proposed in this study. In this paper, the Canny operator, which has higher accuracy in positioning the workpiece edge than Roberts, Sobel, Prewitt, and Laplace operators, was selected to extract the edge features of the large and complex workpiece. The specific steps of template creation are:

1. Perform Gaussian filtering on the original image containing the target workpiece;
2. Extract significant feature areas that matched the target workpiece in the image to obtain \(F^*(a, b)\). Through the subtraction operation shown as Formula 9, the image of the target workpiece was obtained:

\[
D(a, b) = F(a, b) - F^*(a, b)
\]  

(9)

3. In order to achieve the matching between the significant feature areas of the target workpiece and the template under different resolutions, it is necessary to construct an image pyramid containing the target workpiece. The steps for creating a Gaussian pyramid that can be used to create template images with reduced features are as follows: first, the original image after Gaussian filtering is down-sampled with every other rows and columns, and the obtained image is taken as the second layer of the pyramid; Then, repeat the filtering and down-sampling operations on the image of this layer, and the obtained image is taken as the third layer of the pyramid; after that, follow the same steps and repeat the operations, images of other layers could also be obtained until the top of the pyramid.
Suppose: $F_j(a,b)$ represents the $j$-th layer image of the Gaussian pyramid with $\varepsilon$ layers; $W$ and $H$ represent the size of the image, and the size of the adopted window function $q(w, h)$ is $(2d+1)\times(2d+1)$, then the down-sampling process of the image can be expressed by Formula 10:

$$F_j(a,b) = \sum_{w=d}^{d} \sum_{h=d}^{d} q(w,h) F_{j+1}(2a+w,2b+h) \quad (10)$$

$$1 \leq j \leq \frac{W}{2^\varepsilon}, 0 \leq i \leq \frac{H}{2^\varepsilon}, 1 \leq j \leq \varepsilon$$

Next, the shape of the workpiece was recognized according to the difference between the shape features of the workpiece image $TP_j(w,h)$ to be detected and the matching template $DE_j(w,h)$. Suppose $ME(j,i)$ represents the matching error, $TH_\varepsilon$ represents the judgment threshold, when $ME(j,i) > TH_\varepsilon$, it’s judged that the features do not match template; when $ME(j,i) \leq TH_\varepsilon$, it’s judged that the features match the template. The matching degree of features can be calculated by Formula 11:

$$ME(j,i) = \sum_{w=1}^{W} \sum_{h=1}^{H} TP_j(w,h) - DE_j(w,h) \quad (11)$$

The matching judgment criterion can be expressed by Formula 12:

$$\begin{cases} TP_j(w,h) = DE_j(w,h), & ME(j,i) \leq TH \\ TP_j(w,h) \neq DE_j(w,h), & ME(j,i) > TH \end{cases} \quad (12)$$

Generally, it is quite difficult to determine the length and width dimensions of a workpiece with irregular shape features. In order to obtain more accurate feature size measurement results, this paper adopted an edge bounding rectangle fitting method with simpler operations to measure the workpiece. That is, starting from the end point extracted from the edge of the workpiece, scan the boundary pixels of the workpiece image line by line to the left, right, down, and up directions, and determine the maximum and minimum edge pixel coordinates $a'_{\text{min}}, a'_{\text{max}}, b'_{\text{min}},$ and $b'_{\text{max}}$, thereby further obtain the bounding rectangle of the edge contour of the target workpiece. To reduce the amount of computation of size measurement, the four corner points of the bounding rectangle can be used to determine the size of the shape feature image of the workpiece, then, the camera calibration results obtained in previous texts were used to convert the length $J_{W1}$ and width $J_{H1}$ into the actual size values of the shape features of the workpiece, which were represents by $J_{W2}$ and $J_{H2}$, respectively.

### III. Feature Extraction and Classification of Workpiece Cracks Based on Machine Vision

#### A. Construction of the multi-scale attention module

After the shape features of the workpiece were recognized and the size of the workpiece was measured, the crack features could be extracted. This paper establishes a neural network based on the multi-scale dilated convolution module to acquire richer context information from workpiece crack images. Then, a new up-sampling method was proposed, using feature mapping of different resolutions. Finally, the multi-scale attention module was introduced to identify and classify the workpieces with cracks or other surface quality issues.

Figure 2 gives the structure of the constructed convolutional neural network. It can be seen from the figure that, the structural details of convolution layers and pooling layers in the CNN, including kernel size and kernel number, can be adjusted as per actual working conditions. The neural network that introduced the multi-scale dilated convolution module can identify the size of the cracks on the surface of the workpiece and obtain rich semantic information of the crack feature images, and it has very good performance in crack classification.

![Fig. 2 Structure of the constructed convolutional neural network](image)

Suppose $\eta_l$ represents the dilation rate of the convolution kernel $CL_x$ that is used to characterize the number of zeros that can be accommodated between pixels, then for the convolution kernel $CL_x$ with a size of $u \times u$ in the $l$-th layer, the size of its sensible area could be calculated by Formula 13:

$$BR_{\eta_l} = l + (l - 1) \times (\eta_l - 1) \quad (13)$$

According to above formula, since the adopted dilation...
operation had only selected \(w \times u\) pixels for convolution calculation, as the computation amount was decreased, the crack sensible area was increased to a certain extent. In order to obtain more abundant semantic information about the cracks of the workpiece, a multi-scale dilated convolution module could be constructed, that is, to merge the convolutional layers of multiple scales and different convolution kernels and dilation rates, in this way, the crack features at different positions could be expressed in combinations. Figure 3 shows the flow of multi-scale dilated convolution.

Figure 4 gives the structure of the multi-scale dilated attention module. The significant crack features can be effectively extracted by capturing the long-distance contextual semantic information through the global maximum pooling operation. Suppose \(CH=[CH_1, CH_2, \ldots, CH_s]\) represents the high-level features of \(\mathbb{R}^{W \times H \times N}\), \(N\) represents the total number of channels, and \(PO_{\max}\) represents the global maximum pooling operation, then the channel feature vector can be obtained by Formula 14:

\[
FV' = PO_{\max}(CH') = \arg\max_{a_j} CH\left(CH'(j,i)\right)
\]  

(14)

The batch normalization (BN) can make full use of the correlations of image data. Suppose \(a_j, E(a_j),\) and \(V(a_j)\) respectively represent the training sample in the \(j\)-th channel and its expected value and variance value; \(\omega_j\) and \(\xi_j\) represent the connection weight and threshold of the network learning parameter; and \(\sigma\) represents the minimum value that is used to avoid the variance value of 0, then the BN process can be represented by Formula 15:

\[
B_j = \omega_j \frac{a_j - E(a_j)}{\sqrt{V(a_j + \sigma)}} + \xi_j
\]  

(15)

Substitute the \(D\) groups of normalized image data obtained through BN operation into the \(sigmoid\) activation function to complete the mapping of the encoding feature channel vector to \([0, 1]\):

\[
NOR = MAP(FV', w) = sigmoid(b)
\]  

(16)

Perform \(NOR\) weighting on \(CH\), that is, the crack feature map that can be finally output is:

\[
CH'' = NOR \cdot CH'
\]  

(17)

Accurate crack location mapping can be achieved by introducing standard convolutional layers with different sizes of convolution kernels into different branch structures. Suppose \(I[j]\) and \(O[j]\) respectively represent the input and output feature maps, \(FI[l]\) represents the filter of length \(J_{W,1}\), then the dilated convolution of the 2D image can be described by Formula 18:

\[
O[j] = \sum_{l=1}^{J_{W,1}} I[j + \eta \cdot l]FI[l]
\]  

(18)

Suppose \(CI\) represents the input feature map, \(L_i\) represents the linear mapping used to match the input dimensions, then, the output feature maps \(CH_1, CH_2,\) and \(CH_3\) under all connected branches were calculated by the \(ReLU\) activation function:

\[
CH = max\left[0, concat(CH_1, CH_2, CH_3) + L_i CI\right]
\]  

(19)

During the encoding phase, the multi-scale dilated convolution module can convert the input image into rich semantic visual features. However, the spatial resolution of these features is rather coarse. An up-sampling module can be configured to restore these features to the resolution of the input image, laying the basis for predicting the spatial distribution of cracks. Inspired by the up-sampling module in Deeplabv3+, the our up-sampling module mainly contains two inputs: low-resolution features with high-level semantic information, and high-resolution features in the shallow network. In this way, local and global context information can be obtained from the features extracted at different scales.
B. Construction of workpiece crack feature extraction network

![Diagram of workpiece crack feature extraction network]

The multi-scale attention module built in previous text and the up-sampling module that can restore the locations of image pixels based on the high-level and low-level hybrid feature maps were integrated to construct the workpiece crack feature extraction network. Figure 5 gives the structure of the up-sampling module. The training process of the network was divided into two parts: the forward propagation used to calculate the predicted crack type classification results under the current parameters and the back propagation used to update the training parameters. The purpose of the network training was to make the difference between the actual and the expected workpiece crack classification results as small as possible.

In the training process, this paper took the Generalized Dice Loss (GDL) function as the loss function of the network to calculate the deviation between the actual and the expected workpiece crack classification results. Suppose $PS_m$ and $PL_m$ respectively represent the segmented volume pixel value of the reference foreground of the crack in the workpiece edge and the predicted probability map of the foreground label, $q_i$ represents the weight provided by different label set attributes, then the predicted probability map of the foreground label, reference foreground of the crack in the workpiece edge and the workpiece crack classification results. Suppose $F(q_i; a^{(i)}, b^{(i)})$ represents the segmentation volume pixel value of the reference foreground of the crack in the workpiece edge, the weight $q_i$ can be calculated by Formula 21:

$$q_i = \frac{1}{\left(\sum_{m=1}^{M} PS_{lm}\right)^2}$$  \hspace{1cm} (21)

The Stochastic Gradient Descent (SGD) method was adopted to update the weights to reduce the deviation between the actual and the expected workpiece crack classification results. Calculate the gradient $\nabla_q F_{GDL}$ of the loss function according to the weight $q_i$:

$$\nabla_q F_{GDL}(q; a^{(i)}, b^{(i)}) =$$

$$\frac{1}{m} \sum_{j=1}^{M} \left[ a^{(i)}(j) \left(1 \cdot \delta(b^{(i)}=i) - \delta\left(a^{(i)}=i \mid a^{(i)}; q\right)\right) \right]$$

$$+ \mu q_i$$

$$\text{Then, update the rate } u \text{ based on the momentum parameter } \sigma \text{ and the learning rate } \beta:$$

$$u \leftarrow \sigma - \beta \nabla_q F_{GDL}(q; a^{(i)}, b^{(i)})$$  \hspace{1cm} (23)

At last, update the weight $q_i$:

$$q_i \leftarrow q_i + u$$  \hspace{1cm} (24)

C. Workpiece crack type classification and damage degree division

According to the law of spatial distribution, the surface cracks of the workpiece can be classified into four types: horizontal cracks, vertical cracks, blocky cracks, and honeycomb cracks. According to the average width of the crack area and the spacing between branches, this study completed the division and evaluation of the damage degree of the workpiece while extracting the crack features. Based on workpiece samples with multiple types of surface cracks, this paper quantitatively analyzed the differences between different crack features. Figure 6 shows the principle of identifying and classifying cracks based on crack features.

Compared with horizontal and vertical cracks, blocky cracks have more complicated spatial distribution and its feature extraction steps require an extra link of the merged inspection and feature combination of adjacent cracks to make an overall judgement on the crack. This link needs to use connected domain analysis to perform sub crack target division on the binary image after the crack features are extracted, then, based on each sub-target, generate the corresponding minimum bounding rectangle including coordinates $(a, b)$ and length $w$ and width $h$. The average distance between crack branches corresponding to each bounding rectangle can be expressed by Formula 25:

$$\frac{\sum_{m=1}^{M} DIS_m}{M} = DIS_1 + DIS_2 + DIS_3$$

The honeycomb cracks and blocky cracks on the surface of the workpiece can be directly judged by Formula 25. When the number of branches between two sub-targets is greater than the set threshold, the two sub-targets and the branches between them are merged and the feature combination was formed, meanwhile, it’s judged to be a blocky crack that is related to adjacent cracks. If the branch number is less than the set threshold, it is judged to be an independent linear crack, then, according to the size of the angle between the diagonal of the bounding rectangle and the horizontal direction, whether the crack is a horizontal or a vertical crack is further judged. The entire judgement process is carried out in cycles of continuously updating the sub-targets of the crack to be determined, and the uninterrupted execution of the merging operation is performed until it is judged that all bounding rectangles are not adjacent.
According to different types of surface cracks, the damage of cracks was divided into three degrees: slight, moderate, and severe. The criterion was the average crack width or the width of blocky crack $D$. Since it's quite difficult to quantify the width of cracks at different positions, suppose $CR_I$ and $CR_S$ represent the number of crack pixels in the image and in the skeleton, $SR_W$ represents the spatial resolution of the workpiece image, then the average crack width can be calculated by Formula 26:

$$CR_{IS} = \frac{CR_I}{CR_S} \cdot SR_W$$  \hspace{1cm} (26)

Using the coordinate $P (A, B, W, H)$ of the minimum bounding rectangle, the equivalent area of the crack could be calculated; suppose $EF_W$ represents the influenced width, Formula 27 gives the calculation formula for the area of a linear crack:

$$S_{LF} = \sqrt{W^2 + H^2} \cdot EF_W \cdot CR_W^2$$  \hspace{1cm} (27)

As for blocky cracks and honeycomb cracks with higher damage degrees, the area could be obtained by approximating the overall area:

$$S_T = W \cdot H \cdot CR^2$$  \hspace{1cm} (28)

**IV. EXPERIMENTAL RESULTS AND ANALYSIS**

To comparatively analyze the effectiveness of the preliminary processing method of workpiece images, comparative experiments were designed to test the network training performance and the crack recognition accuracy. The data samples are 5,100 images on surface cracks of workpieces. They were divided into a training set and a test set by the ratio of 3:1.

Figures 7 and 8 respectively show the loss function curves of the constructed network model before and after the improvement, and the P-R curves of the constructed network model before and after the improvement. According to the comparative analysis of the figures, after the multi-scale dilated convolution module had been introduced, the model's recognition accuracy of workpiece deformation, cracks and other quality problems had increased by 2.13%, and the training time cost had been reduced by more than a half. The introduction of Gaussian filtering had also adjusted and optimized the accuracy of crack classification to a certain extent.

In order to further verify the performance of the proposed algorithm in terms of workpiece edge extraction, Table 1 compares the test results of different edge extraction methods, including our algorithm, and the edge extraction operators such as Roberts, Sobel, Prewitt, Laplace, and Canny. According to the data in the table, the proposed algorithm achieved higher scores in accuracy, recall rate, F-score, and mean intersection over union, reaching 99.16%, 96.23%, 96.51% and 76.09%, respectively. When detecting images with a resolution of 2480×1560, the workpiece edge extraction speed of the method.

<table>
<thead>
<tr>
<th>Method</th>
<th>P</th>
<th>R</th>
<th>F-score</th>
<th>Mean intersection over union</th>
<th>Time cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roberts</td>
<td>71.25</td>
<td>71.18</td>
<td>72.63</td>
<td>43.27</td>
<td>4.87s</td>
</tr>
<tr>
<td>Sobel</td>
<td>93.73</td>
<td>93.36</td>
<td>94.06</td>
<td>61.39</td>
<td>3.15s</td>
</tr>
<tr>
<td>Prewitt</td>
<td>92.32</td>
<td>92.67</td>
<td>93.18</td>
<td>69.12</td>
<td>0.86s</td>
</tr>
<tr>
<td>Laplace</td>
<td>97.01</td>
<td>95.91</td>
<td>97.67</td>
<td>73.51</td>
<td>1.32s</td>
</tr>
<tr>
<td>Canny</td>
<td>94.14</td>
<td>92.53</td>
<td>98.42</td>
<td>70.23</td>
<td>0.65s</td>
</tr>
<tr>
<td>Proposed method</td>
<td>99.16</td>
<td>96.23</td>
<td>96.51</td>
<td>76.09</td>
<td>0.72s</td>
</tr>
</tbody>
</table>

Fig. 7 Curves of loss function of the constructed network model before and after the improvement

Fig. 8 P-R curves of the constructed network model before and after the improvement

Table 1. Comparison of experimental results of different edge extraction methods
The proposed algorithm was 0.72s per image, while the speeds of other algorithms were slower.

Fig. 9 Curves of mean intersection over union under different dilation modes

For the multi-scale dilated convolution module, comparative experiments were designed in this study and 12-times up-sampling was performed on the network feature mapping, the purpose was to verify the positive impact of the constructed module on the effect of crack feature extraction. Figure 9 shows the curves of the average intersection over union under different dilation modes. After 50 iterations, the values of the average intersection over union of the dilation modes including baseline dilation and spatial pyramid pooling could all reach the stable state, but the proposed model that adopted the up-sampling module and the multi-scale attention module achieved the best results. Figure 10 shows the comparison of the workpiece crack extraction effect under different dilation modes.

Based on the extracted workpiece cracks, all the crack sub-targets were judged and classified by the algorithm proposed in Subsection 3.3. Then, based on the classification results, as well as the crack width, crack area, and branch density data, the damage degree of the workpiece was judged. Figure 11 gives the calculation results of crack area. Table 2 gives the classification results of 11 groups of cracks, Table 3 gives the division results of the damage degree of the workpiece.

According to Table 3, for the simple horizontal or vertical linear cracks, the damage degree of the workpiece can be judged by the width of the cracks. 500 sample images of workpiece surface cracks were used to verify the workpiece damage degree division algorithm, and Table 3 gives the division results of workpiece damage degree. According to the data in the table, the performance of the proposed algorithm in damage degree division was relatively ideal. For blocky, honeycomb, horizontal, and vertical cracks, the correct rates reached 86.7%, 83.1%, 93.2% and 94.5%. The recognition rates of chunk cracks and fractures were slightly low, mainly due to the selection of the threshold for crack spacing.

In summary, our method can minimize the incorrect extraction or omission of cracks through relatively few processes. The results verify the high robustness of our model, which integrates multi-scale dilated convolution module and attention mechanism in the encoding phase. The integrated module captures context information on multiple scales, and realizes accurate extraction of features. Based on the rich semantic information, low-level high-resolution features are used to restore the boundaries of targets, making it possible to extract image cracks more accurately under complex backgrounds, and to segment the image cracks more precisely.

Table 2. Crack classification results

<table>
<thead>
<tr>
<th>Crack No.</th>
<th>Number of branches</th>
<th>Angle</th>
<th>Average distance</th>
<th>Width</th>
<th>Type</th>
<th>Damage degree</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>1</td>
<td>6.32</td>
<td>/</td>
<td>8.9</td>
<td>Horizontal</td>
<td>Severe</td>
<td>0.32s</td>
</tr>
<tr>
<td>1-1</td>
<td>1</td>
<td>82.31</td>
<td>/</td>
<td>7.5</td>
<td>Vertical</td>
<td>Severe</td>
<td>0.33s</td>
</tr>
<tr>
<td>2-2</td>
<td>1</td>
<td>73.57</td>
<td>/</td>
<td>7.5</td>
<td>Horizontal</td>
<td>Slight</td>
<td>0.35s</td>
</tr>
<tr>
<td>2-3</td>
<td>1</td>
<td>17.28</td>
<td>/</td>
<td>7.6</td>
<td>Vertical</td>
<td>Severe</td>
<td>0.33s</td>
</tr>
<tr>
<td>2-4</td>
<td>1</td>
<td>75.64</td>
<td>/</td>
<td>21.8</td>
<td>Horizontal</td>
<td>Severe</td>
<td>0.32s</td>
</tr>
<tr>
<td>3-1</td>
<td>3</td>
<td>/</td>
<td>1.27</td>
<td>21.5</td>
<td>Blocky</td>
<td>Slight</td>
<td>0.37s</td>
</tr>
</tbody>
</table>
This study realized automatic detection and online quality inspection of workpiece surface cracks in the industrial production field based on machine vision technology. In order to achieve online monitoring of workpiece quality problems such as shape defects and size deviations, firstly this paper proposed the vision-field environment calibration method and gave the specific method for workpiece shape feature recognition and size measurement based on machine vision; then, it built and integrated the multi-scale attention module and the up-sampling module that can restore the locations of image pixels based on the high-level and low-level hybrid feature maps; after that, based on the prediction and classification results of the workpiece crack feature network, the study performed workpiece crack feature extraction, crack type classification, and damage degree division; at last, comparative experiments were designed and performed to verify the training performance and crack recognition accuracy of the constructed network; specifically, the experiments compare the P-R curves of the constructed network model before and after improvement, and contrast the results of different edge extraction methods, revealing the superiority of our algorithm in edge extraction from workpieces; besides, different dilation modes were compared in terms of the average intersection over union, and the workpiece crack extraction effect, which verifies the scientific nature of integrating multi-scale attention module and the up-sampling module; the ability of our algorithm to evaluate the degree of damage was proved by the calculation results on cracked area and the classification of the degree of workpiece damage.

ACKNOWLEDGMENT

This paper was supported by National Natural Science Foundation of China (51875164); Henan Education Department Project (21BS35002); Henan Manufacturing Information Technology Service Center

<table>
<thead>
<tr>
<th>Type</th>
<th>Number of correct predictions</th>
<th>Number of incorrect predictions</th>
<th>Correct rate</th>
<th>Division result of damage degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocky</td>
<td>53</td>
<td>6</td>
<td>86.7%</td>
<td>Slight (15)</td>
</tr>
<tr>
<td>Honeycomb</td>
<td>57</td>
<td>10</td>
<td>83.1%</td>
<td>Slight (16)</td>
</tr>
<tr>
<td>Horizontal</td>
<td>127</td>
<td>6</td>
<td>93.2%</td>
<td>Moderate (32)</td>
</tr>
<tr>
<td>Vertical</td>
<td>135</td>
<td>8</td>
<td>94.5%</td>
<td>Slight (45)</td>
</tr>
</tbody>
</table>

Table 3. Division results of the damage degree of the workpiece

V. CONCLUSION

References


Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)
This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US