

*Scientific Paper*Doi: <http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v42n3e20220010/2022>**DYNAMIC POTATO IDENTIFICATION AND CLEANING METHOD  
BASED ON RGB-D****Xiaoming Fu<sup>1</sup>, Zhijun Meng<sup>2\*</sup>, Zedong Wang<sup>1</sup>, Xiaohui Yin<sup>3</sup>, Chang Wang<sup>4</sup>**<sup>2\*</sup>Corresponding author. Beijing Research Center for Information Technology in Agriculture / Beijing, China.E-mail: [zhijunmeng@126.com](mailto:zhijunmeng@126.com) | ORCID: <https://orcid.org/0000-0001-5737-0684>**KEYWORDS**

potato, dynamic identification, cleaning method, RGB-D, Mask R-CNN.

**ABSTRACT**

To solve the problems of a large number of clods remaining in potatoes after mechanized harvesting in northern heavy clay soil planting areas in China and requiring much labor to separate clods from potatoes, which leads to a heavy workload, inefficiency and a low cleaning rate, an RGB-D-based Mask R-CNN dynamic potato identification model is established by using acquired RGB-D image data of untreated potatoes after mechanized harvesting, and a potato cleaning method is presented in this paper. This makes it possible to automatically separate clod impurities from potatoes. The experimental results showed that the prediction accuracy of the identification model is more than 97%. With the increase in cleaning conveyance speed, the prediction accuracy of the model and the actual cleaning precision show a downward trend. Comprehensively considering the potato cleaning efficiency and accuracy, when the speed is set to  $0.4 \text{ m}\cdot\text{s}^{-1}$ , the cleaning precision reaches as high as 96.35%. This research provides a method and theoretical reference for the further study of intelligent potato cleaning systems.

**INTRODUCTION**

At present, the potato planting area in China is approximately 5.6 million acres, and it is one of the four major food crops in China (Xu et al., 2021). With the continuous development of the potato industry, the planting area has expanded unceasingly, and mechanical harvesting has become the main potato harvesting method that lowers the cost of harvesting (Zhao, 2020; Cunha et al. 2011), but there is still a problem that there are many impurities, such as clods and stones, which are similar to the potato in size and quality, which makes it difficult to separate these impurities from the potatoes after mechanized harvesting (Li et al., 2019; Zhang et al., 2021). Especially in the northern heavy clay soil area, the soil viscosity is high, so there is a high impurity rate of soil clods in potatoes after harvesting, which increases the difficulty in separating clods from potatoes (Lu et al., 2017). This leads to further manual sorting and cleaning being needed, which increases the production cost (Zhao et al., 2019b).

In recent years, domestic and foreign scholars and agricultural machinery manufacturing enterprises have conducted much research on potato cleaning methods and equipment. At present, potato cleaning methods include screen type, roller type, pneumatic type and noncontact online detection to realize potato cleaning classification (Wei et al., 2017; Xie et al., 2019; Gan-Mor et al., 1986; Hosainpour et al., 2011). Lv et al. (2021) designed an impurity removal device for a potato receiving hopper. Through the adjustable spiral roller impurity removal device, impurity separation on the potato surface was realized through the rotation friction of the impurity removal roller. Geng et al. (2021) designed a potato-stone separator based on airflow suspension technology for a potato cleaner, which used high-speed airflow suspension and a vibrating screen swing to separate potatoes and stones in the movement process. Al-Mallahi et al. (2008) used machine vision to identify the soil clods in mechanically harvested potatoes and extract color image features for target classification. Geng et al. (2019) used the laser backscattering imaging

<sup>1</sup> College of Engineering, Heilongjiang Bayi Agricultural University/ Daqing, China.<sup>2</sup> Beijing Research Center for Information Technology in Agriculture/ Beijing, China.<sup>3</sup> Daqing Oilfield Co Ltd Fourth Oil Production Plant Planning and Design Research Institute/ Daqing, China.<sup>4</sup> College of Science, Heilongjiang Bayi Agricultural University/ Daqing, China.

Area Editor: Gizele Ingrid Gadotti

Received in: 1-20-2022

Accepted in: 6-1-2022



(LBI) detection method to detect potato-containing impurities and removed clods and stones in the potato dropping process through the impurity removal mechanism. All these studies have solved the problem of cleaning and sorting potatoes with remaining impurities. However, for the main potato-producing areas in northern China, the potatoes after harvest contain a large number of impurities and soil clods (Zhao et al., 2019a; Yang et al., 2020). The quality and size of some soil clods are similar to those of potatoes, while some potatoes have soil on their surface, so the difference between image characteristics and soil clods is not significant. All these problems directly affect potato cleaning quality.

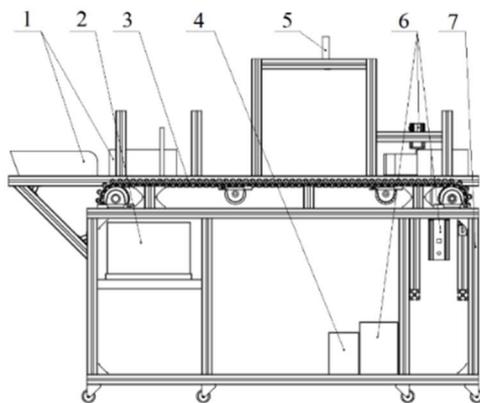
To solve the above problems, the Mask R-CNN dynamic potato identification model based on RGB-D (red green blue depth, RGB-D), which is three-dimensional color image data containing depth information, is established, and

the cleaning control method is presented in this paper. The cleaning effect is verified by the test platform. The identification and separation of impurities in potatoes containing impurities can be realized.

## MATERIAL AND METHODS

### Basic structure of the potato cleaning testing platform and cleaning method

The structure of the potato cleaning testing platform is shown in Figure 1 (a), which includes a feeding device 1, a conveying device 3, an image acquisition system 5, a master computer 2, a control system 4, a cleaning servo system 6 and other major components. Its physical diagram is shown in Figure 1 (b).



(a) Structure Drawing

1. Feeding Device
2. Master Computer
3. Conveying Device
4. Control System
5. Image Acquisition System
6. Cleaning Servo System
7. Body Frame



(b) Practicality Picture

FIGURE 1. Experimental platform of the potato intelligent cleaning system.

During the potato cleaning operation, the potatoes containing impurities are loaded into the feeding device 1. The master computer 2 sends the start signal to the control system 4, and the servo controller controls the conveying device 3 to run at the specified speed. The conveying device 3 transports potatoes with clods to the rear of this platform. After passing through the double baffle, potatoes or clods are transported backward in a line. When potatoes with impurities are conveyed into the image acquisition area, the depth camera of the image acquisition system 5 collects the target images. The master computer 2 begins to classify and predict the target image based on the RGB-D potato image dynamic identification model, and the computer sends the identification result data to the control system 4, which controls the stepper motor of the cleaning servo system 6 by category and location signal. When the target is identified as a potato or clod, the controller controls the stepper motor to work based on the cleaning control model. When the target nears the location of the clearing rotating board, the stepper motor drives the rotating board to rotate clockwise or counterclockwise by a certain angle, and then the potato or clod moves along the rotating board and the guide board right or left to leave the conveyor to separate potatoes and clod impurities that are similar to the potatoes in size and quality.

### Mask R-CNN dynamic potato identification model based on RGB-D

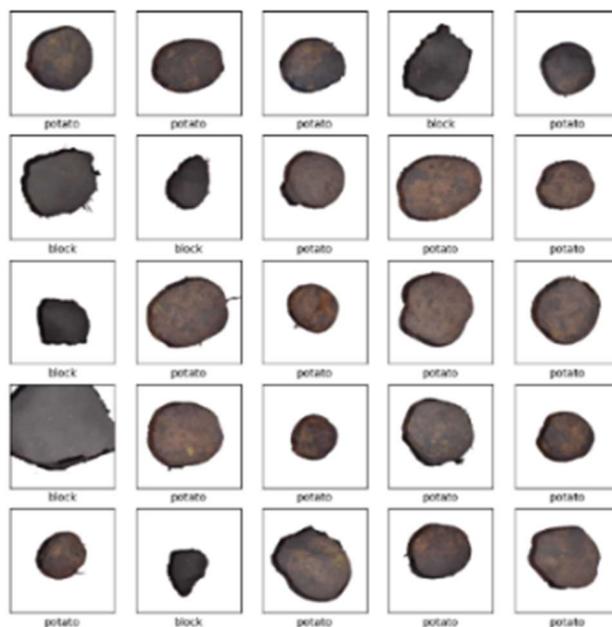
The test samples are potatoes that contain impurities after mechanized harvest in Longzhen Farm, Heilongjiang Province, China, which is a heavy clay soil planting area. Most residual stems and some soil clods were removed after mechanized harvest. However, there are still a large number of soil clods that are difficult to remove. By testing the impurity contents of 500 kg samples after mechanized harvest in the area, the soil clods were 18.22% of the samples, the stones were 0.25%, and the residual stems were 1.17%. The potato cleaning ratio was only 80.36%, which makes it possible for these soil clods to be taken to the factory and leads to partial soil loss. Thus, the dynamic potato identification and cleaning method using three-dimensional machine vision technology to separate soil clods from potatoes is proposed here. This identification model reflects the surface three-dimensional features of the target by adding depth images to improve potato recognition accuracy.

The RGB images and depth images of the samples were dynamically obtained by a Kinect DK depth camera mounted on the test platform of the potato cleaning system, and the images were merged to form four channels of RGB-D image data for sample storage, of which 1,000

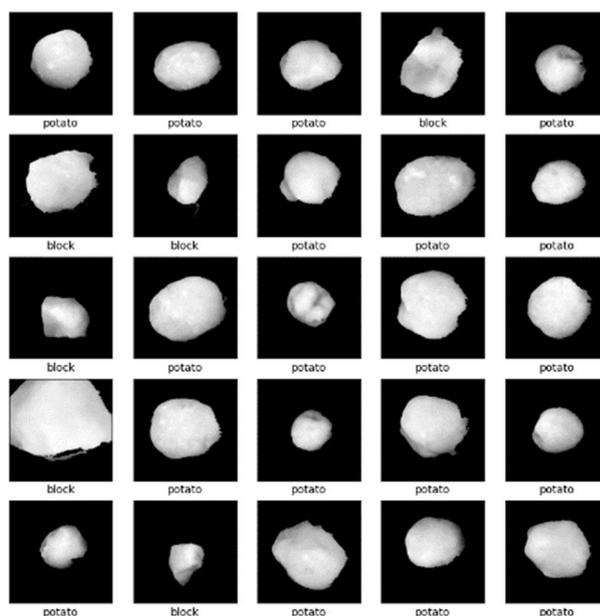
potato RGB-D samples were collected. To improve the model recognition accuracy and stability, 500 clod RGB-D samples were further collected.

To improve the dynamic recognition ability of the model, all samples were strengthened after data preprocessing. The collected images of potatoes and clods were flipped, and the motion blur, chromaticity and brightness of images were adjusted so that 9,000 RGB-D image samples of potato and 4,500 RGB-D image samples of clod were received. Among

them, 1,500 potato samples and 750 soil clod samples were randomly selected to construct the model test set, and the remaining samples were used as the model training set. There was no intersection between the training set and the test set. A portion of the samples is shown in Figure 2. For the convenience of visual display, the 4-channel data were converted into PNG format for visual display (as shown in Figure 2 (a)), and Figure 2 (b) shows the grayscale images of the single-channel depth data of the samples.



(a) RGB-D Images



(b) Gray images of depth data

FIGURE 2. Visualization of Partial Sample Data.

A mask region-convolutional neural network (Mask R-CNN) (He et al., 2020) integrates the improved faster region-convolutional neural network (Faster R-CNN) (Ren et al., 2017) and the fully convolutional network (FCN)

(Long et al., 2017). Feature pyramid networks (FPNs) and anchor technology were used to deal with the detection of targets at a large scale (Dollar et al., 2014; Lin et al., 2017). The Mask R-CNN recognition model based on potato

RGB-D images proposed in this paper first extracts the feature images of the input images and then generates two outputs that are the anchor category judged as foreground or background and the result of border fine-tuning, which

enables the bounding box to better fit the target through a region proposal network (RPN). Then, the ROI is segmented through the region of interest align (ROI align) layer. The model structure is shown in Figure 3.

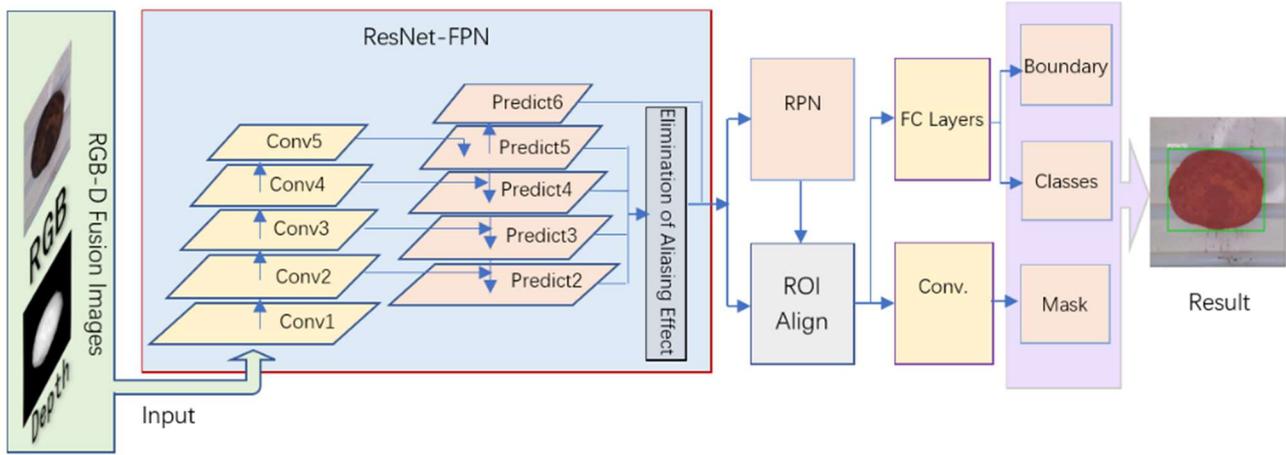


FIGURE 3. Model structure of potato RGB-D image recognition based on Mask R-CNN

In this paper, the ResNet-101 network is used to construct the ResNet-FPN backbone network for image feature extraction (He et al., 2016; Xie et al., 2017). Feature depth is divided into five stages according to the size of the feature map. The last layer of each stage outputs Conv1, Conv2, ..., Conv5. The FPN horizontal connection fuses the convolution operation and upsampling results of each layer in Conv2, Conv3, Conv4, and Conv5 to generate Predict2, Predict3, ..., Predict6 (shown in Figure 4). The fusion results are convoluted to eliminate the aliasing effect, they are used as the input of the RPN network, and regions of interest (ROIs) are separated from the feature map and input to ROI align according to the generated region proposal. The bilinear interpolation method was used to calculate the pixel value of the feature map (Author, 1991; Gribbon & Bailey, 2004). After max pooling, the fixed-size feature map was output (Ou et al., 2019; Bai et al., 2020). Through the fulling convolutional layers (FC layers), the category and boundary coordinate information of the target features are output. Additionally, the ROI align output results are processed through the convolution layer, and then the binary mask is used to generate the masks of target categories. Finally, the sigmoid function (Han & Moraga, 1995) is used to process and calculate the binary loss for each category to avoid competition between categories (Zhang et al., 2020).

The loss function of this model is:

$$L([p_i], [t_i], [m_i]) = L_{cls} + L_{box} + L_{mask} \quad (1)$$

where:

$$L_{cls} = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) \quad (2)$$

$$L_{box} = \lambda \frac{1}{N_{box}} \sum_i p_i^* L_{box}(i, t_i^*) \quad (3)$$

$$L_{mask} = \frac{1}{N_{mask}} \sum_i L_{mask}(m_i) \quad (4)$$

where:

$L$  is the loss;

$L_{cls}$  is the classification loss;

$L_{box}$  is the regression loss;

$L_{mask}$  is the mask loss;

$N_{cls}$  is the number of categories;

$N_{box}$  is the number of anchors;

$N_{mask}$  is the number of masks;

$t_i$  is the prediction offset of the anchor;

$t_i^*$  is the actual offset of the anchor;

$p_i$  is the probability that the anchor is predicted as a target;

$p_i^*$  is the probability that the anchor is predicted as a nontarget, and

$m_i$  is the confidence of the predicted target.

### The cleaning control model

The position of the target center, the anchor frame center, can be determined by the Mask R-CNN target boundary coordinate information. The calculation results of the recognition model are transmitted to the servo controller through the master control computer. When the controller receives the target category and position signals, the stepper motor is controlled by the cleaning control model to execute the corresponding instructions to realize the separation of potatoes and soil clods.

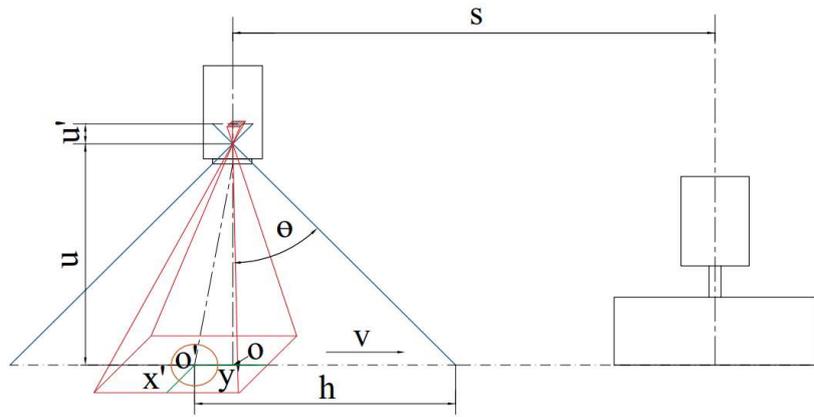


FIGURE 4. Schematic of the target position in the view field of the camera.

Set the coordinate of the center point of the anchor frame of the identified object as  $(x_i, y_i)$  and the vertical pixel of the image as  $p_v$ . If the current target is the  $i_{th}$  object, the vertical pixel distance between the target center and the image center point can be calculated as  $C_i$ :

$$c_i = \frac{1}{2} p_v - y_i \quad (5)$$

Given that the visual angle of the camera is  $\theta$ , the distance between the camera lens and the conveying platform is  $u$  (as shown in Figure 4). The actual horizontal distance between the target object and the center of the camera sensor within the camera's field of view can be calculated as  $C_i$ :

$$C_i = c_i \left( \frac{2u \tan \theta}{p_v} \right) \quad (6)$$

Let the horizontal distance between the center of the image sensor and the axis of the cleaning motor be  $S$ , and let the angular velocity of the main shaft of the conveying device measured by the rotary encoder be  $\omega$ . The radius of

the main sprocket is  $r$ . The vector of conveying time  $[t_i]$  from the recognition position of the target to the position of the cleaning stepper motor can be calculated by eqs (5) and (6) as:

$$[t_i] = \left[ s + \left( \frac{1}{2} p_v - y_i \right) \cdot \left( \frac{2u \tan \theta}{p_v} \right) \right] / \omega \cdot r \quad (7)$$

where:

$S$  is the horizontal distance between the image sensor and the motor axis;

$p_v$  is the number of vertical pixel points in the image;

$y_i$  is the ordinate of the target center point;

$u$  is the distance between the camera lens and object;

$\theta$  is the visual angle of the camera;

$\omega$  is the angular velocity of the main shaft of the conveying device, and

$r$  is the radius of the main sprocket.

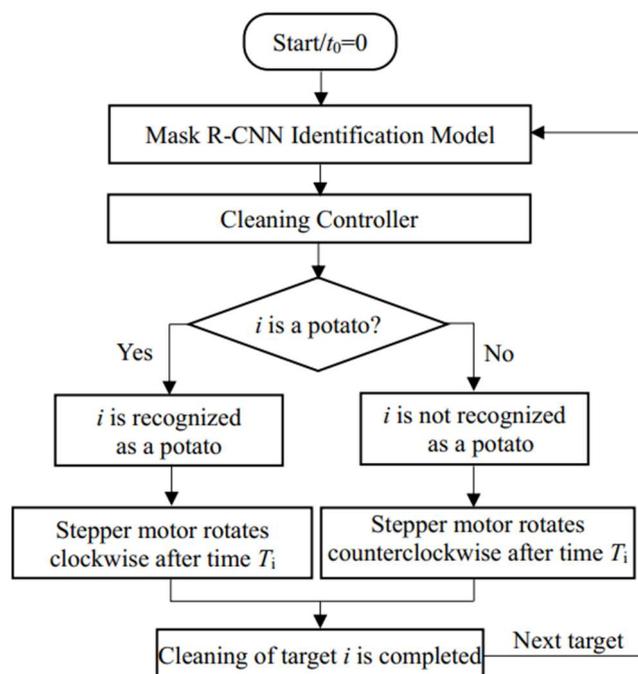


FIGURE 5. Potato cleaning control flowchart.

The time when the system starts to work is denoted by  $t_0$ . The cleaning controller receives the starting operation signal. The prediction time of the identification model is denoted by  $t_{pi}$ . The time interval  $T_i$  that is applied to the intermittent control of the stepper motor when the target nears clearing the rotating board can be obtained as:

$$T_i = t_i - t_{i-1} - t_{pi}, \quad t_0 = 0. \quad (8)$$

The cleaning controller processes different categories of targets based on  $T_i$  ( $i=1,2,3,L,n$ ) to control the stepper motor, and its control process is shown in Figure 5.

## RESULTS AND DISCUSSION

The testing platform is equipped with a portable graphics workstation, Ubuntu 18.04.5LTS operating system, 16 GB memory, Quadro RTX3000 GPU, and 6 GB video memory capacity. All model training and testing in this paper are carried out in this environment. The model

parameters were set as follows: the number of iterations was 100, the number of iteration rounds was 10, and the learning rate was 0.001. The training dataset of the potato Mask R-CNN model was established. To determine the appropriate size of the training set, the influence of the training data size on the prediction accuracy was tested. From Figure 6, when the dataset size was fixed, the distribution of test accuracy for each percentage of division for training and test data was displayed. As seen from the distribution, with the increase in the proportion of the training data, the distribution of the accuracy of the test set decreases sharply, and the accuracy is better. Too small a proportion of the training data results in lower test accuracy, the reason for which may be because the training set is not sufficient to represent the target features. The average accuracy of the model decreased slightly when the percentage of division for training and test data increased from 85% to 90%, probably because the model tends to overfit the training examples. Therefore, the proportion of the training data was set to 85%.

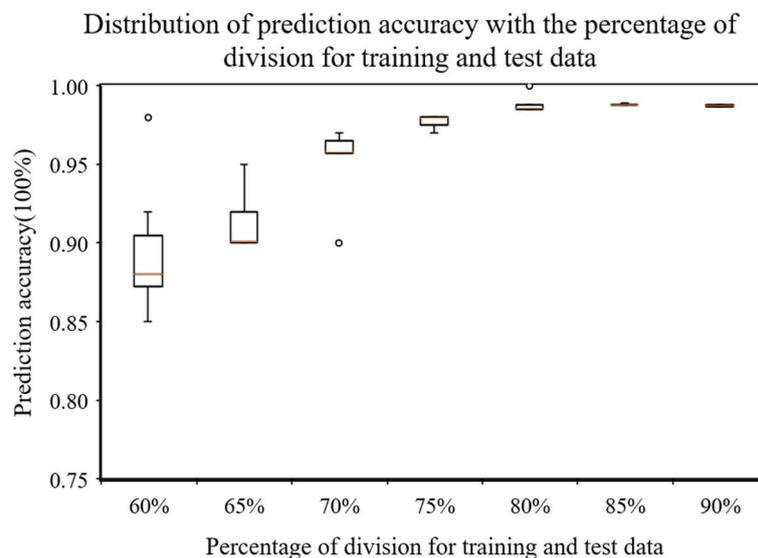
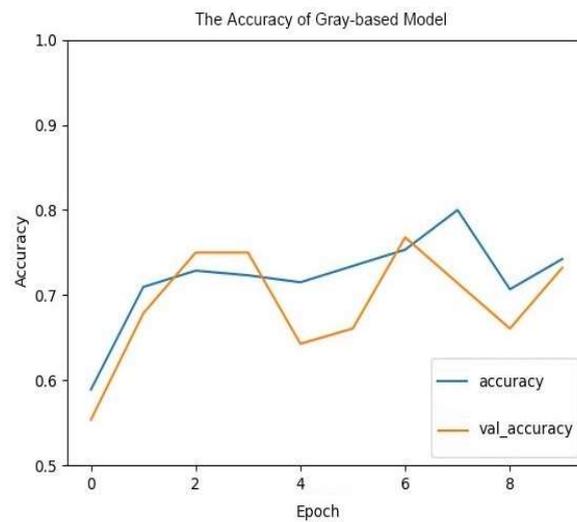


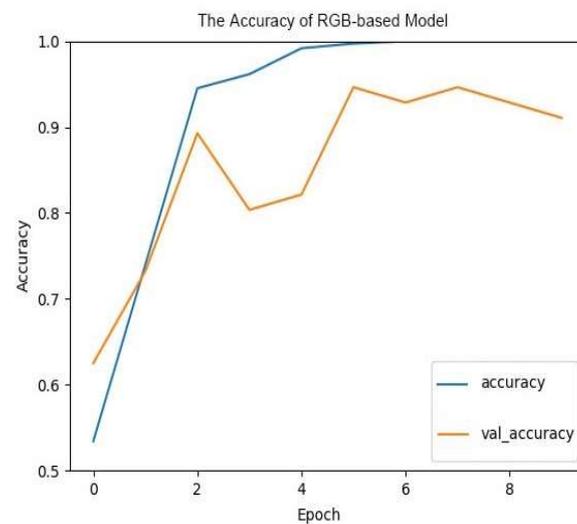
FIGURE 6. Distribution of prediction accuracy with the percentage of division for training and test data.

To analyze the dynamic potato identification model based on RGB-D, the RGB-D-based, RGB-based and gray-based Mask R-CNN models were trained on the potato RGB-D, RGB and gray image data of the same sample in the same model framework environment. As seen in the accuracy convergence curve (as shown in Figure 7), with the increase in the number of training iterations, the accuracy of the gray-based Mask R-CNN model is approximately 70%, which is caused by the insignificant difference between the gray characteristics of potato and soil clods. Therefore, the

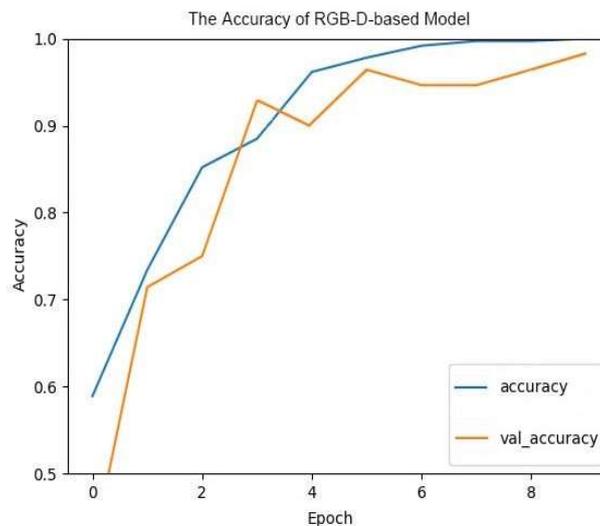
identification accuracy of the potato recognition model constructed from gray images is poor. The prediction accuracy of both RGB-based and RGB-D-based Mask R-CNN models can reach more than 90%. However, because the RGB-D-based Mask R-CNN model contains RGB color information as well as the depth characteristics of the target, the RGB-D Mask R-CNN model has a recognition accuracy of 97% for the test set that does not participate in model training, and its generalization ability is higher than that of the RGB-based Mask R-CNN model.



(a) Gray-based CNN model



(b) RGB-based CNN model



(c) RGB-D-based CNN model

FIGURE 7. Accuracy convergence curve of Mask R-CNN model training and testing.

To determine the best conveying speed, we take the conveying speed as a single factor to test the effect of the conveying speed on the Mask R-CNN model target recognition rate for all test samples. The average recognition rate is  $R$  as follows:

$$R = \frac{1}{T} \sum_i \left( \frac{N - n_i}{N} \times 100\% \right) \tag{9}$$

where:

$T$  is the number of test repetitions;

$N$  is the total number of test samples, and

$n_i$  is the number of targets that are not recognized.

From Figure 8, when the conveying speed is more than  $0.6 \text{ m}\cdot\text{s}^{-1}$ , the target recognition rate decreases with increasing conveying speed. When the conveying speed is more than  $1 \text{ m}\cdot\text{s}^{-1}$ , the recognition rate begins to decrease significantly. Figure 9 intuitively shows that targets can be identified stably under the lower conveying speed. However, when the conveying speed is greatly improved, a few targets cannot be recognized. Therefore, the value range of the conveying speed parameter is set from  $0.2 \text{ m}\cdot\text{s}^{-1}$  to  $1.0 \text{ m}\cdot\text{s}^{-1}$ .

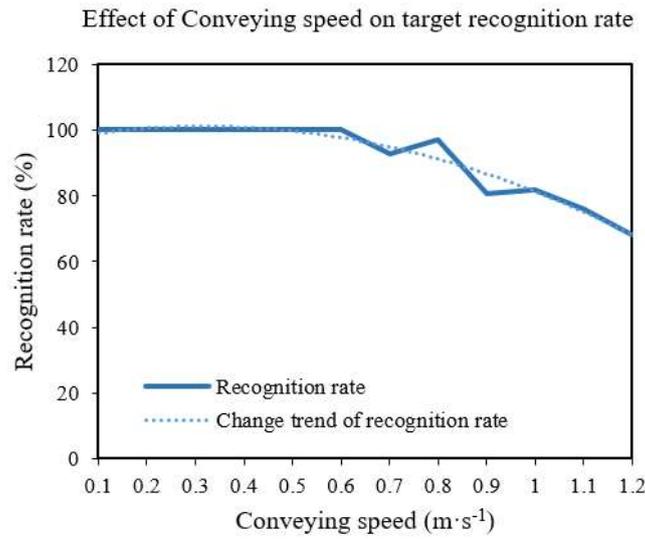
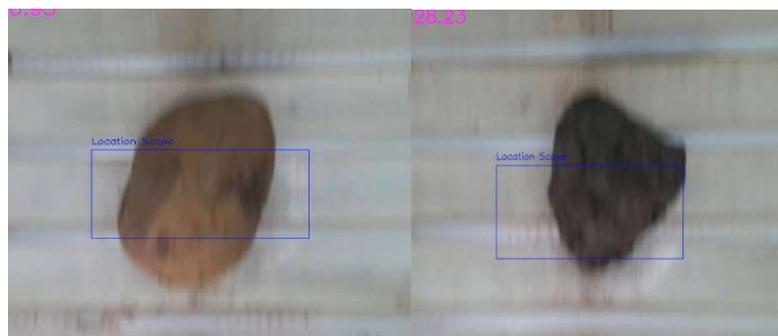


FIGURE 8. Effect of the conveying speed on the target recognition rate.



(a) All potatoes and clods are identified under the lower conveying speed



(b) A few potatoes and clods are not identified under the high conveying speed

FIGURE 9. States of identifying targets under the influence of conveying speed.

The parameters of the test platform remain unchanged, and the cleaning system adopts the potato Mask R-CNN identification model based on RGB-D to test the cleaning control model at different cleaning and conveying

speeds. Each group of speed parameters was repeated 20 times. The average prediction accuracy, average prediction time and average cleaning precision were obtained as shown in Table 1.

TABLE 1. Test result of cleaning.

No.	Conveying speed ( $m \cdot s^{-1}$ )	Average prediction accuracy (%)	Average prediction time (s)	Average cleaning precision (%)
1	0.2	97.85	0.15	96.27
2	0.4	97.43	0.16	96.35
3	0.6	93.58	0.15	90.14
4	0.8	81.24	0.18	70.03
5	1.0	70.43	0.17	45.21

As seen in Table 1, with the increase in conveying speed, the average prediction time has no significant difference. The average prediction accuracy and cleaning precision show a decreasing trend. When the conveying speed exceeds  $0.4 m \cdot s^{-1}$ , the average cleaning precision decreases significantly. Although the sample data are processed by motion blur, the larger increase in the target moving speed leads to the weakening of the image features collected by the current equipment. When the speed reaches  $1.0 m \cdot s^{-1}$ , the target movement interval time is close to the total prediction time of the dynamic identification model and the response time of the controller, which results in a significant decrease in the cleaning precision. Considering the accuracy and efficiency of cleaning, the conveying speed should be improved while the cleaning precision meets the requirements. According to the national standard *Code of Practice for Grading and Inspecting of Commercial Potatoes* (GB/T 31784-2015) (State Administration for Market Regulation, 2015), the potato cleaning percentage is required to be more than 94%. Therefore, when the speed is  $0.4 m \cdot s^{-1}$ , the average prediction accuracy is 97.43%, and the average cleaning precision is 96.35%. The cleaning accuracy and efficiency are higher than those of the traditional potato cleaning method. The cleaning performance was better than that in the national standard, which can meet the actual cleaning operation requirements.

## CONCLUSIONS

1. In this paper, the RGB-D potato image was obtained by a depth camera, and the dynamic identification model based on RGB-D was established. Through the established cleaning control model, the automatic separation of soil clod impurities from potatoes by a cleaning system was realized.

2. The RGB-D dataset was constructed based on the original potato images with impurities after mechanized harvest. Through the RGB-D-based, RGB-based and gray-based potato Mask R-CNN model tests, the results show that the target recognition accuracy of the Mask R-CNN model based on RGB-D is not less than 97%.

3. The cleaning control model was established based on the time it takes for the target object to move from the

identified location to clearing the rotating board. The testing results show that with the increase in the conveying speed of cleaning, the prediction accuracy and the actual cleaning precision have a decreasing trend. When the speed is  $0.4 m \cdot s^{-1}$ , the average prediction accuracy is 97.43%, and the average cleaning precision reaches 96.35%.

## ACKNOWLEDGMENTS

This research was funded by the National key Research and Development Program of China (No. 2017YFD0700705, No. 2016YFD0701600); National Grass-roots Agricultural Technology Promotion Support Project of China (No. JCTG17-04); Program for Young Scholars with Creative Talents of Heilongjiang Bayi Agricultural University (No. CXRC2017017); Natural Science Talent Support Plan of Heilongjiang Bayi Agricultural University (No. ZRCPY202022).

## REFERENCES

- Al-Mallahi A, Kataoka T, Okamoto H (2008) Discrimination between potato tubers and clods by detecting the significant wavebands. *Biosystems Engineering* 100:329-337.
- Author U (1991) Different view: Digital image warping. *IEEE Computer Graphics & Applications* 11:4-5.
- Bai T, Pang Y, Wang J, Han K, Luo J, Wang J, Lin J, Wu J, Zhang H (2020) An optimized faster R-CNN method based on DRNet and roi align for building detection in remote sensing images. *Remote Sensing* 12:762.
- Cunha JPAR da, Martins DH, Cunha WG da (2011) Operational performance of the mechanized and semi-mechanized potato harvest. *Engenharia Agricola* 31(4): 826-834.
- Dollar P, Appel R, Belongie S, Perona P (2014) Fast feature pyramids for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36:1532-1545.
- Gan-Mor S, Zacharin A, Galili N, Feller R, Magolin E (1986) Absorbing stone impact to enable separation from potatoes. *Transactions of the ASAE - American Society of Agricultural Engineers* 29:1526-1529.

- Geng DY, Su GL, Wei ZC, Tan DL, Li XQ, Liu Y (2021) Design and experiment of potato-stone separator based on airflow suspension technology. *Transactions of the Chinese Society for Agricultural Machinery* 52:102-110.
- Geng J, Xiao L, He X, Rao X (2019) Discrimination of clods and stones from potatoes using laser backscattering imaging technique. *Computers and Electronics in Agriculture* 160:108-116.
- Gribbon KT, Bailey DG (2004) A novel approach to real-time bilinear interpolation. In: *IEEE International Workshop on Electronic Design, Test and Applications*. Perth, IEEE, Proceedings...
- Han J, Moraga C (1995) The Influence of the sigmoid function parameters on the speed of backpropagation learning. In: *International Workshop on Artificial Neural Networks: from Natural to Artificial Neural Computation*. Proceedings...
- He KM, Gkioxari G, Dollar P, Girshick R (2020) Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42: 386-397.
- He KM, Zhang XY, Ren SQ, Sun J (2016) Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*:770-778.
- Hosainpour A, Komarizade MH, Mahmoudi A, Shayesteh MG (2011) High speed detection of potato and clod using an acoustic based intelligent system. *Expert Systems with Applications* 38:12101-12106.
- Li ZH, Wen XY, Lu JQ, Li JQ, Yi SJ, Qiao D (2019) Analysis and prospect of research progress on key technologies and equipments of mechanization of potato planting. *Transactions of the Chinese Society for Agricultural Machinery* 50: 1-16.
- Lin TY, Dollar P, Girshick R, He K, Hariharan B, Belongie S (2017) Feature pyramid networks for object detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 936-944.
- Long J, Shelhamer E, Darrell T (2017) Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39(4):640-651.
- Lu JQ, Sun H, Dui H, Pang MM, Yu JY (2017) Design and experiment on conveyor separation device of potato digger under heavy soil condition. *Transactions of the Chinese Society for Agricultural Machinery* 48:146-155.
- Lv JQ, Du CL, Liu ZY, Li JC, Li ZH, Li ZY (2021) Design and test of impurity removal device of potato receiving hoppe. *Transactions of the Chinese Society for Agricultural Machinery* 52:82-90+61.
- Ou P, Lu K, Zhang Z, Liu ZY (2019) Object recognition and spatial localization based on Mask RCNN. *Computer Measurement & Control* 27: 172-176.
- Ren SQ, He KM, Girshick R, Sun J (2017) Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 39:1137-1149.
- State Administration for Market Regulation (2015) Code of practice for grading and inspecting of commercial potatoes. National Standard GB/T 31784-2015.
- Wei ZC, Li XQ, Sun CZ, Liu WZ, Zhang YF (2017) Analysis of potato mechanical damage in harvesting and cleaning and sorting storage. *Journal of Agricultural Science and Technology* 19: 63-70.
- Xie S, Girshick R, Dollar P, Tu Z, He K (2017) Aggregated residual transformations for deep neural network. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 5987-5995.
- Xie SS, Wang CG, Deng WG (2019) Experiment of a swing separating sieve on a potato digger. *Engenharia Agricola* 39:548-554.
- Xu N, Zhang HL, Zhang RH, Xu YK (2021) Situation and prospect of potato planting in China. *Chinese potato Journal* 35:81-96.
- Yang HG, Xie HX, Yan JC, Wei H, Wang HO, Zhang HJ (2020) Overview of mechanized cleaning technology of post harvest potato. *Journal of Chinese Agricultural Mechanization* 41: 115-120.
- Zhang YQ, Chu J, Leng L, Miao J (2020) Mask-Refined R-CNN: A Network for Refining Object Details in Instance Segmentation. *Sensors* 20:1010.
- Zhang ZG, Wang HY, Li YB, Yang X, Ibrahim I, Zhang ZD (2021) Design and experiment of multi-stage separation buffer potato harvester. *Transactions of the Chinese Society for Agricultural Machinery* 52:96-109.
- Zhao QL (2020) Research status and development prospect of potato harvesting machinery at home and abroad. *Agricultural Engineering* 10:7-10.
- Zhao SN, Ma YC, Gao RW, Feng Q, Chen SY, Zhang RG (2019a) Research status of potato storage technology at home and abroad. *Storage and Process* 19: 153-158.
- Zhao SX, Liu CL, Hu J, Sa L, Yu HY (2019b) Research status of harvesting machinery for domestic Potato. *Journal of Heilongjiang Bayi Agricultural University* 31:78-83.