Metrol. Meas. Syst., Vol. 30 (2023) No. 3, pp. 531–548
DOI: 10.24425/mms.2023.146425

METROLOGY AND MEASUREMENT SYSTEMS
Index 330930, ISSN 0860-8229
www.metrology.wat.edu.pl

VISUAL DETECTION OF MILLING SURFACE ROUGHNESS BASED ON IMPROVED YOLOV5

Xiao Lv[1], Huaian Yi[1], Runji Fang[1], Shuhua Ai[1], Enhui Lu[2]

*1) School of Mechanical and Control Engineering, Guilin University of Technology, Guilin, 541006,
People's Republic of China* (✉ *yihuaian@126.com*)
2) School of Mechanical Engineering, Yangzhou University, Yangzhou, 225009, People's Republic of China

Abstract

Workpiece surface roughness measurement based on traditional machine vision technology faces numerous problems such as complex index design, poor robustness of the lighting environment, and slow detection speed, which make it unsuitable for industrial production. To address these problems, this paper proposes an improved YOLOv5 method for milling surface roughness detection. This method can automatically extract image features and possesses higher robustness in lighting environments and faster detection speed. We have effectively improved the detection accuracy of the model for workpieces located at different positions by introducing Coordinate Attention (CA). The experimental results demonstrate that this study's improved model achieves accurate surface roughness detection for moving workpieces in an environment with light intensity ranging from 592 to 1060 lux. The average precision of the model on the test set reaches 97.3%, and the detection speed reaches 36 frames per second.

Keywords: Surface roughness, improved Yolov5, detection speed, attentional mechanisms.

1. Introduction

Surface roughness is an important indicator to judge the surface quality of a workpiece. It affects the service life of the workpiece and the stability of the overall equipment during operation, especially in the field of medical and health care, aerospace, electronic equipment, military industry and other high-precision technologies and there are strict requirements for the surface roughness of the workpiece. The non-contact measuring is flexible and provides a vast detection area as advantages. Electronic, optical, and machine vision techniques are the non-contact measurement technologies most frequently utilized in industrial production [1]. Electronic and optical measurement equipment is expensive and susceptible to environmental changes such as light intensity and air humidity. The advantages of the machine vision approach include great efficiency and the ability to be integrated and automated. It transforms picture

Article history: received March 4, 2023; revised May 19, 2023; accepted May 25, 2023; available online September 20, 2023.

information into digital information, which is then processed and analyzed. As a result, research on surface roughness measurement using machine vision has increased recently.

As the method of predicting the surface roughness of workpieces based on first-order texture features ignores the position information between pixels, Gadelmawla *et al.* improve detection accuracy by using the 2D and 3D plots of the grey level co-occurrence matrix [2]. However, the grayscale image is a degraded image and the sensitivity of the designed index to the roughness parameter is weakened. To improve roughness detection accuracy, Yi *et al.* established a mathematical model describing the relation between the surface roughness of the grinding workpiece and the image sharpness by constructing an RGB image sharpness evaluation algorithm based on the color space difference [3]. Zhang *et al.* proposed a chromatic aberration index based on the difference of brightness of the virtual image formed by the reflection of red and green light sources from the workpiece surface, and a mathematical model covering the relation between the workpiece surface roughness and the chromatic aberration index was established using a support vector machine [4]. Considering the influence of illumination angles on roughness detection, Somthong *et al.* obtained the relationship between light source irradiation angle and surface roughness by a coordinate measuring machine and the photometric stereo method, thereby obtaining the optimal lighting conditions for measuring surface roughness [5]. Each of these methods uses artificially designed indexes to measure the surface roughness of a workpiece. Although the prediction accuracy is good, this index-based machine vision method is susceptible to the effects of the imaging environment, including light intensity, specimen position, and shooting angle. For example, the experimental procedures in the literature [3, 4] were performed in a dark environment and kept the light source and specimen position unchanged; the literature [5] had strict and inefficient requirements for image acquisition of the workpiece surface. Hence, the roughness detection techniques based on index design discussed above are not appropriate for use in industrial production settings.

Convolutional neural networks have a wide range of potential applications in image processing, and the 2012 release of Alex-Net solidified their significant role in computer vision [6]. Images are stored in computers as digital matrices and convolutional neural networks are used to automatically extract image features by convolutional operations on the digital matrices by convolutional kernels, and then different images are classified by classifiers. To reduce the complexity and prediction time of the prediction system, Rifai *et al.* used a 10-layer convolutional neural network to classify the workpiece surface roughness [7]. To prevent unnecessary image interference, such as background, on the accuracy of roughness detection, He *et al.* proposed an ROI extraction method based on *random wanderer* (RW) image segmentation to extract target regions and feed them into a convolutional neural network training to evaluate the roughness [8]. Deep AlexCORAL, a milling surface roughness class classification model based on deep migration learning, was proposed by Su *et al.* and uses deep migration learning to reduce the quantity of data needed by the model and the difference in data distribution between the training and test sets [9]. However, all the above methods require global feature extraction and classification of surface roughness levels for the entire image, resulting in long processing time and slow detection speed. Furthermore, these methods can only perform classification detection on individual workpiece images and are unable to perform multi-workpiece detection and workpiece positioning, which limits their practical applicability.

It should be emphasized that robustness to lighting environment and measurement speed are essential for surface roughness detection of workpieces in industrial production. The variation of lighting environment directly affects the quality of surface images and the accuracy of feature extraction by the model, which in turn affects the surface roughness detection results of

*Metrol. Meas. Syst.,*Vol. 30 (2023), No. 3, pp. 531–548
DOI: 10.24425/mms.2023.146425

unknown milling samples. Detection speed is one of the requirements in industrial production as through fast detection, product quality information can be obtained in a timely manner, ensuring product quality stability and consistency, improving production efficiency and reducing production costs.

To achieve a higher detection speed and robustness to lighting environment, this paper proposes a YOLOv5-based milling surface roughness detection method with the addition of the *Coordinate Attention* (CA) mechanism. Compared with traditional machine vision methods that require manual feature extraction, this model has better lighting robustness and detection speed. During the training phase, the collected dataset of milling workpieces undergoes preprocessing and data augmentation, and then the model is trained on the processed data to learn the image features of workpieces with different surface roughness categories. In the testing phase, the model is capable of accurately identifying the surface roughness categories of milling workpieces under different light intensities. In addition, after comparing with Faster RCNN (*Region-based Convolutional Neural Network*), *Single-Shot Multibox Detection* (SSD), and the original YOLOv5 algorithm [10–12], it was found that the improved YOLOv5 model has advantages as it comes to both volume and detection speed, and the detection speed can reach 36 frames per second.

2. YOLOv5 algorithm and improvement

2.1. YOLOv5 algorithm

The detection principle of the surface roughness detection model based on the improved YOLOv5 is to divide the input surface image of the milling workpiece into a grid of cells. If the center of the milling workpiece falls within a certain grid cell, the network predicts its roughness level. The model consists of four parts: Input, Backbone, Neck, and Head [13]. Here is a brief introduction to these four parts.

Figure 1 shows the Yolov5 algorithm's overall structure, which is divided into four sections.

1. Input. In order to reduce computational complexity, increase the size of the training set, and improve the model's generalization ability, we preprocessed the milling workpiece images in the training set, including operations such as size reduction, flipping, cropping, rotation, and contrast adjustment. We set anchor for the size of the milling workpiece in the image, which is used to generate the predicted bounding boxes and ground truth boxes, and to calculate the difference between them. By updating the size of the predicted boxes in reverse, we further improve the detection performance of the model.
2. Backbone. We used CSPDarknet53 as the backbone network of our model to extract features automatically from the surface images of the milling workpiece. We used the Focus Principle in the network to create low-resolution feature maps by slicing and concatenating high-resolution milling images, as shown in Fig. 2. This allows the model to operate on images of varying sizes and captures both local and global features early on, aiding the feature extraction process. Additionally, we employed the C3 structure to address the issue of gradient information repetition during the training of milling workpiece surface images, which accelerated the training speed while maintaining accuracy [14].
3. Neck. To detect the surface roughness of milling workpieces of different sizes, we used the *Feature Pyramid Networks* (FPN) and *Path Aggregation Network* (PAN) structures. These structures can perform convolution on milling workpiece images to form feature maps of different scales. In this structure, the low-resolution high-semantic information feature map is

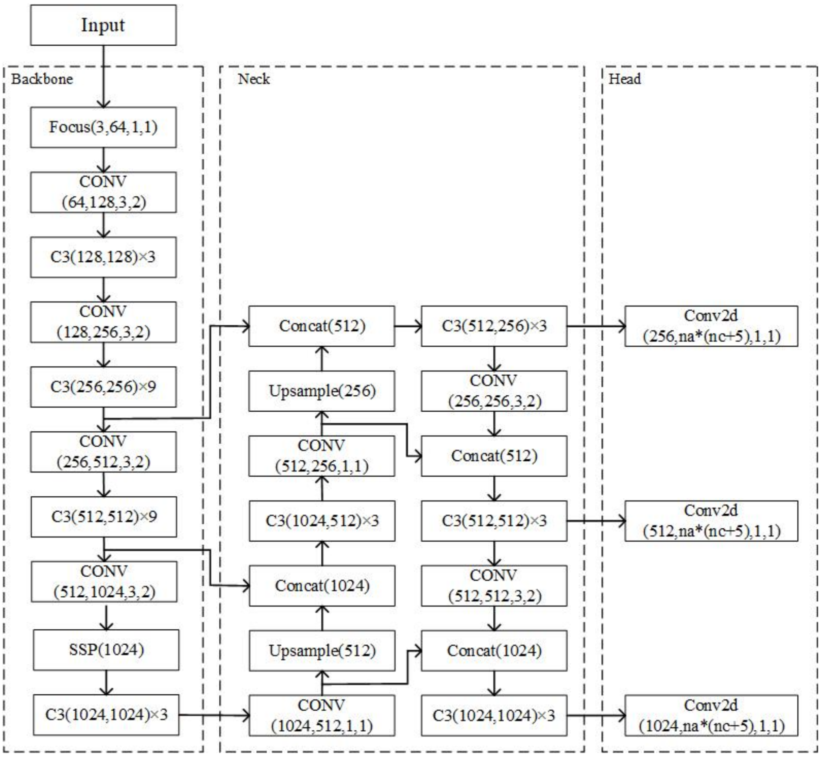


Fig. 1. Yolov5 model structure.

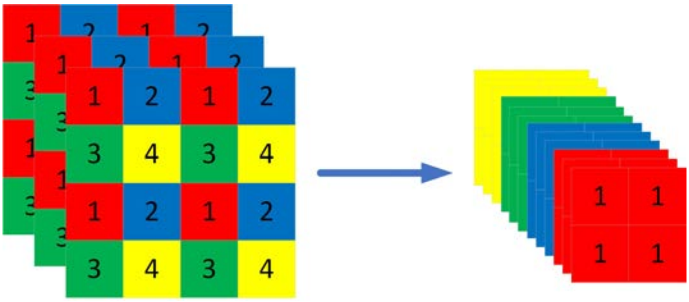


Fig. 2. Focus Principle diagram.

fused with the high-resolution low-semantic information feature map, enabling the model to obtain more comprehensive semantic information [15]. In deep neural networks, the positional information of the feature maps is continuously diminished. With the PAN structure, we can propagate the shallow layer's positional information to the deep layer's feature maps, enhancing the network's ability to recognize milling workpieces of different scales [16].

4. Head. The model used three types of loss functions during the training phase, which were used for the classification, localization, and confidence calculation of the milling workpieces. Binary cross-entropy was used as the loss function for classification and confidence [17], as

Metrol. Meas. Syst., Vol. 30 (2023), No. 3, pp. 531–548
DOI: 10.24425/mms.2023.146425

shown in equations (1) and (2). Due to the milling workpieces being rectangular in the images, and the CIOU_Loss function taking into account the overlap area, center point distance, and aspect ratio between the predicted bounding boxes and ground truth boxes, it can serve as a localization loss function that helps the predicted boxes to better match the ground truth boxes.

$$\begin{aligned}\mathrm{Loss}_{\mathrm{conf}} = &-\sum_{i=0}^{S\times S}\sum_{j=0}^{M} I_{ij}^{\mathrm{obj}}\left[\hat{C}_i^j \log C_i^j + \left(1-\hat{C}_i^j\right)\log\left(1-C_i^j\right)\right] \\ &-\lambda_{\mathrm{noob}}\sum_{i=0}^{S\times S}\sum_{j=0}^{M} I_{ij}^{\mathrm{noobj}}\left[\hat{C}_i^j \log C_i^j + \left(1-\hat{C}_i^j\right)\log\left(1-C_i^j\right)\right],\end{aligned} \tag{1}$$

$$\mathrm{Loss}_{\mathrm{class}} = \sum_{i=0}^{S\times S} I_i^{\mathrm{obj}} \sum_{j=0}^{M}\left[\left(p_i(c)-\hat{p}_i(c)\right)^2\right], \tag{2}$$

$$\mathrm{Loss}_{\mathrm{CIOU}} = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v. \tag{3}$$

In equations (1)–(2), $S \times S$ represents the number of grids, M represents the number of bounding boxes in each grid, I_{ij}^{obj} represents the presence or absence of the milling workpieces in the bounding box, $\hat{C}_i^j$ represents the prediction confidence of the bounding box in the grid, C_i^j represents the true confidence of the bounding box in the grid, $\hat{p}_i(c)$ represents the probability that the milling workpiece is predicted as the roughness class c, while $p_i(c)$ represents the probability that a milling workpiece is roughness class c. In equation (3), bb^{gt} represents the Prediction box and Ground truth box, ρ represents the distance between the centroids of the two boxes, and v represents the similarity of the aspect ratio of the two boxes.

2.2. Improved Yolov5 algorithm

In the field of image processing, neural networks learn features through large amounts of data, while all features do not differ for neural networks and they do not pay too much attention to certain aspects such as the temporal domain, spatial domain, channel, hybrid domain, *etc.* By utilizing attention mechanisms in neural networks, the model can focus more on specific features such as the Squeeze-and-Excitation attention mechanism that emphasizes the relationships between channel features and allows the model to automatically learn the importance of different channel features [18]. The Convolutional Block Attention Module mechanism combines channel and spatial attention in a sequential manner, enabling the model to focus more on the recognition of objects themselves [19].

In the task of milling surface roughness detection, the position of the workpiece is often distributed in local and different regions, while there may exist some noise or irrelevant information in other regions, which can interfere with the learning of the model. Therefore, this paper uses the CA mechanism to better adapt to this situation. The CA mechanism can learn the position information of the milling workpiece in the image by utilizing the position encoding vector of the workpiece, thus achieving more accurate roughness detection [20].

As shown in Fig. 3, CA was added to the bottleneck of the C3 module in the backbone network. Firstly, an average pooling operation was performed along the height and width directions of the milled workpiece image to obtain two feature maps. The output of the c channel at height h and

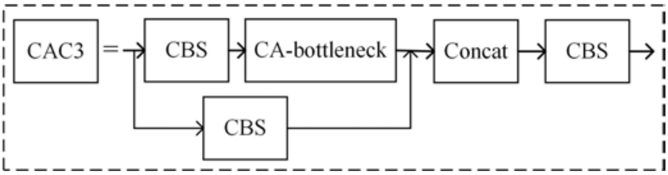


Fig. 3. Location of the added CA.

width w can be expressed as follows:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \le i < W} x_c(h, i), \tag{4}$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \le j < H} x_c(j, w). \tag{5}$$

By cascading the two feature maps above and transforming them using a shared 1 × 1 convolution, the dimension is reduced to the original C/r, the feature map f is generated by the Sigmoid activation function, and the feature maps are convolved 1 × 1 according to the original dimensions to obtain the feature maps f^h and f^w with the same number of channels as the original ones. The Sigmoid activation function is used once more to acquire its attention weights g^h and g^w in the height and width directions. Finally, the feature maps with attention weights in the height and width directions are obtained with weighting calculations. The process is shown in Fig. 4, and the equation for CA output is as follows:

$$f = \delta\left(F_1\left(\left[z^h, z^w\right]\right)\right), \tag{6}$$

$$g^h = \sigma\left(F_h\left(f^h\right)\right), \tag{7}$$

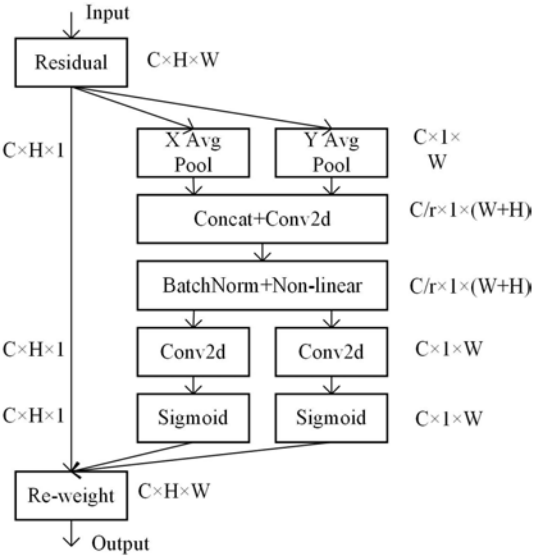


Fig. 4. CA structure diagram.

Metrol. Meas. Syst., Vol. 30 (2023), No. 3, pp. 531–548
DOI: 10.24425/mms.2023.146425

$$g^w = \sigma\left(F_w\left(f^w\right)\right), \tag{8}$$

$$y_c(ij) = x_c(ij) \times g_c^h(i) \times g_c^w(j). \tag{9}$$

3. Experimental design

In industrial applications of roughness detection, the first two challenges that need to be addressed are the robustness to lighting environments and fast detection speed. The robustness to lighting environments is verified by designing variable lighting and image acquisition positions, and the detection speed is verified by comparison tests with the Faster RCNN model and the SSD model, and the experimental flow is shown in Fig. 5, which is divided into five parts. The experiments are based on the PyTorch framework and use the *graphics processing unit* (GPU) to speed up the training. The training environment is shown in Table 1.

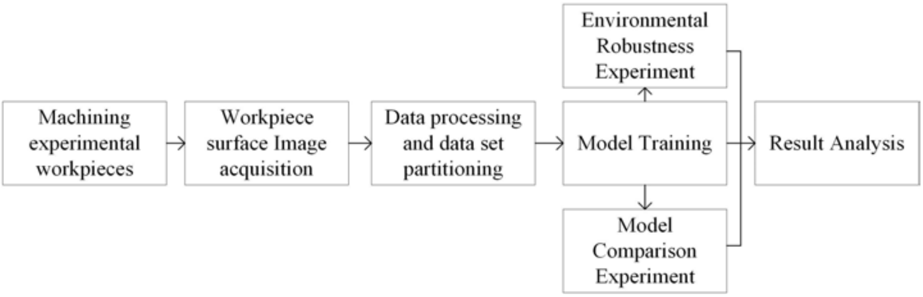

Machining experimental workpieces
Workpiece surface Image acquisition
Data processing and data set partitioning
Environmental Robustness Experiment
Model Training
Model Comparison Experiment
Result Analysis

Fig. 5. Experimental flow diagram.

Table 1. Software and hardware conditions.

Name	Configuration
Operating System	Windows11 (×64)
CUDA	11.3.1
CPU	Inter(R) Core^{TM} i5 –12400 F
Python	3.8
GPU	NVIDA GeForce RTX 3050
RAM	2 × 8 GB

3.1. Preparation of the milling workpieces

Thirty milling workpieces with different surface roughness were processed using the material equipment and processing parameters in Table 2 in the experiment.

The contact-type roughness detector obtains surface roughness information by detecting the variation of the probe caused by sliding over the workpiece surface. As the roughness information of the same workpiece surface at different locations fluctuates to some extent, the roughness of six different areas on each workpiece was measured randomly using a contact-type roughness gauge. The average value was calculated as the surface roughness of the workpiece. In the team's previous work [9], the following measurement results of each workpiece were obtained as shown in Table 3.

Table 2. Experimental materials and selected processing parameters.

Machine tool	**Milling cutter**	**Cutting depth**
XHS7145	TAP400R100-32-6T	0.1 mm
Spindle speed	Feed rate	Workpiece material
600 r/min	200–1100 mm/min	45# steel
Workpiece size	surface roughness tester	Workpiece roughness
60 × 40 × 10 mm	TR210	1–3.6 µm

Table 3. Surface roughness of milled workpiece (unit: µm).

No.	**First**	**Second**	**Third**	**Fourth**	**Fifth**	**Sixth**	**Average**
1	1.310	1.289	1.295	1.290	1.260	1.313	1.293
2	1.063	1.130	1.120	1.122	1.095	1.107	1.106
3	1.077	1.065	1.108	1.028	1.105	1.084	1.078
4	1.378	1.381	1.304	1.306	1.351	1.332	1.342
5	1.214	1.222	1.284	1.213	1.297	1.262	1.249
6	1.221	1.200	1.164	1.154	1.195	1.159	1.182
7	1.483	1.466	1.484	1.447	1.496	1.493	1.478
8	1.440	1.599	1.529	1.483	1.435	1.561	1.508
9	1.595	1.524	1.527	1.558	1.563	1.605	1.562
10	1.433	1.404	1.363	1.471	1.475	1.497	1.441
11	1.593	1.518	1.747	1.629	1.583	1.564	1.606
12	1.489	1.416	1.335	1.448	1.471	1.466	1.438
13	1.931	1.968	1.918	1.948	1.968	2.034	1.961
14	2.085	2.088	1.985	2.054	2.046	1.987	2.041
15	2.054	2.052	2.018	2.060	2.079	1.991	2.042
16	1.899	1.975	2.066	1.852	2.044	1.986	1.970
17	2.195	2.216	2.215	2.158	2.178	2.251	2.202
18	2.118	2.081	2.135	2.156	2.201	2.121	2.135
19	2.809	2.842	2.836	2.849	2.775	2.759	2.812
20	2.845	2.932	2.928	2.991	2.904	2.889	2.915
21	2.880	2.786	2.886	2.738	2.827	2.867	2.831
22	2.754	2.785	2.721	2.852	2.734	2.778	2.771
23	2.795	2.789	2.752	2.787	2.694	2.702	2.753
24	2.702	2.704	2.624	2.672	2.711	2.672	2.681
25	3.363	3.309	3.368	3.379	3.360	3.299	3.346
26	3.173	3.208	3.283	3.276	3.274	3.129	3.224
27	3.353	3.499	3.450	3.391	3.566	3.328	3.431
28	3.377	3.387	3.390	3.362	3.407	3.467	3.398
29	3.319	3.221	3.313	3.360	3.153	3.209	3.263
30	3.417	3.387	3.503	3.484	3.627	3.645	3.511

*Metrol. Meas. Syst.,*Vol. 30 (2023), No. 3, pp. 531–548
DOI: 10.24425/mms.2023.146425

3.2. Acquisition of the workpiece surface image

In order to simulate the variation of illumination in industrial production and evaluate the robustness of the model in different lighting environments, we used the experimental platform shown in Figure 6 to capture images of the milled workpiece surface.

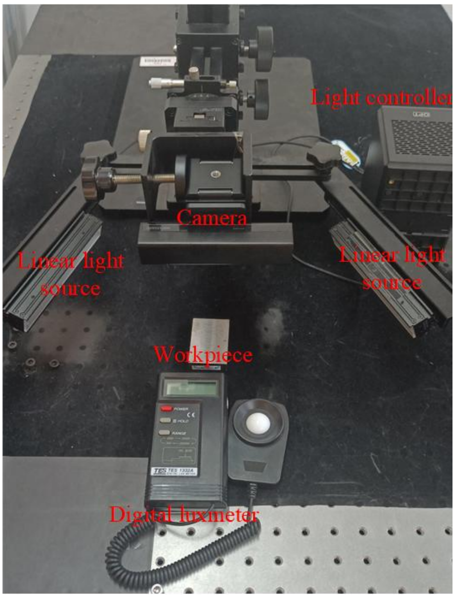


Fig. 6. Workpiece surface image acquisition platform.

To ensure the reproducibility and repeatability of the experiments, the size of the platform used in the experiment was 80 × 60 cm, and the camera used for the image capture was fixed horizontally 20 cm above the platform, with the camera lens vertical to the platform and the workpiece being captured. Two 15 cm-long strip light sources were fixed at both ends of the camera, with an angle of 130 degrees between them. Due to surfaces with different roughness being affected by light in different ways, with lower roughness surfaces having stronger reflection capabilities, we set different levels of lighting intensity in the experimental setup, ranging from 592 to 1060 LUX, and randomly adjusted the position of the workpieces during taking the photographs. The positioning of the light sources and cameras was designed to ensure that the intensity of the light shining on the surface of the workpiece varied at different positions. The images acquired with different illumination and at different positions are shown in Fig. 7.

The experimental equipment in Fig. 6 is a 4K pixel camera, OPTLI14030 white linear light source, light controller, 45# steel milling workpiece, digital luxmeter, and experimental platform.

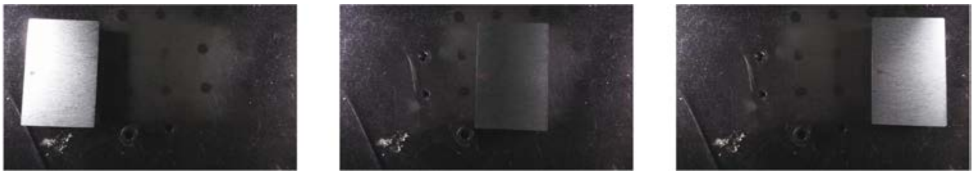


Fig. 7. Some images of the milled workpieces.

3.3. Data pre-processing and data set partitioning

As surface roughness grading of workpieces in actual industrial production follows the standards of the International Organization for Standardization (ISO 1302), which divides roughness grades into ranges of 0–0.4 µm, 0.4–0.8 µm, 0.8–1.6 µm, 1.6–3.2 µm, and 3.2–6.3 µm, the 30 machined workpieces listed in Table 2 were classified into five roughness grade intervals, namely 1.0–1.4 µm, 1.4–1.9 µm, 1.9–2.5 µm, 2.5–3.1 µm, and 3.1–3.7 µm, according to the ISO standard in this experiment. These five groups were denoted as R1, R2, R3, R4, and R5, respectively. For each roughness grade interval, six workpieces were selected, of which four were used as training samples and the remaining two were used as validation samples.

A total of 1090 images of milled workpiece surfaces were captured in the experiment, including 664 images in the training set and 426 images in the validation set. As neural network models typically require many image samples for training, we employed data augmentation techniques such as random flipping, translation, and adjustments of image hue, saturation, and exposure to expand the training dataset. This approach saves costs and improves the model's generalization ability. Therefore, we performed data augmentation on the training set and increased the sample size to 3320 images. To simulate the possible uneven sample distribution, the number of images captured for the R4 grade workpieces was relatively small compared to the other roughness levels. The roughness class classification and the number of data sets are shown in Table 4.

Table 4. Data set partition table (5 categories).

Roughness (unit: µm)	1.0–1.4	1.4–1.9	1.9–2.5	2.5–3.1	3.1–3.7
Number of train set images	688	688	728	488	728
Number of validation set images	92	92	92	58	92

3.4. Model Training

The experiments use the YOLOv5s network with the addition of CA to train the augmented dataset and judge its performance by evaluation metrics. Since the stochastic gradient descent optimizer converges slowly in the training process and may fall into local minima [21], the Adaptive Moment Estimation is used as the optimizer in this experiment.

In the training process, too large a learning rate may make the model unstable, and too small a learning rate may make the model converge slowly and reach a local minimum, so the Warmup learning rate is used to prevent overfitting during training. A first-order linear interpolation algorithm is used in the Warmup phase to update the learning rate for each round [22], and thereafter, a cosine annealing algorithm is used to update the learning rate for each round to ensure the stability of the training process. The larger the image size of the input network, the better the training effect, but it requires more training time and better computer hardware configuration. Combined with the current computer configuration, the input image is adjusted from shooting pixel size 3840 × 2160 to 800 × 800, and the initial learning rate is 0.01, the batch size is 6, and the number of training iterations is 400. some of the training hyperparameters are shown in Table 5.

Table 5. Partial training hyperparameters.

Initial learning rate	**Batch size**	**Epoch**	**Image size**	**Iou_Loss**	**Cls_Loss**
0.001	6	400	800 × 800	0.05	0.5

*Metrol. Meas. Syst.,*Vol. 30 (2023), No. 3, pp. 531–548
DOI: 10.24425/mms.2023.146425

4. Experimental results and analysis

4.1. Evaluation indicators

In target detection algorithms, the goodness of a model is often evaluated by the *average precision* (AP), *mean average precision* (mAP), detection speed, and model size, where AP is related to mAP and accuracy (P), and recall (R). The specific formulas for accuracy and recall are as follows.

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{10}$$

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \tag{11}$$

where TP represents the amount of data in the data set that are actually positive and classified as positive by the classifier, FP represents the amount of data in the data set that are incorrectly predicted as positive, TN represents the amount of data that are correctly predicted as negative, and FN represents the amount of data that are incorrectly predicted as negative. The formula shows that accuracy represents the proportion of positive samples with correct predictions among the predicted samples to all positive samples, and the recall represents the proportion of positive samples with correct predictions among the predicted samples to all samples.

In this paper, accuracy indicates how many of the predicted roughness classes of milled workpieces are correct, and recall indicates how many of all milled workpieces are correctly predicted to have a roughness class. Therefore, accuracy and recall are inversely proportional, and the higher the accuracy, the lower the recall. AP is the area under the Precision-Recall curve for each category, and mAP is the average value of AP for each category. The larger the mAP value, the better the detection effect. The formula is shown below, where N represents the total number of categories, and in this experiment, the roughness of all milling workpieces is divided into 5 categories, so $N = 5$.

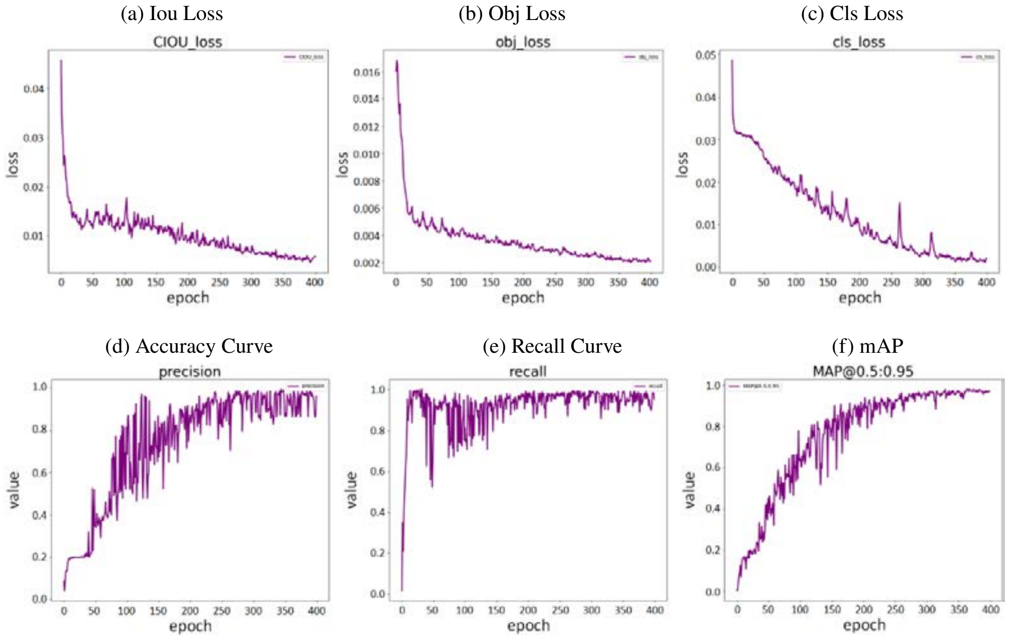
$$\mathrm{AP} = \int_0^1 \mathrm{P(R)dR} \tag{12}$$

$$\mathrm{mAP} = \frac{\sum_{i=1}^{N} \mathrm{AP}_i}{N}. \tag{13}$$

4.2. Experimental results

4.2.1. Analysis of training results

The evaluation results of the model after 400 iterations of training are shown in Fig. 8. From 8c, 8d, 8e and 8f, we can see that the model tends to stabilize after the first 40 rounds of iterative training, although with occasional fluctuations, and starts to stabilize around 200 rounds and converges around 250 rounds. The fluctuations may be caused by two reasons: (1) it is just at the end of the Warmup phase when the learning rate is adjusted by the cosine annealing algorithm to make the curve slowly stabilize. (2) The data set is unbalanced in all types of samples. Although the data set is expanded by data enhancement, the percentage of samples with the 4th level of roughness is still too small compared to other samples due to the limited experimental conditions.


(a) Iou Loss (b) Obj Loss (c) Cls Loss

(d) Accuracy Curve (e) Recall Curve (f) mAP

Fig. 8. Losses on the training set and mAP on the validation set.

After testing the model with the validation set images, the overall accuracy of the model can reach 96.2% and the recall rate can also reach 95.1%. The model has a better detection effect for static milling workpiece images, and some of the static milling workpiece image detection results are shown in Fig. 9. Different roughness workpieces can be correctly identified by the model with a high confidence level.

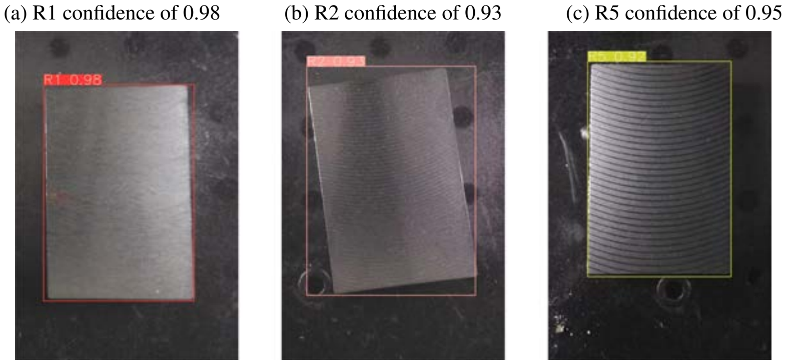

(a) R1 confidence of 0.98 (b) R2 confidence of 0.93 (c) R5 confidence of 0.95

Fig. 9. Static workpiece image detection results.

4.2.2. Model comparison

We created a comparative test with the Faster RCNN model and the SSD model to verify the effectiveness of the method in this study for real-time recognition of surface roughness of milled workpieces.

Metrol. Meas. Syst., Vol. 30 (2023), No. 3, pp. 531–548
DOI: 10.24425/mms.2023.146425

1. Faster RCNN was proposed in 2016 on the basis of RCNN and Fast RCNN as a representative of two-stage target detection models which are still used in various aspects of target detection today. From the paper [9], it is known that the standard ResNet50 backbone network in the COCO (*Common Objects in Context*) target detection dataset can produce higher detection accuracy, so in this paper, the backbone network VGG16 of Faster RCNN is replaced with ResNet50. To make the model converge faster and better, this experiment utilizes the official PyTorch Faster RCNN pre-training weights for migration learning. The learning rate is 0.01, the number of epochs is 100, and the batch size is 6.
2. SSD is a one-stage target detection method, which has the advantages of fast detection speed and high detection accuracy, and its backbone network is based on VGG16 with added multiple convolutional layers and an average pooling layer. As in the experiments above, the backbone network is replaced with ResNet50, and migration learning is performed using pre-trained weights. The learning rate is 0.005, the number of epochs is 100 rounds, and the batch size is 6.

The experiments above were conducted under the same hardware configuration and data set conditions, and the parameters were adjusted by training several times until the model achieved better results. The experimental results are evaluated in terms of mAP, model size, and FPS metrics for comparison The specific experimental evaluation results are shown in Table 6.

Table 6. Performance comparison of different object detection networks.

Networks	**mAP/%**	**Model size/M**	**FPS**
Faster RCNN	97.1	315	7
SSD	92.2	104	34
Yolov5	96.1	13.7	36
Yolov5+CA	97.3	13.2	36

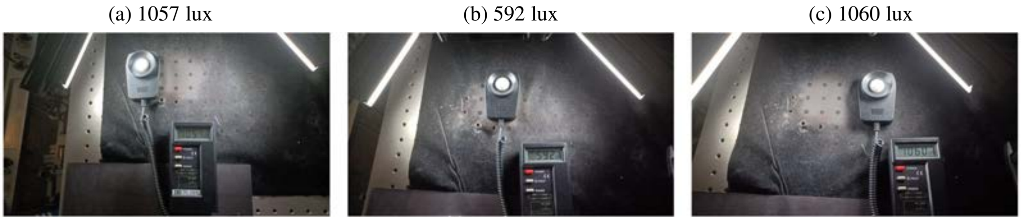
It is obvious from the above table that (1) the Faster RCNN model with migration learning by using pre-trained weights has an mAP close to the improved Yolov5 with added CA, but the detection speed is only 7 frames per second, which is far from the standard of real-time detection. In addition, the model size of 315M makes the Faster RCNN unable to be deployed for use on mobile devices. (2) The SSD model can detect up to 34 frames per second, but its mAP is lower than the rest of the models, and the model size is one-third of Faster RCNN, but still not up to what is needed for mobile deployment. (3) The mAP of the improved Yolov5 model reached 97.3%, which is a 5.2% and 1.2% improvement compared to SSD and Yolov5. The detection speed has improved by 80.6% over Faster RCNN to 36 frames per second, which satisfies the demand for real-time detection. In addition, the model size of 13.2 M is also 95.8% and 87.3% less than that of Faster RCNN and SSD, demonstrating its suitability for mobile deployment.

5. Discussion

5.1. Impact of the lighting environment on the model

As surface roughness measurement based on machine vision relies on optical imaging principles, it involves capturing images of the workpiece surface with a camera, followed by manual or algorithmic extraction of features in the image that are related to surface roughness parameters. The predicted roughness values of the unknown workpiece surfaces are then obtained based on these features. However, different lighting intensities and environments can directly affect the distribution of the features in the workpiece surface images, thus influencing the detection results.

To test the detection speed and robustness of the model under different lighting conditions, we used a high-definition camera with a resolution of 3840 × 2160 and a frame rate of 30FPS to capture videos of milling workpieces with different roughness moving at a constant speed in an environment with a light intensity of 592-1060LUX. During the movement, the light intensity on the surface of the workpiece gradually decreased and then increased again, as shown in Fig. 10. The detection results under different lighting conditions were used to evaluate the model's robustness to changes in light sources.

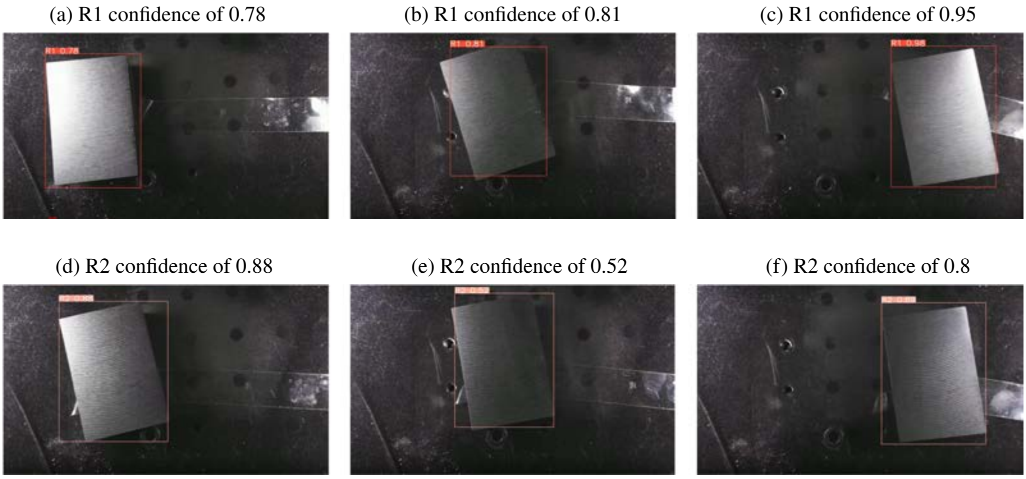

(a) 1057 lux

(b) 592 lux

(c) 1060 lux

Fig. 10. Process of light intensity change.

The results show that when the confidence threshold is set to 0.25, the roughness levels of R1, R2, and R3 milling workpieces can be accurately recognized in the experiment. Some real-time detection processes are shown in Fig. 11. However, during the training process, there were relatively few milling workpieces with roughness level R4, resulting in an imbalanced distribution of the dataset. When the confidence threshold was set low, there was a phenomenon of multiple labels for a single sample when detecting workpieces with higher roughness, as shown in Fig. 12. When the confidence threshold is readjusted to 0.7, the model can accurately recognize workpieces with various levels of roughness. The model can also accurately recognize the roughness levels of workpieces when the surface light intensity changes, indicating that the model has good imaging environment robustness.


(a) R1 confidence of 0.78

(b) R1 confidence of 0.81

(c) R1 confidence of 0.95

(d) R2 confidence of 0.88

(e) R2 confidence of 0.52

(f) R2 confidence of 0.8

Fig. 11. Real-time detection process diagram.

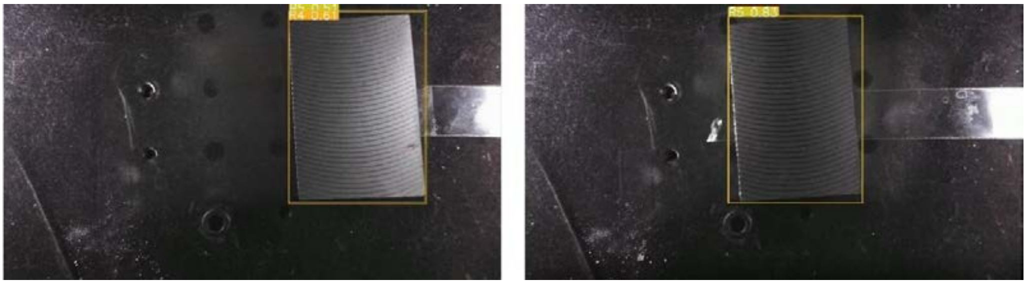


Fig. 12. Multiple labeling phenomenon in real-time detection.

5.2. Impact of model detection speed on industrial production

Real-time detection can help identify problems on the production line in a timely manner and improve production efficiency and product quality. To achieve real-time detection, the detection speed of the model is very important, but it should be emphasized that the required detection speed for different industrial production lines may vary, and there are no standardized specifications. Generally speaking, real-time detection requires a detection speed of at least several dozen frames per second. The detection speed is affected by various factors, such as hardware equipment and model algorithms.

In the hardware environment shown in Table 1, the detection speed of the model is 36FPS when detecting the video data captured in Section 5.1, which is higher than the detection speed of Faster RCNN with higher detection accuracy, indicating its potential for industrial production. As the index-based machine vision method requires a high demand for the light source environment, it is unsuitable for industrial production and thus not compared.

6. Conclusions

The paper proposes a method that can provide visual detection of the surface roughness of milled workpieces, which is based on the Yolov5 model improved by adding the coordinate attention mechanism (CA). In the 5-class roughness level target detection, the following achievements have been made:

1. Average accuracy improved from 96.1% to 97.3%, which is a 5.2% improvement compared to the SSD algorithm.
2. The detection speed is improved by 80.6% compared to Faster RCNN, reaching 36 frames per second.
3. The detection of uniformly moving workpieces in an environment with a light intensity of 592-1060 LUX proves the model's better robustness to light environments.

In future work, we will further optimize the algorithm model to address various challenges existing in real production environments, aiming to enhance its generalization and environmental robustness. We will focus on detection of finishing workpieces with weak surface feature information, such as the detection of grinding workpieces, so that the algorithm can have higher detection accuracy and higher detection speed.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC) (Grant No. 52065016).

Appendix I. Nomenclature

Abbreviations	**Definition**
YOLO	You Only Look Once Network
CA	Coordinate Attention
SVM	support vector machine
CIOU	Complete-IoU
SE	Squeeze-and-Excitation
CBAM	Convolutional Block Attention Module
FPN	Feature Pyramid Networks
PAN	Path Aggregation Network
Ra	Roughness parameters

Symbol	**Definition**
S	number of grids
M	number of bounding boxes in each grid
I_{ij}^{obj}	presence or absence of objects in the bounding box
$\hat{C}_i^j$	prediction confidence of the bounding box in the grid
C_i^j	true confidence of the bounding box in the grid
$\hat{p}_i(c)$	probability of predicting the detected object as class c
$p_i(c)$	enhancement node group
c	Objective Category
b	Prediction box
b^{gt}	Ground truth box
ρ	Euclidean Distance between the centroids of the two boxes
v	the similarity of the aspect ratio of the two boxes
f	feature map
AP	average precision
mAP	mean average precision
TP	Ture Positives
FP	False Positives
TN	Ture Negatives
FN	False Negatives

Parameter Name	**Setting**
Initial learning rate	0.001
Image size	800 × 800
Iou_Loss	0.05
Cls_Loss	0.5
Batch size	6
Epoch	400
Optimizers	Adam

*Metrol. Meas. Syst.,*Vol. 30 (2023), No. 3, pp. 531–548
DOI: 10.24425/mms.2023.146425

References

[1] Kiran, M. B., Ramamoorthy, B., & Radhakrishnan, V. (1998). Evaluation of surface roughness by vision system. *International Journal of Machine Tools and Manufacture*, *38*(5–6), 685–690. https://doi.org/10.1016/S0890-6955(97)00118-1

[2] Gadelmawla, E. S. (2004). A vision system for surface roughness characterization using the gray level co-occurrence matrix. *NDT & e International*, *37*(7), 577–588. https://doi.org/10.1016/j.ndteint.2004.03.004

[3] Huaian, Y. I., Jian, L. I. U., Enhui, L. U., & Peng, A. O. (2016). Measuring grinding surface roughness based on the sharpness evaluation of colour images. *Measurement Science and Technology*, *27*(2), 025404. https://doi.org/10.1088/0957-0233/27/2/025404

[4] Zhang, H., Liu, J., Lu, E., Suo, X., & Chen, N. (2019). A novel surface roughness measurement method based on the red and green aliasing effect. *Tribology International*, *131*, 579–590. https://doi.org/10.1016/j.triboint.2018.11.013

[5] Somthong, T., & Yang, Q. (2016, May). Surface roughness measurement using photometric stereo method with coordinate measuring machine. In *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings* (pp. 1–6). IEEE. https://doi.org/10.1109/I2MTC.2016.7520329

[6] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

[7] Rifai, A. P., Aoyama, H., Tho, N. H., Dawal, S. Z. M., & Masruroh, N. A. (2020). Evaluation of turned and milled surfaces roughness using convolutional neural network. *Measurement*, *161*, 107860. https://doi.org/10.1016/j.measurement.2020.107860

[8] He, Y., Zhang, W., Li, Y. F., Wang, Y. L., Wang, Y., & Wang, S. L. (2021). An approach for surface roughness measurement of helical gears based on image segmentation of region of interest. *Measurement*, 183, 109905. https://doi.org/10.1016/j.measurement.2021.109905

[9] Su, J., Yi, H., Ling, L., Wang, S., Jiao, Y., & Niu, Y. (2022). A surface roughness grade recognition model for milled workpieces based on deep transfer learning. *Measurement Science and Technology*, *33*(4), 045014. https://doi.org/10.1088/1361-6501/ac3f86

[10] Li, W., Li, F., Luo, Y., & Wang, P. (2020, December). Deep domain adaptive object detection: a survey. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1808–1813). IEEE. https://doi.org/10.1109/SSCI47803.2020.9308604

[11] Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE*. https://doi.org/10.48550/arXiv.1905.05055

[12] Wu, X., Sahoo, D., & Hoi, S. C. (2020). Recent advances in deep learning for object detection. *Neurocomputing*, *396*, 39–64. https://doi.org/10.1016/j.neucom.2020.01.085

[13] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 779–788). https://doi.org/10.1109/CVPR.2016.91

[14] Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., & Yeh, I. H. (2020). CSPNet: A new backbone that can enhance learning capability of CNN. In *Proceedings of The IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 390–391).

[15] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2117–2125).

[16] Liu, S., Qi, L., Qin, H., Shi, J., & Jia, J. (2018). Path aggregation network for instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 8759–8768).

[17] Zhao, Z., Yang, X., Zhou, Y., Sun, Q., Ge, Z., & Liu, D. (2021). Real-time detection of particle board surface defects based on improved YOLOV5 target detection. *Scientific Reports*, *11*(1), 21777. https://doi.org/10.1038/s41598-021-01084-x

[18] Jie, H., Li, S., Gang, S., & Albanie, S. (2017). Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(8), 2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372

[19] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 3–19). https://doi.org/10.1007/978-3-030-01234-2_1

[20] Hou, Q., Zhou, D., & Feng, J. (2021). Coordinate attention for efficient mobile network design. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13713–13722). https://doi.org/https://doi.org/10.48550/arXiv.2103.02907

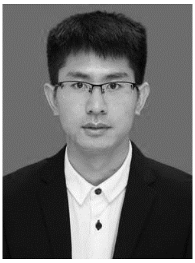
[21] Ruder S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

[22] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

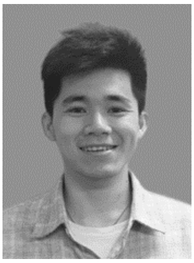
Xiao Lv is a postgraduate student at the College of Mechanical and Control Engineering at Guilin University of Technology. His main research area is machine vision.

Aihua Shu is a teacher at the College of Foreign Languages of Guilin University of Technology.

Huaian Yi is an associate professor at the College of Mechanical and Control Engineering at Guilin University of Technology and his main research area is machine vision.



Enhui Lu received his Ph.D. in Mechanical Engineering from Hunan University. Currently, he is a teacher at the College of Mechanical Engineering at Yangzhou University and his main research interests are machine vision and artificial intelligence algorithms for surface quality assessment.



Runji Fang received his B.Sc. degree from Guilin University of Technology in 2020. At present, he is a current graduate student of Guilin University of Technology. His main research interest is machine vision.