ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20250214002373
Original scientific paper

Steel Metallurgy Crack Detection and Material Pile Volume Measurement Based on 3D Point Cloud and Improved SSD Model

Wei DENG*, Bingquan ZHU

Abstract: For crack detection and stockpile volume measurement in steel metallurgy production process, an improved single shot multibox detector model and Delaunay triangulation technique based on greedy projection triangulation are proposed, respectively. Among them, the study mainly adopts ShuffleNet V2 module and efficient channel attention network to optimize the detection efficiency and accuracy of the original detection model. Meanwhile, based on the 3D point cloud dissection results, the triangular projection with depth is used to realize the pile volume measurement. The results showed that compared with other advanced models, the improved single shot multibox detector model achieved a mean average precision of 98.68% in steel plate crack detection, which was 26.7% higher than the faster region based convolutional neural network. The area under the curve of the receiver operating characteristic curve of the improved model was as high as 0.8576, and the detection time was only 29.85 ms. Meanwhile, the mean relative error of the proposed stack metrology method in the metrology experiments was only 2.15%, while the mean relative error values of the Digital Surface Modeling Method and volume balance method were as high as 3.94% and 6.71%, respectively. This indicates that the proved crack detection and stockpile metrology methods for steel plates have significant performance advantages and provide effective technical support for quality control in the production process of iron and steel metallurgy.

Keywords: 3D point cloud; cracks; iron and steel metallurgy; material pile; SSD; volume

1 INTRODUCTION

Iron and steel metallurgy industry is an important pillar of the global industry, and in its production process, crack detection (CD) on the surface of steel plate is an important link to ensure product quality. Automated and high-precision CD technology is crucial in the quality control of iron and steel metallurgy industry [1-3]. At the same time, the production process of steel plate is very complex and has high requirements on the variety and quantity of raw materials. However, due to the limitation of the storage site, the production pile needs to be volumetrically measured and monitored during the production process. Therefore, efficient volumetric metering of production stockpiles is essential [4]. Currently, the detection methods of steel plate cracks are mainly of two categories: physical detection and computer vision detection. The former mainly utilizes ultrasonic waves or lasers to carry out physical detection, which requires the setting of complex sensors, and the cost is high, and the applicable scenes are limited [5]. Computer vision detection methods are mainly used for feature extraction by processing images, which are more widely used at present. However, the traditional image processing technology is difficult to adapt to the complex and changing production environment, the detection accuracy (DA) is low, to be further improved [6]. At present, most of the measurement methods for the production pile are stacked into a regular three-dimensional shape for manual calculation, which is time-consuming, expensive. Moreover, due to the different shapes and sizes of different raw materials, the accuracy of measurement is low. Therefore, more advanced metering methods need to be proposed.

Nowadays, CD technology is widely applied in many different industries. Meng S. et al. presented an automated real-time CD approach based on unmanned aircraft in an attempt to address the low accuracy and efficiency of standard CD methods in complicated situations. The method fully integrated lightweight classification, rough and fine segmentation, and crack width measurement algorithms. The results demonstrated that the method effectively improved the crack edge DA under non-ideal shooting conditions [7]. To achieve CD in asphalt pavements, Nguyen S. D. et al. evaluated the accuracy of various deep learning (DL) based crack recognition. The results showed that faster region-based convolutional neural network (Faster-R-CNN) worked best in crack object detection, and DenseNet and ResNet performed well in crack classification [8]. Xiang X. et al. proposed a DL based pavement CD network to address the complexity and uneconomicality of traditional pavement CD methods. The network incorporated the latest Transformer module to enhance the effectiveness of CD. Meanwhile, the long-range correlation of cracks was recorded using the Transformer. The results indicated that the method achieved CD accuracy F1 values of 0.6739 and 0.6650 on two pavement datasets [9]. Ye W. et al. created a pixel-level segmentation network that combined multi-scale data and suggested a methodical pixel-level crack segmentation-quantization technique for slab track CD. Meanwhile, the errors of crack widths predicted by the method were only 0.13 and 0.12 mm, respectively [10].

Stereo vision technology has a large advantage in the application of volume and length measurement. To achieve diameter detection of sorghum stalks in the field, Xiang L. et al. proposed a high-throughput plant phenotyping method based on portable stereo imaging. The module used a stand-alone embedded design and automatically extracted stems and quantified stem diameters by masked region convolutional neural network (CNN) and semi-global block matching algorithms. The results showed that the correlation coefficient between the stem diameter generated by this method and the actual value was as high as 0.97, with an average absolute error of only 1.44 mm [11]. Vargas R. et al. addressed the issue of low precision of measuring 3D shapes in streak projection contouring by proposing a measurement technique that combines phase coordinate mapping with inverse projection stereo vision. According to the results, the hybrid approach achieved high-precision 3D shape measurement and outperformed

the conventional stereo vision method in terms of reconstruction time and accuracy [12]. To effectively measure the length of live fish in aquaculture, Silva C. et al. proposed an underwater visual measurement system. The system mainly realized the length measurement of live fish by estimating the angle of incidence of light in the water, and at the same time, a short baseline design was used to improve the 3D reconstruction accuracy. The results indicated that the measurement error of this system for live fish length was less than 1% and had high robustness [13]. For the measurement of athletes' movement displacement parameters, Li H. et al. proposed a non-contact displacement measurement system. The system incorporated wireless sensing-based technology and binocular stereo vision technology. The system mainly utilized template matching to realize the extraction of the coordinates of the athlete's image and used Euclidean 3D reconstruction to recover the spatial coordinates [14].

To summarize, numerous efficient methods for measuring CD and volume length have been put forth by researchers both domestically and internationally. However, the current CD methods still have problems such as low DA and complicated calculation. Meanwhile, there is still much room for progress in volume measurement methods. In recent years, with the rapid development of 3D point cloud (3D-PD) technology and deep learning methods, detection methods based on 3D-PCs have gradually become an effective means to solve these problems. In the area of CD in steel metallurgy, three-dimensional point cloud data can accurately reflect the shape and location of cracks while providing more detailed geometric information for volume measurement. Furthermore, the single shot multibox detector (SSD) model, a classic object detection algorithm, has achieved notable results in numerous domains. However, there is a paucity of research on its application to CD. Therefore, the SSD model is introduced in the detection of steel plate cracks. To create the ShuffleNet V2-ECA-SSD model, the ShuffleNet V2 module with efficient channel attention (ECA) module is presented. Meanwhile, binocular stereo vision technology is introduced, and Delaunay triangulation technique based on greedy projection triangulation (GPT) algorithm is adopted to realize 3D-PC dissection of the pile. The volume of the pile is measured according to the triangular projection and depth, in an attempt to realize efficient quality control in iron and steel metallurgy production. The innovation of the study is to replace the backbone network (BN) of the SSD model with the ShuffleNet V2 module, which realizes the light weight of the model. Moreover, the ECA module is used to increase the model's DA even more. In addition, a non-contact method of volumetric measurement of stockpile is realized by using Delaunay triangular profiling technique based on GPT algorithm.

2 METHODS AND MATERIALS

The study first proposes an SSD model for steel plate CD in steel and metallurgy production plants. To increase the detection model's accuracy and efficiency, the ShuffleNet V2-ECA-SSD model is created by introducing the ShuffleNet V2 and ECA modules in turn. Then, to realize the volume measurement of the pile in the steel metallurgy production, binocular stereo vision technology is introduced. Moreover, Delaunay triangular sectioning technology based on GPT algorithm is used to realize 3D-PC sectioning of the pile, and the volume measurement of the pile is carried out according to the triangular projection and depth.

2.1 CD in Steel Metallurgy Based on Improved SSD Modeling

For CD of steel plates in steel metallurgy, the study proposes a one-stage CNN model known as the SSD model. The input of this model contains three color channels, and the input image will be adjusted to a uniform size to facilitate multi-scale target detection. Fig. 1 depicts the SSD model's network structure.

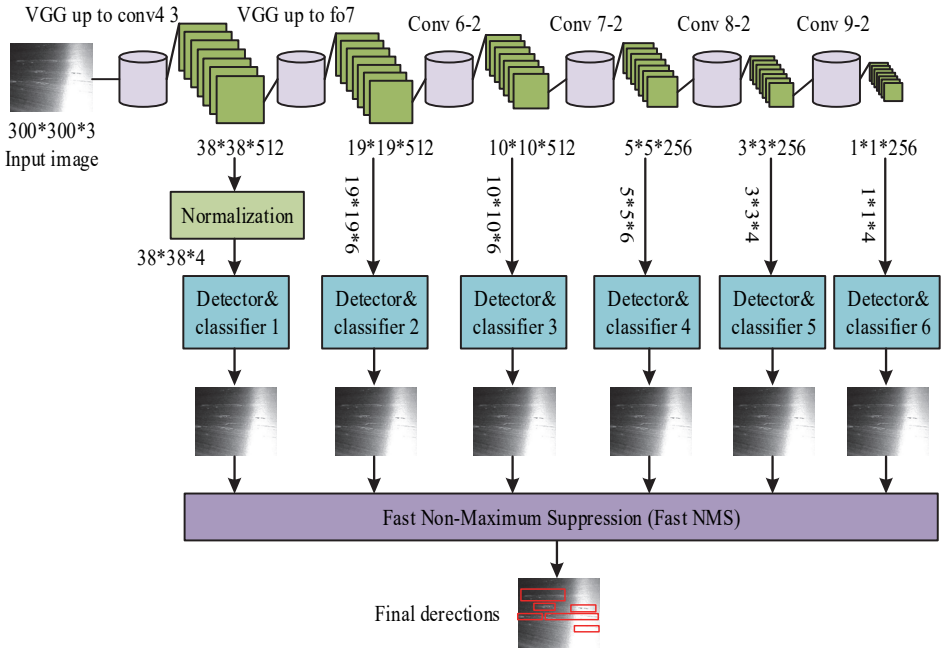


Figure 1 The network structure of SSD model

In Fig. 1, the BN of the SSD model is VGG16 for feature extraction. The effective feature layers include Conv4_3, Conv6_2, Conv8_2, Conv9_2, and Fc7 layers. Num_anchors represent the number of prior frames associated with each point of the feature layer. Each layer of feature extraction is convolved with 4 times Num_anchors for the prediction of geometric transformation parameters. The size of the a priori frames has a linear relationship with the feature map (FM), as shown in Eq. (1).

$$S_e = S_{\min} + \frac{S_{\max} - S_{\min}}{m-1}(e-1),\ e \in [1, m] \tag{1}$$

In Eq. (1), a_r represents the e +1-th FM. a_r is the total quantity of FMs. S_e is the scale of the a priori frame on the e +1-th FM. $S_{\max}$ and $S_{\min}$ denote the upper and lower limits used for scale scaling. The width of the a priori box is calculated as shown in Eq. (2).

$$w_e^a = S_e\sqrt{a_r} \tag{2}$$

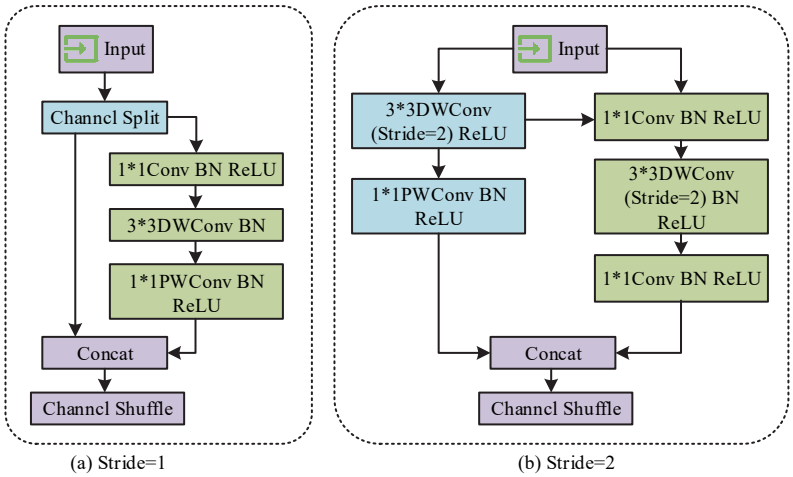
In Eq. (2), generation w_e^a represents the a priori frame width. a_r denotes the set of aspect ratios of the a priori frame, taking the values shown in Eq. (3).

$$a_r = \{1, 2, 3, 1/2, 1/3\} \tag{3}$$

The length of the a priori frame is calculated as shown in Eq. (4).

$$h_e^a = \frac{S_e}{\sqrt{a_r}} \tag{4}$$

In Eq. (4), h_e^a represents the a priori frame length. Although the traditional SSD algorithm has high DA, it has high computational complexity and is difficult to be deployed in resource-constrained environments [15, 16]. Considering the lightweight requirement of the model, the study optimizes its structure and adopts the ShuffleNet V2 module to replace the BN in the original SSD model for feature extraction. This network has superior feature extraction capability and extraction speed compared to VGG. ShuffleNet V2 adopts the Channel Shuffle mechanism and contains three grouped convolutions. It can realize the information fusion between different groups and has a significant advantage in DA. The channel shuffle mechanism is mainly responsible for reorganizing the feature matrix after segmentation. Furthermore, ShuffleUnit is the core module of the mechanism, containing two processing paths. In Fig. 2, the particular structure is displayed.



(a) Stride=1 (b) Stride=2

Figure 2 Shuffle unit structure diagram

In Fig. 2, "BN" denotes the batch normalization process and "DWConv" denotes the depth divisible convolutional layer. The step size of Fig. 2a is 1. In this case, the ShuffleUnit structure contains two paths, and the first path will not process any data. The second path contains two 1 × 1 Conv layers and a 3 × 3 DWConv layer. The last two paths are fused with each other to generate the final result. Fig. 2b has a step size of 2. where the input feature matrix is not channel-split. The first path contains a 3 × 3 DWConv layer with a 1 × 1 one Conv layer. The second path is consistent with Fig. 2a. The number of channels in the output is doubled compared to the input. With this network, the lightweighting of the model can be effectively realized. In the actual steel metallurgy steel plate CD, there are a large number of small cracks. Detecting these small targets with the enhanced SSD model utilizing ShuffleNet V2 is challenging [17]. To address this problem, the study further introduces ECA, which employs a one-dimensional convolution for the extraction of channel-to-channel dependencies and obtains localized channel interaction information without dimensionality reduction. Fig. 3 displays the ECA module's flowchart.

Fig. 3 reveals that the input image has C channels. Processing using global average pooling (GAP) is done first. The spatial information of each channel is compressed into a single value to obtain a 1 × 1 × C feature vector. The specific calculation is shown in Eq. (5).

$$z_c = \frac{1}{H \times W}\sum_{h=1}^{H}\sum_{w=1}^{W} x_{h,w,c}, \quad \forall c = 1, 2, ..., C \tag{5}$$

In Eq. (5), z denotes the global average information of each channel. $x_{h,w,c}$ denotes the pixel value of the h-th row,

w-th column, and *c*-th channel. *H* and *W* denote the height and width of the FM. The convolution kernel (CK) size is then adaptively chosen based on the input FM's channel count, and the result is displayed in Eq. (6).

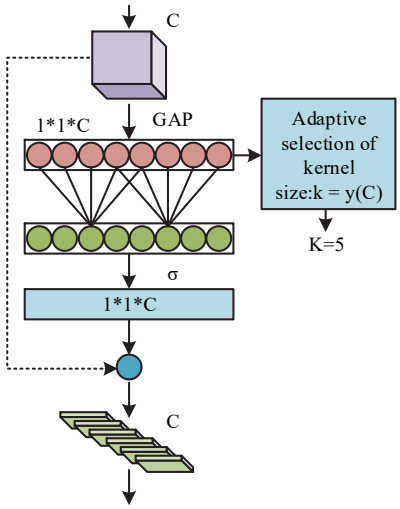


Figure 3 Flow chart of ECA module

$$k = \left[\delta \cdot C^{\alpha} \right] + 1 \tag{6}$$

In Eq. (6), k denotes the CK size. δ and α denote the hyperparameters, which take the values of 0.5 and 2, respectively. Then the feature vector after GAP is passed through the one-dimensional convolutional layer. Moreover, the weights of each channel are determined by applying the Sigmoid activation function to the one-dimensional convolution's output. Eq. (7) displays the determination of the specific weight.

$$\omega = \sigma(\mathrm{Conv1d}_s(y)) \tag{7}$$

In Eq. (7), ω represents the weights. σ represents the Sigmoid activation function. $\mathrm{Conv1d}_s$ denotes one-dimensional convolution operation and s is the CK size. y denotes the feature vector obtained after GAP. Next, the original input FM is multiplied with the weights computed through the Sigmoid function to weight each channel, and the expression is shown in Eq. (8).

$$X' = X \otimes \omega \tag{8}$$

In Eq. (8), X' denotes the weighted FM. $\otimes$ denotes the element-by-element multiplication on the channel dimension, and the final output FM. Eq. (9) provides the formula for updating the FM following the combination of the ECA module.

$$\hat{X} = X + \gamma \cdot X' \tag{9}$$

In Eq. (9), γ denotes the scaling factor, which is employed to adjust the degree of integration between the output FM and the original FM. By introducing the ShuffleNet V2 and the ECA module into the original SSD model, the ShuffleNet V2-ECA-SSD model is thus generated. This model not only reduces the computational overhead and improves the inference efficiency compared with the SSD model in steel plate CD in steel metallurgy, but also improves the DA in small target CD.

2.2 Volumetric Measurement of Stockpile Based on 3D-PC

In the domain of steel metallurgy, the processes of CD and pile volume measurement are of paramount importance. The former is employed to monitor the quality of steel plates, while the latter ensures the rational utilization and accurate calculation of raw materials [18]. In the aforementioned study, the ShuffleNet V2-ECA-SSD model is utilized to detect steel plate cracks. Subsequent research is conducted to combine three-dimensional point clouds for material pile volume measurement. This is done to achieve the dual goals of quality control and resource optimization in the steel metallurgy production process. To realize the volume measurement of the stockpile, the study mainly uses binocular cameras to acquire the material images, obtain the corresponding 3D-PC information, and then measure the volume of the stockpile. This process involves binocular stereo vision technology, by utilizing two cameras to simultaneously capture the image of material pile from different viewpoints, to obtain the parallax value and parallax depth of each 2D point in the material image. Then the study combines the principle of triangulation to further derive the 3D-PC dataset of the stockpile from the known 2D point information obtained from multiple viewpoints. It is assumed that the projected coordinates of a point in two viewpoints are (x_1, y_1) and (x_2, y_2) (x_2, y_2), respectively. The depth information Z of the point is expressed in Eq. (10).

$$Z = \frac{B \cdot f}{d} \tag{10}$$

In Eq. (10), B denotes the baseline length between the two cameras. f denotes the camera focal length. d is the parallax. After obtaining Z, then the 2D projected points can be transformed into 3D point coordinates in space. The specific transformation calculation is shown in Eq. (11).

$$\begin{cases} X = \dfrac{(x_1 - c_x) \cdot Z}{f} \\ Y = \dfrac{(y_1 - c_y) \cdot Z}{f} \\ Z = \dfrac{B \cdot f}{d} \end{cases} \tag{11}$$

In Eq. (11), (X, Y, Z) denotes the spatial point coordinates. Through this relation, each 2D projected point can be converted to its coordinates in 3D space to obtain the complete 3D-PC dataset of the stockpile. However, this approach can only obtain the decentralized 3D-PC [19]. To realize the 3D reconstruction of the stockpile, the study introduces the Delaunay triangular dissection algorithm based on GPT to dissect the 3D-PC. This algorithm is an optimized triangular dissection method that improves the accuracy and computational efficiency of point set partitioning [20]. In this algorithm, the attributes of the

points are classified into two categories, dead points and live points. Live points are located in the boundary or internal partitions of Delaunay triangular dissection and are points that participate in the construction of triangles during the dissection process. By updating the live points, the algorithm is able to progressively optimize the dissection structure and maximize the dissection result at each step in a "greedy" way, making sure that no point falls inside the outer circle of any triangle. Dead points are points that have been completely processed or no longer affect the current triangulation. They are no longer active at a given stage of the dissection calculation and their properties no longer change. Dead points are mainly responsible for providing the algorithm with a static regional division of the point set to avoid redundant calculations. The attributes of edges are categorized as dead edges, live edges, and boundary edges. Dead edges can be used as optimal edges and live edges are edges that need to be involved in forming other triangles. Boundary edges are the edges located at the boundary of the region, which connect the live and dead points. The exact flow of the algorithm is shown in Fig. 4.

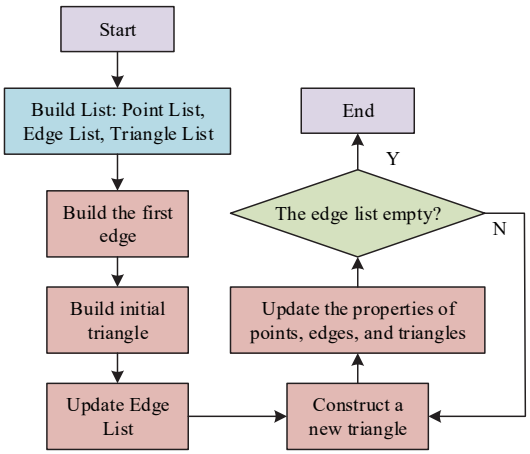


Figure 4 3D-PC partitioning process based on GPDT algorithm

In Fig. 4, when the machine performs triangular sectioning, the list of points, edges, and triangles first needs to be constructed. Then initial triangle edges are generated and saved to the edge list. Subsequently, a live point is taken from the live edge list to form the optimal initial triangle. At this point, the edge list is updated. A live edge is taken from the list to form a new triangle mesh. Subsequently, update the attributes of each point and edge as well as the list. It ends when all the live edges have been updated as dead edges. In addition, the 3D-PC dissection operation based on the GPT algorithm is mainly performed in the point cloud library, where the file format is in PCD format. Therefore, the 3D-PC format obtained from the study needs to be converted from TXT to PCD [21]. Fig. 5 displays the schematic diagram of the stockpile's 3D profiling mesh that is produced using 3D-PC.

A regular triangular section is illustrated in Fig. 5. Each triangle shares edges with other triangles to form a continuous mesh structure. Once the GPT profiling is performed on the 3D-PC coordinates, the calculation of the pile volume is then performed based on the triangular projection with depth. It is assumed that A, B, and C denote the vertices of the triangular cylinders, and the corresponding depth values are denoted as h_a, h_b, and h_c, respectively. The volume of the triangular column is calculated as shown in Eq. (12).

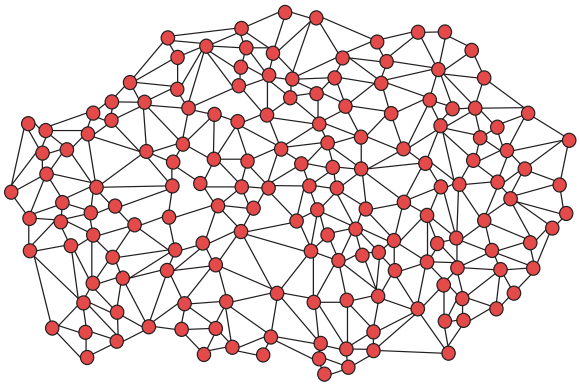


Figure 5 Schematic diagram of three-dimensional mesh division of material pile

$$V = \frac{S \cdot (h_i + h_j + h_k)}{3} \tag{12}$$

In Eq. (12), S represents the triangular area of the triangular column projection. The volume of the whole material pile is calculated as shown in Eq. (13).

$$V_{\text{total}} = \sum V \tag{13}$$

The overall volume of the pile can be calculated by transforming the captured 3D-PC into individual volume calculation units. In this case, the schematic of the column generated after the individual triangular projections is shown in Fig. 6.

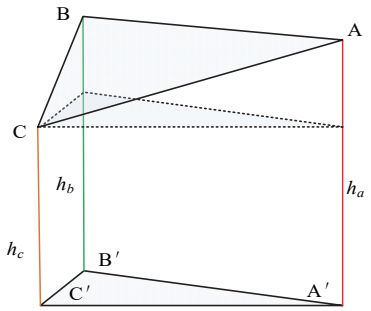


Figure 6 Schematic diagram of the cylinder generated after triangular projection

In Fig. 6, the triangle consisting of vertices A, B, and C is the triangle after the GPT algorithm is dissected. The triangle formed by vertices A', B', and C' is then the projection of the triangle. The volume of the whole material pile is mainly calculated by summing up all the triangular cylinder volumes. In addition, the influence of noise and outliers must also be considered when measuring the volume of material piles. In this study, the radius filtering method is applied to reduce the influence of noise. Specifically, the basic idea of radius filtering is to traverse each point in the point cloud and check whether there are enough neighboring points within a spherical area of a certain radius. If the number of neighboring points around a point is less than the predetermined threshold, the point is considered a noise point and is removed. The process of anomaly detection includes extraction of difference points, clustering of difference points, and screening of effective

anomaly areas. The key to difference extraction lies in calculating the Euclidean distance between points. If the Euclidean distance between a point in the point cloud to be detected and any point in the template point cloud exceeds a preset threshold, the point is identified as a difference point. On the contrary, it is not considered as a point of difference. Next, the density-based spatial clustering algorithm (DBSCAN) is used for density clustering of noisy applications. Through this process, adjacent differential points are clustered into point cloud clusters to identify potential anomalous areas. Finally, based on the geometric characteristics of the anomalous regions, effective anomalous regions are further screened out.

3 RESULTS

This study first verifies the DA and efficiency of the ShuffleNet V2-ECA-SSD model in steel plate CD, using two more advanced detection algorithms to compare with it. Ablation tests are carried out concurrently to confirm each module's efficacy. Then, the performance of 3D-PC profiling technique based on GPT algorithm in stockpile volume metering is tested to verify the volume metering accuracy of the method.

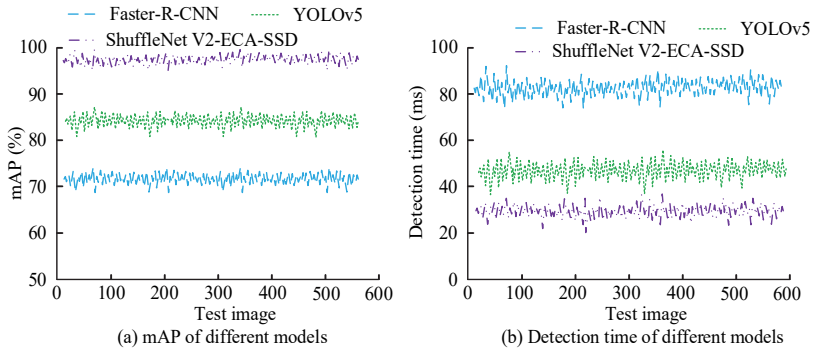
3.1 Analysis of Steel Plate CD Based on ShuffleNet V2-ECA-SSD Modeling

In the steel plate CD experiment, the dataset used in the study is obtained from steel plate images taken by a certain steel metallurgical plant. These images are taken in actual production environments to capture cracks and other possible defects on the surface of the steel plate. In the production process of iron metallurgical plants, processes such as steel melting, rolling, and cooling all involve high temperatures. The production temperature of steel is typically over 1500 °C. To prevent equipment overheating, industrial cameras and light sources that capture images typically require special high-temperature protective housings to avoid direct exposure to high-temperature areas. At the same time, to reduce the effect of vibration on the image quality, the imaging equipment is mounted on an anti-vibration mount. At the beginning of the experiment, the researchers obtain 2926 images of steel plate surfaces at different production stages, capturing a variety of crack types, including cracks, scratches, etc. In the process of organizing the dataset, the researchers perform image filtering, removing images with too dark or too bright lighting. To improve the robustness of the model and reduce the risk of overfitting, data enhancement is performed on the training set. Augmentation operations include random cropping, rotation, mirror flipping, brightness adjustment, noise addition, etc. In the data pre-processing stage, all images are normalized. After a series of processing, a total of 22200 images are obtained. Twenty percent of them are utilized for testing the model, and eighty percent are used for training it. Tab. 1 displays the configurations of the experimental setup.

Table 1 Experimental environment setup

Hardware and software	Setting
Deep learning framework	Pytorch 1.8
CPU	AMD Ryzen 7 5800H
GPU	GeForce RTX 3060
Memory	8GB
Operating system	Windows 7 64bits
Development language	Python 3.6

The more sophisticated you only look once v5 (YOLOv5) model and the Faster-R-CNN model are used in the study to examine the performance of the suggested ShuffleNet V2-ECA-SSD model in the identification of steel plate cracks. Mean average precision (mAP) and detection time (DT) are two evaluation indices that are utilized for comparison. Fig. 7 displays the outcomes of comparing the mAP value and DT for various models. In Fig. 7a, the ShuffleNet V2-ECA-SSD model has a high mAP value of 98.68% in steel plate CD. While the Faster-R-CNN model is only 71.98%, which is 26.7% lower compared to the ShuffleNet V2-ECA-SSD model. From Fig. 7b, the DT of ShuffleNet V2-ECA-SSD model in steel plate CD is only 29.85 ms, which is reduced by 16.29 ms and 53.23 ms compared to YOLOv5 and Faster-R-CNN model, respectively. It displays the high DA and detection efficiency of the suggested steel plate CD model. Among them, the ECA module improves the DA of the model, and the ShuffleNet V2 module realizes the lightweight of the model.



(a) mAP of different models (b) Detection time of different models

Figure 7 The mAP values and detection time of different models

The next step in the study is to draw each model's receiver operating characteristic curve (ROC) and compare each model's area under the curve (AUC). The comparative results are displayed in Fig. 9. The AUC value of the ShuffleNet V2-ECA-SSD model is as high as 0.8576. It is followed by the YOLOv5 model, which has an AUC value of 0.7536. On the other hand, the AUC value of the Faster-R-CNN model is only 0.6854, which is significantly lower than that of the proposed ShuffleNet V2-ECA-SSD model. The reason is that the YOLOv5 model is difficult to capture

fine differences when dealing with certain complex scenes or small cracks. Faster-R-CNN has high computational complexity in the model, which affects the accuracy and efficiency. In contrast, the detection model proposed in the study has stronger generalization ability and DA.

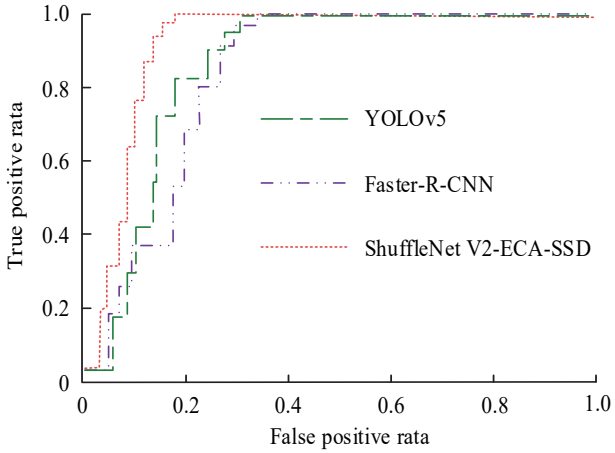


Figure 8 ROC curves of various models

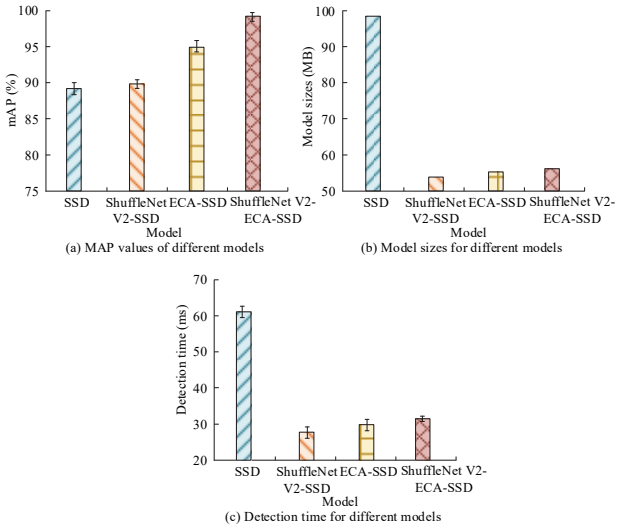


Figure 9 Analysis results of ablation experiment

The study continued with ablation experiments on the ShuffleNet V2-ECA-SSD model, which is compared with the SSD model, the ShuffleNet V2-SSD model, and the ECA-SSD model. The metrics used include mAP value, model size and DT. Fig. 9 displays the outcomes of the ablation tests. In Fig. 9a, the ShuffleNet V2-ECA-SSD model proposed by the study has the highest mAP with a value of 98.52%. The SSD model has the lowest mAP value of only 88.79%. When compared to the SSD model, the mAP of the ECA-SSD model with the ECA module added is noticeably higher, suggesting that the ECA module can successfully raise the DA. In Fig. 9b, the model size of the ShuffleNet V2-SSD model is only 52.47 MB, which are lower than the rest of the models. It indicates that the ShuffleNet V2 module achieves the lightweight of the detection model. Meanwhile, the suggested model in the study satisfies the lightweight criteria because the ShuffleNet V2-ECA-SSD model size is just 55.34 MB, which is only 2.87 MB larger than the ShuffleNet V2-SSD model. In Fig. 9c, the ShuffleNet V2-ECA-SSD model has the lowest DT, which is only 28.41 ms. The ShuffleNet V2-ECA-SSD model is only 32.54 ms, which reduces 71.25 ms and 67.12 ms compared to the SSD model, respectively. It indicates that the ShuffleNet V2-ECA-SSD model has lower detection efficiency with higher DA.

The study concludes by comparing the precision-recall curve (P-R) of each model in the ablation experiments, and the results are shown in Fig. 10. The ShuffleNet V2-ECA-SSD model has a high average precision (AP) of 98.54%, and its P-R AUC is significantly larger than the remaining models. The SSD model has the lowest AP of 87.21% and its P-R AUC is the smallest. The AP value of the ShuffleNet V2-SSD model is only 88.52%, which is a 10.02% decrease compared to the ShuffleNet V2-ECA-SSD model. While the AP value of the ECA-SSD model is 95.35%, the DA is significantly improved compared to the SSD model. It verifies the advantage of the ECA module for detecting small target cracks. Overall, the ShuffleNet V2-ECA-SSD model proposed by the study has higher CD accuracy.

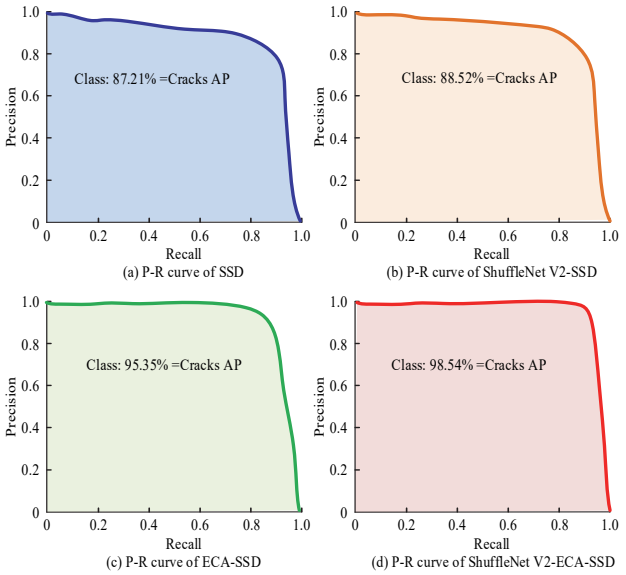


Figure 10 P-R curves of various models

To evaluate the adaptability of the steel plate CD model used in the study in actual dynamic scenarios, a high-speed industrial camera is also used to collect 2000 images of steel plates with cracks in the actual dynamic environment of the production line, and the adaptability of the model is tested under different motion speeds (0.5 m/s, 1 m/s, 2 m/s). Dynamic measurement accuracy is evaluated using mAP values, mean relative error (MRE), and root mean square error (RMSE) to analyze and study the detection performance of the proposed model in the dynamic environment of a production line. The experimental results are shown in Tab. 2. At different motion speeds, the mAP value of the CD model gradually decreases due to the image blurring caused by faster motion speeds, which affects certain detection performance. However, the ShuffleNet V2-ECA-SSD model maintains a high accuracy of over 90% at various motion speeds on the production line. Meanwhile, the MRE of the ShuffleNet V2-ECA-SSD model is relatively low, with only 3.5% at a motion speed of 2.0 m/s. In addition, the AUC values of the model in different motion

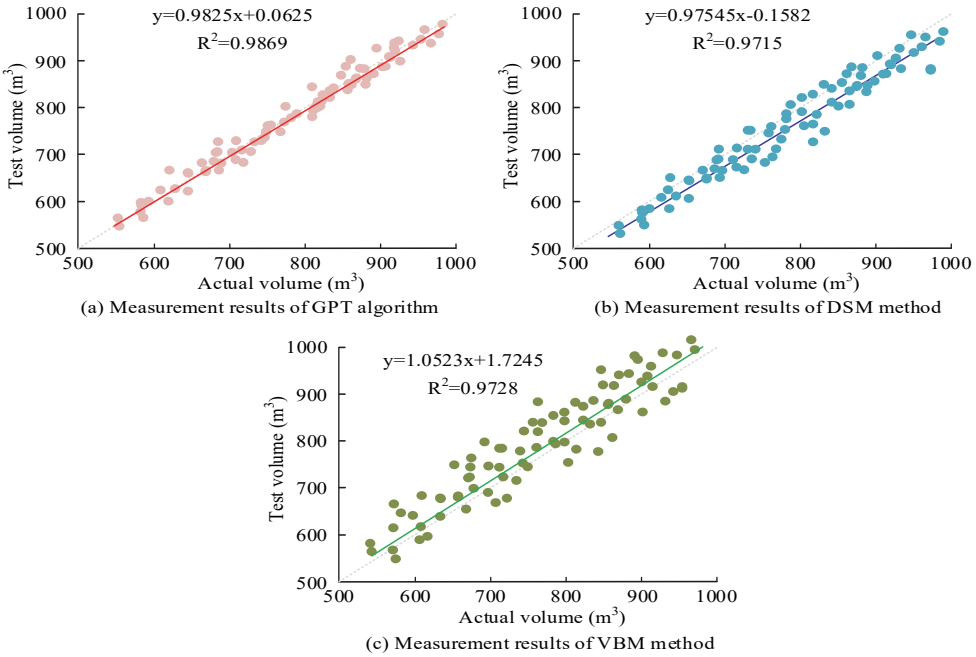
scenarios are all higher than 0.81, indicating high detection stability.

Table 2 Crack detection results in dynamic scenes

Movement speed / m/s	mAP / %	MRE / %	AUC
0.5	97.64	2.1	0.8439
1.0	96.58	2.9	0.8357
2.0	94.85	3.5	0.8196

3.2 Experimental Analysis of Stack Volume Metering Based on GPT Algorithm

The study uses the digital surface model (DSM) and volume balance method (VBM) to compare with the stockpile's volume measurement method based on the GPT algorithm in an attempt to confirm its efficacy. The dataset used comes from the images of stockpiles taken in an iron and steel metallurgical plant, totaling 1894 images, and each stockpile is labeled with the real volume. The study fits the stockpile measurements of each method to the true values, and a regression comparison of the three different methods is performed. The results are shown in Fig. 11. In Fig. 11a, the regression line of the GPT algorithm is closest to the standard regression line. The regression equation (RE) is $y = 0.9825 \cdot x + 0.0625$, and the coefficient of determination (COD) $R^2 = 0.9938$. It indicates that the algorithm's measurements are very close to the real values and the measurement accuracy is high. In Fig. 11b, the RE of DSM method is $y = 0.97545 \cdot x - 0.1582$ and the COD $R^2 = 0.9715$. Compared to the GPT algorithm, its measurement accuracy is marginally poorer. In Fig. 11c, the RE of the VBM method is $y = 1.0523 \cdot x + 1.7245$, and the COD R^2 is 0.9728, which is reduced by 0.021 compared to the GPT algorithm. It indicates that the proposed method of volumetric measurement of the stockpile in the study has a high measurement accuracy.

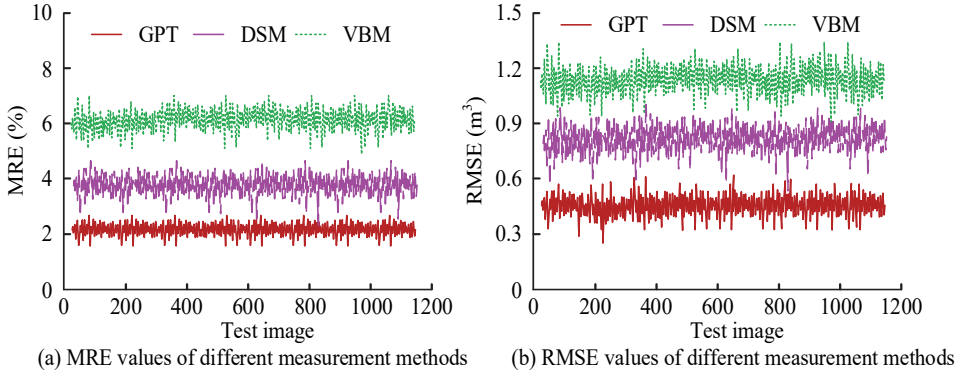


(a) Measurement results of GPT algorithm (b) Measurement results of DSM method

(c) Measurement results of VBM method

Figure 11 Comparison of regression results of different methods

The study's findings are displayed in Fig. 12 and compare the measurement accuracy of each approach using MRE and RMSE. In Fig. 12a, the MRE value of the GPT algorithm proposed in the study is only 2.15%. Whereas, the MRE values of DSM method and VBM method are as high as 3.94% and 6.71%, respectively. In Fig. 12b, the RMSE value of the GPT algorithm is only 0.42 m^3, which is reduced by 0.42 m^3 and 0.75 m^3 compared with the DSM method and the VBM method, respectively. It shows that the proposed GPT algorithm-based stockpile volumetric measurement method of the research has significant performance advantages.



(a) MRE values of different measurement methods (b) RMSE values of different measurement methods

Figure 12 Measurement accuracy of material piles using various methods

4 DISCUSSION AND CONCLUSION

Aiming at CD and stockpile volume measurement in steel metallurgy production, the study proposed ShuffleNet V2-ECA-SSD model and 3D-PC dissection technique based on GPT algorithm. The results revealed that the DT of ShuffleNet V2-ECA-SSD model in steel plate CD was only 29.85 ms, which was reduced by 16.29 ms and 53.23 ms compared to YOLOv5 and Faster-R-CNN model, respectively. The AUC value of the ShuffleNet V2-ECA-SSD model was as high as 0.8576, while the AUC values of the YOLOv5 model and the Faster-R-CNN model were only 0.7536 and 0.6854, respectively. The reason for this was that the YOLOv5 model was difficult to capture the fine differences when dealing with certain complex scenes or small cracks. Faster-R-CNN had high computational complexity in the model, which affected the accuracy and efficiency. In contrast, the detection model proposed by the study had stronger generalization ability and DA. In the ablation experiment, the lowest mAP value of the SSD model was only 88.79%. The mAP value of the ECA-SSD model with the addition of the ECA module was significantly improved compared to the SSD model, which verified the DA of the ECA module. Meanwhile, the ShuffleNet V2-SSD model had the lowest DT of 28.41 ms, which was 71.25 ms less compared to the SSD model. The lightweight advantage of the ShuffleNet V2 module was verified. In addition, the regression line of the GPT algorithm proposed in the study was closest to the standard regression line, with a RE of $y = 0.9825 \cdot x + 0.0625$ and a COD of $R^2 = 0.9938$, which was significantly better than the rest of the methods. The RMSE value of the GPT algorithm of this algorithm was only 0.42 m^3, which was reduced by 0.42 m^3 and 0.75 m^3 compared to the DSM method and VBM method, respectively. It indicated that the proposed ShuffleNet V2-ECA-SSD model and the 3D-PC dissection technique based on GPT algorithm had significant advantages in CD and stockpile volume metrology problems in iron and steel metallurgy production. Nevertheless, the accuracy of the proposed metrology method for 3D-PC profiling in dynamic environment will be affected, and further filtering and optimization algorithms can be introduced to improve the robustness of the point cloud data.

In addition to the field of iron and steel metallurgy, the improved SSD modeling and 3D-PC processing techniques proposed in this study have potential applications in other industrial fields, such as machinery manufacturing and construction. In mechanical manufacturing, these technologies are expected to improve the automation level of component inspection and quality control. In the field of architecture, 3D-PC segmentation and measurement technology can be used for structural analysis and building detection, enabling more accurate building information model generation.

In a static environment, the model proposed by the research can fully leverage its advantages and achieve efficient and accurate CD and material pile volume measurement. However, in the context of steel metallurgy production, the environment is characterized by its complexity. Variables such as equipment vibration, obstruction, and changes in viewing angle can all exert an influence on the accuracy of detection and measurement. Future research can explore the introduction of more advanced noise reduction techniques and dynamic adaptation mechanisms. It can be considered to combine time series data and use time series models in deep learning to better handle changes in dynamic scenes and improve the model's adaptability to unexpected situations. In addition, adaptive learning methods are being explored in response to environmental changes to improve model stability in real-time production. Meanwhile, with the advent of Industry 4.0, real-time data processing and intelligent feedback systems are becoming increasingly important. Future research can further combine edge computing technology to sink computing tasks to the device side to achieve rapid data processing and feedback.

5 REFERENCES

[1] Asadi, P., Mehrabi, H., Asadi, A., & Ahmadi, M. (2021). Deep convolutional neural networks for pavement crack detection using an inexpensive global shutter RGB-D sensor and ARM-based single-board computer. *Transportation Research Record*, *2675*(9), 885-897. https://doi.org/10.1177/03611981211004974

[2] Ha, J., Park, K., & Kim, M. (2021). A development of road crack detection system using deep learning-based segmentation and object detection. *The Journal of Society for e-Business Studies*, *26*(1), 93-106. https://doi.org/10.7838/JSEBS.2021.26.1.093

[3] Safaei, B., Onyibo, E. C., Goren, M., Kotrasova, K., Yang, Z. C., Arman, S., & Asmael, M. (2023). Free vibration investigation on RVE of proposed honeycomb sandwich beam and material selection optimization. *Facta Universitatis, Series: Mechanical Engineering*, *21*(1), 31-50. https://doi.org/10.22190/FUME220806042S

[4] Yuan, C., Xiong, B., Li, X., Sang, X., & Kong, Q. (2022). A novel intelligent inspection robot with deep stereo vision for three-dimensional concrete damage detection and quantification. *Structural health monitoring*, *21*(3), 788-802. https://doi.org/10.1177/14759217211010238

[5] Chen, D. R. & Chiu, W. M. (2023). Deep-learning-based road crack detection frameworks for dashcam-captured images under different illumination conditions. *Soft Computing*, *27*(19), 14337-14360. https://doi.org/10.1007/s00500-023-08738-0

[6] Xu, M. (2023). Solar cell defect detection based on improved G-SSD Network. *International Journal of Energy*, *2*(1), 68-71. https://doi.org/10.54097/ije.v2i1.5618

[7] Meng, S., Gao, Z., Zhou, Y., He, B., & Djerrad, A. (2023). Real-time automatic crack detection method based on drone. *Computer-Aided Civil and Infrastructure Engineering*, *38*(7), 849-872. https://doi.org/10.1111/mice.12918

[8] Nguyen, S. D., Tran, T. S., Tran, V. P., Lee, H. J., Piran, M. J., & Le, V. P. (2023). Deep learning-based crack detection: A survey. *International Journal of Pavement Research and Technology*, *16*(4), 943-967. https://doi.org/10.21275/sr24805160228

[9] Xiang, X., Wang, Z., & Qiao, Y. (2022). An improved YOLOv5 crack detection method combined with transformer. *IEEE Sensors Journal*, *22*(14), 14328-14335. https://doi.org/10.3390/math11102377

[10] Ye, W., Ren, J., Zhang, A. A., & Lu, C. (2023). Automatic pixel-level crack detection with multi-scale feature fusion for slab tracks. *Computer-Aided Civil and Infrastructure Engineering*, *38*(18), 2648-2665. https://doi.org/10.1111/mice.12984

[11] Xiang, L., Tang, L., Gai, J., & Wang, L. (2021). Measuring stem diameter of sorghum plants in the field using a high-

throughput stereo vision system. *Transactions of the ASABE*, *64*(6), 1999-2010. https://doi.org/10.13031/trans.14156
[12] Vargas, R., Marrugo, A. G., Zhang, S., & Romero, L. A. (2020). Hybrid calibration procedure for fringe projection profilometry based on stereo vision and polynomial fitting. *Applied optics*, *59*(13), D163-D169. https://doi.org/10.1364/AO.383602
[13] Silva, C., Aires, R., & Rodrigues, F. (2024). A compact underwater stereo vision system for measuring fish. *Aquaculture and Fisheries*, *9*(6), 1000-1006. https://doi.org/10.1016/j.aaf.2023.03.006
[14] Li, H. & Zhang, B. (2021). Application of integrated binocular stereo vision measurement and wireless sensor system in athlete displacement test. *Alexandria Engineering Journal*, *60*(5), 4325-4335. https://doi.org/10.1016/j.aej.2021.02.033
[15] Li, H., Wang, W., Wang, M., Li, L., & Vimlund, V. (2022). A review of deep learning methods for pixel-level crack detection. *Journal of Traffic and Transportation Engineering (English Edition)*, *9*(6), 945-968. https://doi.org/10.1016/j.jtte.2022.11.003
[16] Zhu, G., Liu, J., Fan, Z., Yuan, D., Ma, P., Wang, M., & Wang, K. C. (2024). A lightweight encoder-decoder network for automatic pavement crack detection. *Computer-Aided Civil and Infrastructure Engineering*, *39*(12), 1743-1765. https://doi.org/10.1111/mice.13103
[17] Kim, B., Yuvaraj, N., & Sri, P. K. R. (2021). Surface crack detection using deep learning with shallow CNN architecture for enhanced computation. *Neural Computing and Applications*, *33*(15), 9289-9305. https://doi.org/10.1007/s00521-021-05690-8
[18] Xiang, L., Gai, J., Bao, Y., Yu, J., Schnable, P. S., & Tang, L. (2023). Field-based robotic leaf angle detection and characterization of maize plants using stereo vision and deep convolutional neural networks. *Journal of Field Robotics*, *40*(5), 1034-1053. https://doi.org/10.1002/rob.22166
[19] Adil, E., Mikou, M., & Mouhsen, A. (2022). A novel algorithm for distance measurement using stereo camera. *CAAI Transactions on Intelligence Technology*, *7*(2), 177-186. https://doi.org/10.1049/cit2.12098
[20] Wang, P., Wang, Z., Xin, S., Gao, X., Wang, W., & Tu, C. (2022). Restricted delaunay triangulation for explicit surface reconstruction. *ACM Transactions on Graphics*, *41*(5), 1-20. https://doi.org/10.1145/3533768
[21] Chen, H., Wu, H., Yang, N., Huang, H., & Liang, W. (2024). Irregular object measurement method based on improved adaptive slicing method. *Multimedia Tools and Applications*, *83*(17), 50557-50580. https://doi.org/10.1007/s11042-023-17342-1

Contact information:

Wei DENG
(Corresponding author)
Intelligent Manufacturing Division,
WISDRI Engineering and Research Incorporation Limited,
Wuhan, 430223, Hubei, China
E-mail: WeiDeng_Wisdri@163.com

Bingquan ZHU
Intelligent Manufacturing Division,
WISDRI Engineering and Research Incorporation Limited,
Wuhan, 430223, Hubei, China
E-mail: 10242@wisdri.com