Hindawi
Journal of Food Quality
Volume 2020, Article ID 8815896, 9 pages
https://doi.org/10.1155/2020/8815896

Research Article

Potato Quality Grading Based on Depth Imaging and Convolutional Neural Network

Qinghua Su,[1,2] Naoshi Kondo,[2] Dimas Firmanda Al Riza,[2] and Harshana Habaragamuwa[2]

[1] *Key Laboratory of Modern Measurement and Control Technology, Ministry of Education, Beijing Information Science & Technology University, Beijing, China*
[2] *Graduate School of Agriculture, Kyoto University, Kitashirakawa-Oiwakecho Sakyo-ku, Kyoto, Japan*

Correspondence should be addressed to Qinghua Su; suqinghua1985@qq.com

Received 27 July 2020; Revised 6 November 2020; Accepted 13 November 2020; Published 23 November 2020

Academic Editor: Eduardo Puértolas

As a cost-effective and nondestructive detection method, the machine vision technology has been widely applied in the detection of potato defects. Recently, the depth camera which supports range sensing has been used for potato surface defect detection, such as bumps and hollows. In this study, we developed a potato automatic grading system that uses a depth imaging system as a data collector and applies a machine learning system for potato quality grading. The depth imaging system collects 3D potato surface thickness distribution data and stores depth images for the training and validation of the machine learning system. The machine learning system, which is composed of a softmax regression model and a convolutional neural network model, can grade a potato tube into six different quality levels based on tube appearance and size. The experimental results indicate that the softmax regression model has a high accuracy in sample size detection, with a 94.4% success rate, but a low success rate in appearance classification (only 14.5% for the lowest group). The convolutional neural network model, however, achieved a high success rate not only in size classification, at 94.5%, but also in appearance classification, at 91.6%, and the overall quality grading accuracy was 86.6%. The quality grading based on the depth imaging technology shows its potential and advantages in nondestructive postharvesting research, especially for 3D surface shape-related fields.

1. Introduction

Potatoes, with over 18.9 million hectares planted globally every year, are one of the most important crops in the world [1]. After harvest, grading based on quality is important in classifying products into different levels, improving packing and other postharvest operations, and allowing the farmer to obtain higher prices. During the grading process, potatoes are separated into different homogeneous groups according to tube-specific characteristics such as shape, mass, color, and deformities. Potatoes are a difficult crop to grade in the postharvest process because of their wide diversity in shape, deformity, and mass, and the grading process thus still relies on experienced workers nearby the conveyor system [2]. Manual grading is a tedious, expensive, and time-consuming process, and it is often affected by a shortage of labor during the harvest season [3]. In addition, inconsistent sorting and grading errors often occur during the manual grading process because workers are easily influenced by the surrounding environment [4, 5].

Machine vision, as a nondestructive measurement, provides a high level of repeatability and accuracy at a low cost. Therefore, it draws modern manufactures' attention to apply the machine vision grading system in postharvesting work [6]. Previous research has successfully detected many key features relating to potato quality using different types of imaging device, such as CCD camera, hyperspectral camera, ultraviolet camera, and X-ray CT. The machine vision system is already capable of predicting potato physical size, including length, width, and mass, and can detect inner and external defects, such as green skin, sprouts, bruises, mechanical injury, black heart, and water core [7–16]. In recent years, research has begun to obtain surface data in 3D space, using methods such as a stereo vision system and a V-shaped mirror system [17, 18]. Another innovative method is to equip a range-sensing device on a machine vision system, such as a depth camera. A depth camera senses object appearance information using time of flight (TOF) and light-coding technologies [19–23], and it has already been applied in motion tracking, automatic driving, indoor 3D mapping, robot navigation, gesture control, potato 3D model rebuilding, and other areas [24–32].

However, in the past, potato classification algorithms have relied largely on the image processing technology, which cooperates tightly with hardware, such as a specific light source. Once the hardware environment changes, the entire algorithm which must be upgraded is difficult to maintain. In addition, potatoes have a variety of forms and their growth is greatly affected by the natural environment. Therefore, the classification accuracy of a fixed classification algorithm changes each year. The ideal classification algorithm should be able to enlarge its knowledge database by training with a small number of manually graded products each year to ensure classification accuracy. Machine learning thus shows its advantages here.

Machine learning uses computational models that exhibit similar characteristics to those of the neocortex, such as neural networks, for information representation. A computer could optimize a performance criterion based on example data or experience with proper programming [33]. A key problem in image data understanding, one of the important application fields for machine learning, is the discovery of effective relevant information from input data. While the performance of conventional, handcrafted features has plateaued recently, new developments in deep compositional architectures have led the performance level to improve continuously. Deep models have shown outstanding performance in many domains compared to hand-engineered feature representations [34, 35].

Many machine learning achievements have been reported and widely used, such as softmax regression and the convolutional neural network. Softmax regression, also called multinomial logistic regression, is as generalization of logistic regression to cases in which multiple classes must be classified. This model is used to predict the probabilities of the different possible outcomes of the categorically distributed dependent variable, given a set of independent variables (which may be real values, binary values, category values, and so on) [36, 37]. Softmax regression was applied as a classifier for the MNIST digit recognition task, in which the goal was to distinguish between 10 different numerical digits [38].

Convolutional neural networks (CNNs), as an excellent machine learning method, are a kind of multilayer neural network specially used for two-dimensional data (including images and videos) [39, 40]. These neutral networks represent the first truly successful deep learning method, in which many layers of the hierarchy are trained differently and successfully in a robust way [41]. In CNNs, a small part of the images are regarded as inputs to the lowest layer of the hierarchical structure and information transmits through the different layers of the network, whereby at each layer, digital filtering is applied in order to obtain salient features of the observed data [42]. In addition, CNNs also provide a certain degree of translation, scaling, and rotation invariance because the local receiving field allows processing units to access basic features, such as directional edges or corners [41]. Currently, CNNs have been applied in various areas of study related to machine learning, including face detection [43, 44], document analysis [45, 46], speech recognition [47, 48], medical examination [49, 50], and precision agriculture [51–53].

Since potato quality classification based on the depth imaging technology and machine learning has merely been reported, the overall objective of this study is to develop a system that automatically grades potato tubers of diverse size and appearance based on machine vision, depth image processing, and machine learning technology. This grading system captures sample depth images by a depth camera system, develops a potato depth image processing algorithm, builds the machine learning models, and evaluates the potato quality level automatically. In addition, the results of two different machine learning models will be compared and analyzed to determine whether machine learning is suitable for potato quality classification. Overall, this method requires the development of fast algorithms to analyze tube appearance and predict the sample mass with high accuracy but less processing time and resource consumption.

2. Materials and Methods

2.1. Potato Samples. In total, 296 potatoes (Jizhangshu no. 8) were purchased from Beijing Qinghe Agricultural Market. By randomly choosing potatoes with diverse masses and appearances, the reliability of our experiment could be ensured. All potatoes were cleaned and washed individually to remove all clay and dirt, and they were then separated into normal (with spherical or ellipsoidal shape) and abnormal (including bumpy, hollow, mechanical injury, and sprout) groups by experienced farmers. Next, according to the Chinese Official Grades and Specifications of Potatoes [54], the potatoes were divided into three categories based on mass: small (<100 g), medium (∈[100 g, 300 g]), and big (≥ 300 g). Six classes were thus used to grade sample quality: Abnormal Big (AB), Abnormal Medium (AM), Abnormal Small (AS), Normal Big (NB), Normal Medium (NM), and Normal Small (NS).

2.2. Depth Machine Vision System. The machine vision system design was similar to designs used in previous research [32], including six main parts: one depth camera system (Primesense Carmine 1.09), two fluorescent lamps (Philips, 18W, 6400 K), one black box, one sample holder, and one PC with Intel i5 CPU, Windows 10 Operating System, and 16G RAM, as shown in Figure 1.

2.3. Camera System Setting and Depth Image Preprocessing. The Primesense camera senses the range using the "light-coding" technology, and the camera resolution was set to 640 $*$ 480 pixels, while the frequency was 30 frames per second [55]. The camera was installed on a black box with 60 cm above the box bottom, and its view field was 45° vertically and 57.5° horizontally.

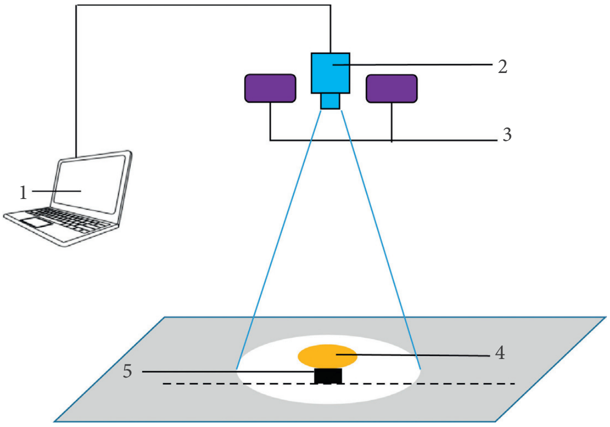

Figure 1: Machine vision system: 1, computer; 2, camera system; 3, light source; 4, potato; 5, sample holder.

The potato and its holder were randomly placed under the camera within the view field, after which the color image and depth image were captured with one shot, as shown in Figures 2(a) and 2(b). Example images for samples with normal and abnormal features (including bumps, hollows, mechanical injuries, and sprouts) are shown in Figure 2. The gray level on each pixel indicated the surface thickness (the distance above the ground) on the potato surface area. As seen in Figure 2(a), some deformities were difficult to recognize in color images because of very subtle color differences in the deformity area and normal surface area. Even when the defect color difference was clear, such as in the case of a mechanical injury, the depth of the injury area still could not be sensed. However, as shown in Figure 2(b), the gray level in the depth image gradually decreased from center peak to boundary in the normal potato image, whereas the gray level increased or decreased differently around a deformity area in abnormal potato depth images. Image enhancement resulted in a group of images with a clearer thickness distribution, as seen in Figure 2(c). This feature indicates that depth images display variance in object surface information in ways that color images cannot [29].

As previous research indicated [32], images from a depth camera could not be directly used. The original depth image recorded the range from potato surface to camera. To calculate the potato surface thickness in the raw image, a difference between the background depth image and the original depth image must be calculated pixel by pixel, and inside potato area, an additional holder height must be subtracted.

The raw image included much noise and many holes, which were caused by the absence of a reflected beam. Therefore, each raw potato image had to be captured three times and an average operation should be conducted pixel by pixel to fill holes. After that, the erosion, dilation, Gauss smoothing, and big noise clearing operations, all part of the depth image preprocessing module, were applied individually to create one valid potato depth image, as shown in Figure 3.

A total of 7084 depth images were captured in this study, and for each potato, depth data were extracted into one image with a resolution of 200 $*$ 200 pixels using OpenCV (http://opencv.org) to reduce the computing time in a convolutional network.

2.4. Machine Learning Models. Two machine learning models were developed: the softmax regression (SR) model and a convolutional neural network (CNN) model. Both were created by the deep learning package Keras, which runs the Tensorflow machine learning package in the background. We trained both models using an Adam optimizer for stochastic optimization, and the initial learning rate and stopping were set to 0.001 and 2, respectively. The loss function used for optimization was a categorical cross-entropy function.

The training image dataset classes were defined as “AB,” “AM,” “AS,” “NB,” “NM,” and “NS” based on each potato manual grading label. In order to improve the performance of the network, each image was randomly augmented in each epoch: random yes or no horizontal and vertical flip, random rotation 0–90°, and random horizontal and vertical shifts. For model training, 500 epochs were included and 5691 images were randomly chosen as a training dataset. Model prediction accuracy and loss in the validation process were performance indices.

2.4.1. Softmax Regression Model. An SR model with two layers was developed for potato classification, including a fully connected layer and a classification layer, as shown in Figure 4.

Since the input depth image was two dimensional, as x [200][200], it had to be converted into a one-dimensional shape x[40000] by a reshape operation to fit the model data input format. Evidence, as a key output for correct image class determination from the fully connected layer, was calculated by image-weighted summation, as shown in equation (1). $W_{i,j}$ is a weight element in weight matrix W; b_{i} is a bias element in bias array B for potato type i; i indicates the potato type (0-AB, 1-AM, 2-AS, 3-NB, 4-NM, and 5-NS); and j shows the pixel index of input image x for the pixel summation. A bias array B[b_0, b_1,. . . b$_5$] and a weight matrix W were created after model training. In addition, a rectified linear (ReLU) activation and a dropout layer ($p = 0.5$) were added after the fully connected layer to avoid the problems of the gradient blowing up and of overfitting, respectively.

$$\text{Evidence}_i = \sum_j W_{i,j} x_j + b_i. \tag{1}$$

The classification layer included a softmax activation function, which converts the linear function output into a six-class probability distribution, as shown in equation (2). Then, the probability array Y[y_0, y_1, ..., y_5] from the classification layer indicated the correct class for input images x, as shown in equation (3):

$$\text{softmax (evidence)}_i = \frac{\exp\left(\text{evidence}_i\right)}{\sum_j \exp\left(\text{evidence}_j\right)}, \tag{2}$$

$$Y = \text{softmax (evidence)}. \tag{3}$$

2.4.2. CNN Model. Our CNN structure is shown in Figure 5. This network has five layers of learned weights: three convolutional layers, one fully connected layer, and one

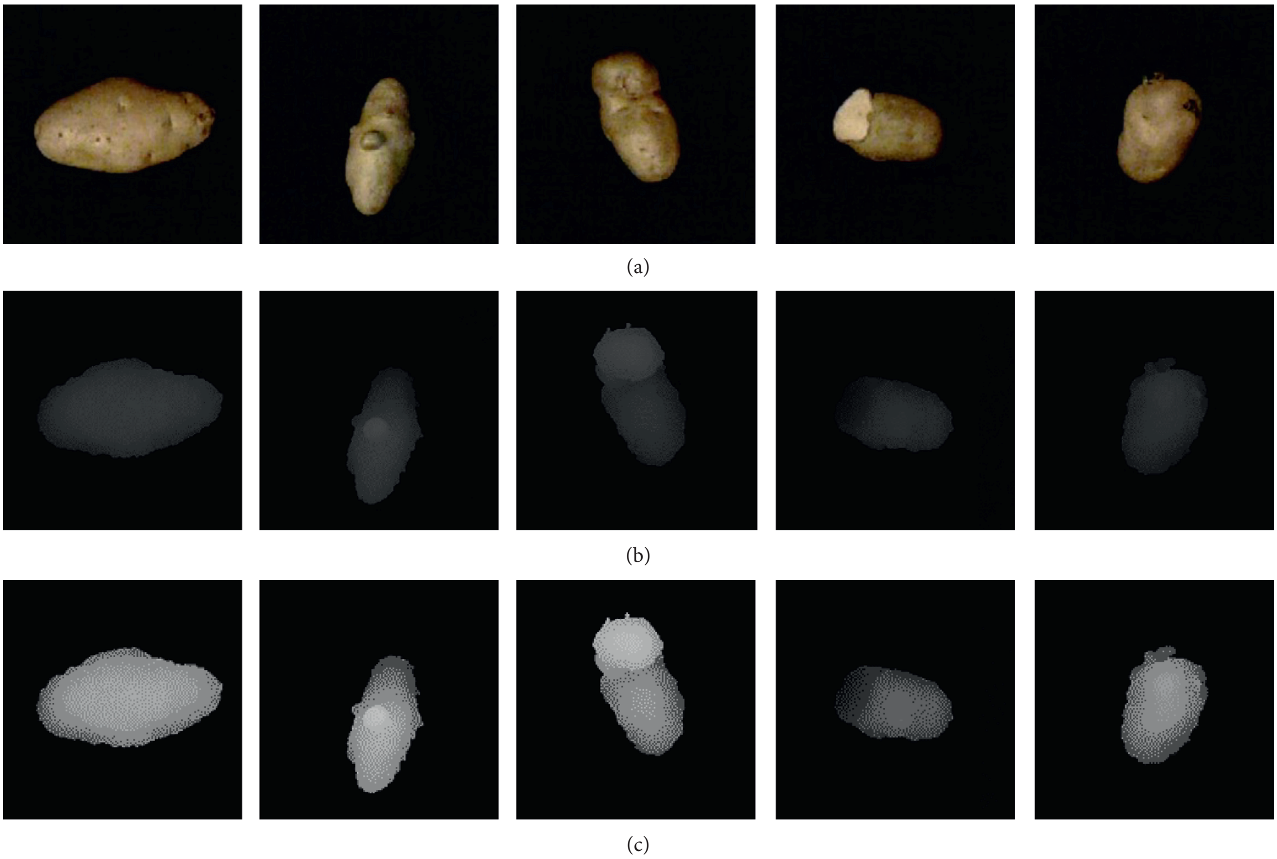

Figure 2: Image comparison for potato features. (a) Color images of normal, bump, hollow, machinery injury, and sprout potatoes. (b) Depth images of normal, bump, hollow, machinery injury, and sprout potatoes. (c) Enhanced depth images of normal, bump, hollow, machinery injury, and sprout potatoes.

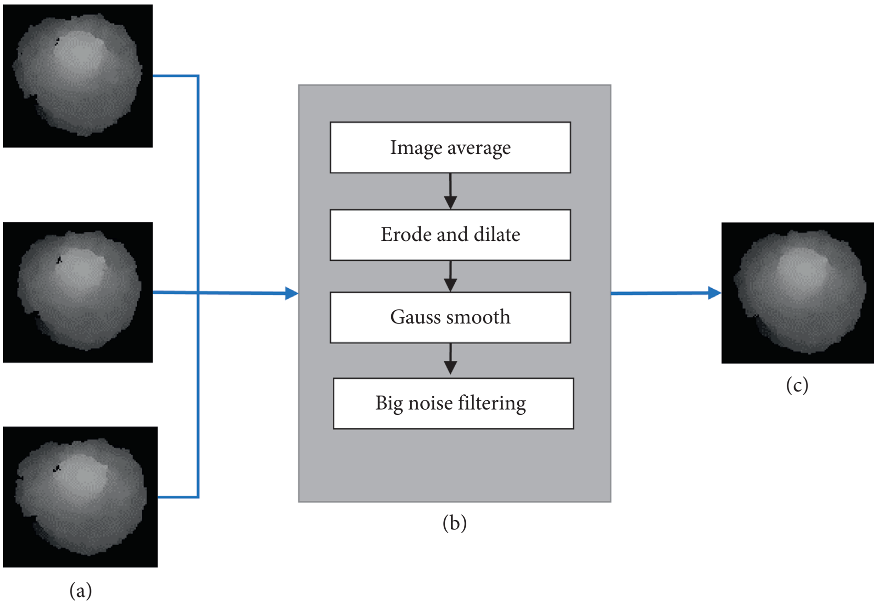

Figure 3: Depth image preprocessing flowchart. (a) Three enhanced potato surface depth images. (b) Potato image processing module. (c) Enhanced potato surface depth image.

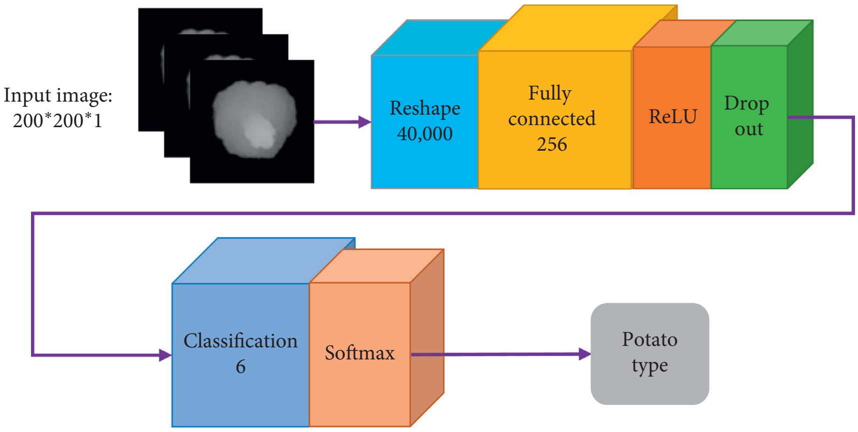

Figure 4: Softmax regression model architecture.

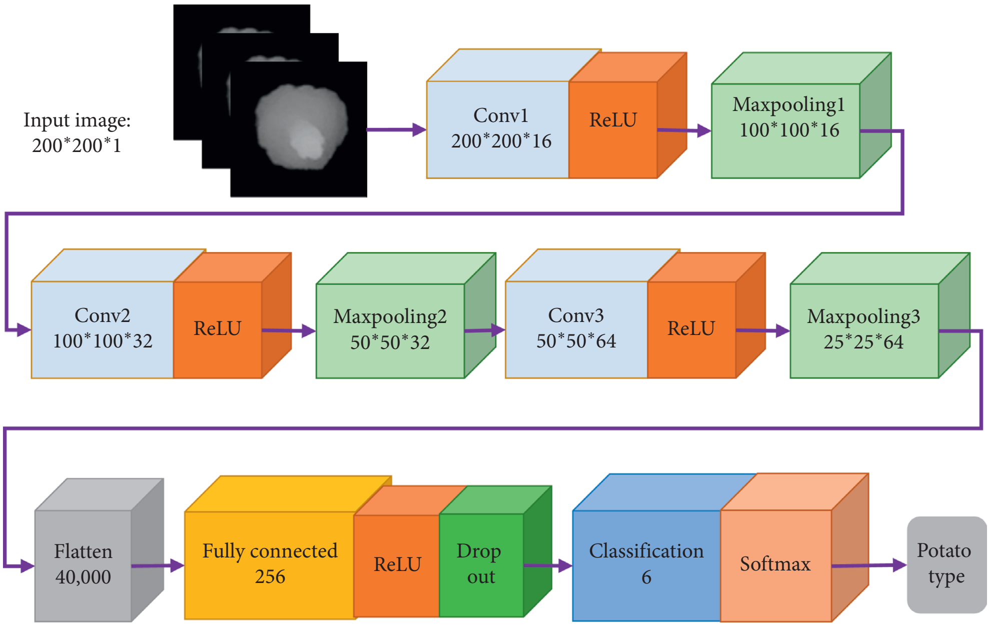

Figure 5: CNN model architecture.

classification layer, with approximately 10 million trainable parameters in total. The CNN model was an improved SR model with three added convolutional layers and a flatten layer used to replace the reshape layer. ReLU activation followed each convolutional layer. Max pooling was performed with a kernel size of $2 * 2$ strides to resize input data into half its size. After the final convolutional layer, the network was flattened to one dimension. To avoid overfitting this model, we included a dropout ($p = 0.5$) in the first fully connected layer and the last stage of the convolutional layers.

3. Results and Discussion

Another group of 1,393 images was used for the validation. Running the validation dataset on two models took 7 and 56 seconds, respectively, for the SR and CNN models, and Figures 6 and 7 show the learning curves. As more epochs were processed, it was obvious that the loss for the CNN models for both training and validation was gradually decreased, whereas the prediction accuracy increased until it stabilized at 500 epochs. However, it was a little different for the SR model, while the trends for both loss and accuracy were the same as in the CNN model in the first 100 epochs; after that, the model was almost stable. Ultimately, the validation accuracy and validation loss of the SR model were 67.2% and 0.777, respectively, after 500 epochs' training, while these were 86.6% and 0.304, respectively, for the CNN model.

The confusion matrixes for both models are shown in Figures 8 and 9, in which the classifications in the network were defined numerically as follows: 0-AB, 1-AM, 2-AS, 3-NB, 4-NM, and 5-NS. It was clear that the prediction accuracy of the SR model was lower than that of the CNN model because only one fully connected layer was used in the SR model and the potato appearance gradient change continuity was lost when an image with two-dimensional

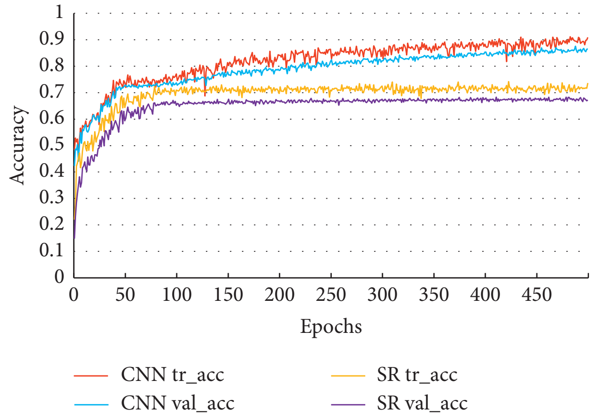

Figure 6: Accuracy curves.

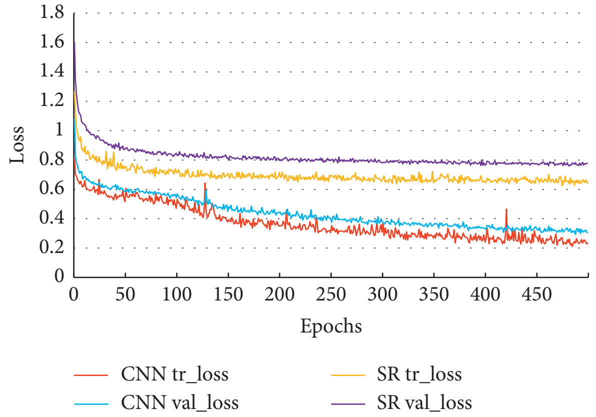

Figure 7: Loss curves.

True labels \ Predicted labels	AB	AM	AS	NB	NM	NS
AB	162	25	0	10	0	0
AM	7	480	7	0	37	11
AS	0	6	236	0	16	43
NB	43	0	0	10	0	0
NM	0	138	4	0	22	2
NS	0	0	108	0	0	26

Figure 8: Confusion matrix of the SR model.

True labels \ Predicted labels	AB	AM	AS	NB	NM	NS
AB	178	16	0	3	0	0
AM	21	451	13	0	57	0
AS	0	6	262	0	0	33
NB	3	1	0	45	4	0
NM	0	5	0	6	152	3
NS	0	0	15	0	6	113

Figure 9: Confusion matrix of the CNN model.

data was reshaped to a one-dimensional array. As a result, the SR model was sensitive for sample size detection (94.4% samples could be grouped into the appropriate size group), while it had low sensitivity for appearance recognition. The success rates for normal potato appearance classification were only 18.9%, 14.5%, and 19.4% for NB, NM, and NS, respectively, because a sample would be grouped into a higher priority class when it had the same prediction probability for different appearance classes (abnormal appearance labels were 0, 1, and 2 and had higher priority, whereas normal appearance labels were 3, 4, and 5 and had lower priority). For instance, a potato (manually marked as NB) that has a 36% chance of being classified as either AB or NB by the SR model will be classified as AB because AB has a priority 0, which is higher than the priority of 3 for the NB class.

With the addition of convolutional layers, the CNN model could process the two-dimensional depth image, extract potato features, and achieve feature mapping. The test result in Figure 9 indicates that the CNN model not only recognized sample appearance and size but also obtained a high success rate for the six-class classifications. In total, 94.5% of potatoes were grouped into the right size classes, which is slightly higher than previous research [16], which has classified the tube size based on the calculated boundaries from three color images. In addition, deformity in appearance was detected in 91.6% of samples, while previous research using image processing has achieved only 88% detection [32]. Overall, 86.6% potatoes were classified according the correct quality level in terms of both appearance and size features. This is slightly lower than previous results, which achieved an 89% success rate [56], but this could be improved by extending the training dataset in the future.

Several hardware-related problems might explain the CNN model grading errors: unexpected noise on the image, missing edge area, and undetected small bumps by sprouts. Figure 10(a) shows a normal potato with unexpected noise on the edge area. A noise area appearing on the bottom right

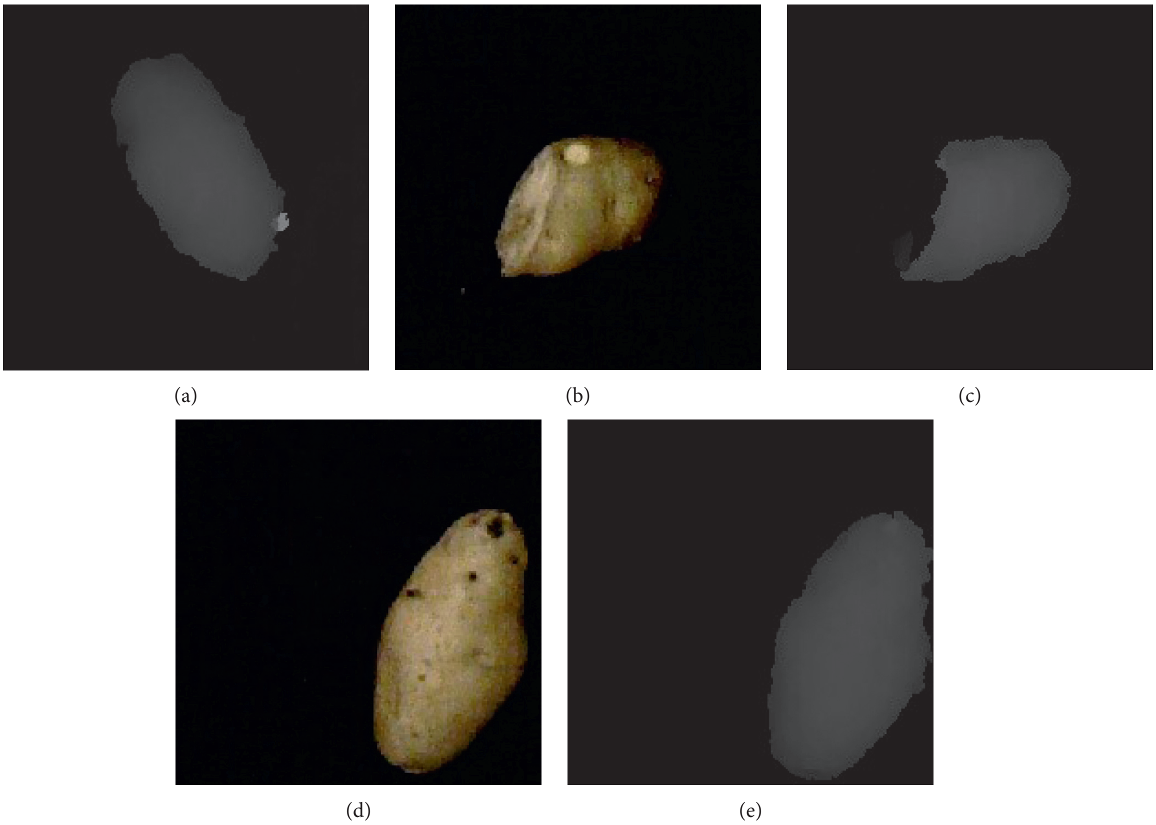

Figure 10: Mistake grading samples by the CNN model. (a) Unexpected noise. (b) Color image of machinery injury potato. (c) Depth image of machinery injury potato. (d) Color image of sprout potato. (e) Depth image of sprout potato.

edge area increased the local average gray level sharply, and our model thus classified this normal potato as an abnormal one. Figures 10(b) and 10(c) illustrate one potato with a machinery injury that lost some edge area, with the surface gradient changing sharply; hence, this AM sample was graded to the AS class. This edge loss problem was also reported in previous studies [29] and was caused by beam loss with too large incident angle. The potato in Figures 10(d) and 10(e) was manually graded as AB due to some sprouts on the surface; however, these small sprouts (less than around 5 mm) could not be detected as small bumps on the depth image, whereas they were obviously darker in the color image.

4. Conclusion

We propose a new potato quality grading system based on a machine vision system and machine learning models. Depth images, which include 3D potato appearance data, were captured and used for quality grading by a machine vision system and machine learning models. The results indicate that a machine learning model with a softmax network has a high sensitivity for sample size detection, with 94.4% accuracy, but at a low rate of appearance classification. The machine learning model with a convolutional neural network achieved a high success rate for size and appearance classification, at 94.5% and 91.6%, respectively, and abnormal defects were successfully detected, and potatoes were correctly grouped according to size and quality level in 86.6% of samples. Therefore, the advantages of this potato grading system are summarized as follows: (1) it is a cost-effective solution. Currently, many manufacturers sell depth camera products and the price has decreased greatly. In addition, the depth camera can be an independent device or can be integrated with a color camera based on the budget and experiment requirements. (2) The system is less affected by ambient light. The depth camera includes a near-infrared light source and can work stably around other lights, such as LEDs and fluorescent lamps. (3) It nondestructively acquires 3D appearance data. This system calculates sample 3D surface shape information based on the light-coding technology and is harmless for the tube surface. (4) It features automatic classification based on human experience. This system is developed and trained based on manual classification experience, and therefore, the classification accuracy can be promoted in the future while extending the training dataset.

This system can capture potato surface shape information such as bumps, hollows, and machinery injury, but it is not sensitive enough to detect small sprouts on the surface; however, these defects are clear in the color images. Therefore, a 4D model combining color and 3D shape information for the nondestructive postharvesting of potatoes may be a potential solution for small sprout detection, and this method is also expected to increase the accuracy of deformity prediction.

Data Availability

All the data presented and analyzed in the manuscript were obtained from laboratory tests at Beijing Information Science and Technology University in Beijing, China. All the laboratory testing data were presented in the figures and tables in the manuscript. We will be very pleased to share all the raw data. Requests for access to these data should be made to suqinghua1985@qq.com.

Conflicts of Interest

The authors have no conflicts of interest.

Acknowledgments

The authors wish to appreciate the great support from Beijing Information Science and Technology University with Qin Xin Talents Cultivation Program (QXTCPA201903 and QXTCPB201901) and School Research Fund (2025041).

References

[1] FAO, "Food and Agriculture Organization Statistics," FAO, New York, NY, USA, 2004.

[2] G. Elmasry, S. Cubero, E. Moltó, and J. Blasco, "In-line sorting of irregular potatoes by using automated computer-based machine vision system," *Journal of Food Engineering*, vol. 112, no. 1-2, pp. 60–68, 2012.

[3] D. S. Narvankar, S. K. Jha, and A. Singh, "Development of rotating screen grader for selected orchard crops," *Journal of Agricultural Engineering*, vol. 42, no. 4, pp. 60–64, 2005.

[4] N. Razmjooy, B. S. Mousavi, and F. Soleymani, "A real-time mathematical computer method for potato inspection using machine vision," *Computers & Mathematics with Applications*, vol. 63, no. 1, pp. 268–279, 2012.

[5] L. Zhou, V. Chalana, and Y. Kim, "PC-based machine vision system for real-time computer-aided potato inspection," *International Journal of Imaging Systems and Technology*, vol. 9, no. 6, pp. 423–433, 1998.

[6] C. Sylla, "Experimental investigation of human and machine-vision arrangements in inspection tasks," *Control Engineering Practice*, vol. 10, no. 3, pp. 347–361, 2002.

[7] A. Al-Mallahi, T. Kataoka, and H. Okamoto, "Discrimination between potato tubers and clods by detecting the significant wavebands," *Biosystems Engineering*, vol. 100, no. 3, pp. 329–337, 2008.

[8] A. Al-Mallahi, T. Kataoka, H. Okamoto, and Y. Shibata, "An image processing algorithm for detecting in-line potato tubers without singulation," *Computers and Electronics in Agriculture*, vol. 70, no. 1, pp. 239–244, 2010.

[9] A. M. Rady and D. E. Guyer, "Rapid and/or nondestructive quality evaluation methods for potatoes: a review," *Computers and Electronics in Agriculture*, vol. 117, pp. 31–48, 2015.

[10] H. Wang, J. Xiong, Z. Li, J. Deng, and X. Zou, "Potato grading method of weight and shape based on imaging characteristics parameters in machine vision system," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 32, no. 8, pp. 272–277, 2016.

[11] A. Dacal-Nieto, E. Vázquez-Fernández, A. Formella, and F. Martin, "A genetic algorithm approach for feature selection in potatoes classification by computer vision," *Industrial Electronics*, vol. 48, pp. 1955–1960, 2009.

[12] Y. Kong, X. Gao, H. Li et al., "Potato grading method of mass and shapes based on machine vision," *Nongye Gongcheng Xuebao/transactions of the Chinese Society of Agricultural Engineering*, vol. 28, no. 17, pp. 143–148, 2012.

[13] N. Razmjooy, V. Vieira Estrela, and H. J. Loschi, "A survey of potatoes image segmentation based on machine vision," in *Applications of Image Processing and Soft Computing Systems in Agriculture*, pp. 1–38, IGI Global, Hershey, PA, USA, 2019.

[14] N. Razmjooy, *Automatic Sorting of Potatoes Using Soft Computing*, LAP LAMBERT Academic Publishing, 2018.

[15] P. Moallem, N. Razmjooy, and B. S. Mousavi, "Robust potato color image segmentation using adaptive fuzzy inference system," *Iranian Journal of Fuzzy Systems*, vol. 11, no. 6, pp. 47–65, 2014.

[16] P. Moallem, N. Razmjooy, and M. Ashourian, "Computer vision-based potato defect detection using neural networks and support vector machine," *International Journal of Robotics and Automation*, vol. 28, pp. 1–9, 2013.

[17] R. Runge, *Mobile 3D Computer Vision: Introducing a Portable System for Potato Size grading*, Radbound University, Nijmegen, Netherlands, 2014.

[18] Z. Zhou, Y. Huang, X. Li, D. Wen, C. Wang, and H. Tao, "Automatic detecting and grading method of potatoes based on machine vision," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 28, no. 7, pp. 178–183, 2012.

[19] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE Journal of Quantum Electronics*, vol. 37, no. 3, pp. 390–397, 2001.

[20] T. Leyvand, C. Meekhof, Y. C. Yi-Chen Wei, J. Jian Sun, and B. Baining Guo, "Kinect identity: technology and experience," *Computer*, vol. 44, no. 4, pp. 94–96, 2011.

[21] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE Multimedia*, vol. 19, no. 2, pp. 4–10, 2012.

[22] Depth Image, http://blog.csdn.net/sdau20104555/article/details/40740683.

[23] Range imaging, https://en.wikipedia.org/wiki/Range_imaging.

[24] Y. Mori, N. Fukushima, T. Yendo, T. Fujii, and M. Tanimoto, "View generation with 3d warping using depth information for ftv," *Signal Processing: Image Communication*, vol. 24, no. 1-2, pp. 65–72, 2009.

[25] M. Ye, X. Wang, R. Yang, L. Ren, and M. Pollefeys, "Accurate 3D pose estimation from a single depth image. International Conference on Computer Vision," *IEEE Computer Society*, vol. 24, pp. 731–738, 2011.

[26] H. Du, P. Henry, X. Ren et al., "Interactive 3D modeling of indoor environments with a consumer depth camera," in *Proceedings of the 2011: Ubiquitous Computing, International Conference, UBICOMP 2011*, pp. 75–84, Beijing, China, September 2011.

[27] L. Xia, C. C. Chen, and J. K. Aggarwal, "Human detection using depth information by kinect," *Applied Physics Letters*, vol. 85, no. 22, pp. 5418–5420, 2011.

[28] Z. Ren, J. Meng, J. Yuan, and Z. Zhang, "Robust hand gesture recognition with kinect sensor," in *Proceedings of the International Conference on Multimedia 2011*, pp. 759-760, Scottsdale, AZ, USA, December 2011.

[29] Q. Su, N. Kondo, M. Li, H. Sun, and D. F. Al Riza, "Potato feature prediction based on machine vision and 3D model rebuilding," *Computers and Electronics in Agriculture*, vol. 137, pp. 41–51, 2017.

[30] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "Rgb-d mapping: using kinect-style depth cameras for dense 3d

modeling of indoor environments," *The International Journal of Robotics Research*, vol. 31, no. 5, pp. 647–663, 2012.

[31] J. Pajarinen and V. Kyrki, "Robotic manipulation in object composition space," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 5336, pp. 11–16, 2014.

[32] Q. Su, N. Kondo, M. Li, H. Sun, D. F. Al Riza, and H. Habaragamuwa, "Potato quality grading based on machine vision and 3D shape analysis," *Computers and Electronics in Agriculture*, vol. 152, pp. 261–268, 2018.

[33] E. Alpaydin, *Introduction to Machine Learning (Adaptive Computation and Machine Learning). Introduction to Machine Learning*, MIT Press, Cambridge, MA, USA, 2004.

[34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks,"-vol. 25, pp. 1097–1105, in *Proceedings of the International Conference on Neural Information Processing Systems*, vol. 25, pp. 1097–1105, Curran Associates Inc., Lake Tahoe, CA, USA, 2012.

[35] Y. Jia, E. Shelhamer, J. Donahue et al., "Caffe: convolutional architecture for fast feature embedding," in *Proceedings of the ACM International Conference on Multimedia. ACM.*, pp. 675–678, New York; NY, USA, November, 2014.

[36] J. Engel, "Polytomous logistic regression," *Statistica Neerlandica*, vol. 42, no. 4, p. 233, 1988.

[37] W. H. Greene, *Econometric Analysis*, Pearson Education, vol. 803, 806 pages, London, UK, 2012.

[38] Softmax Regression, http://deeplearning.stanford.edu/wiki/index.php/Softmax_Regression.

[39] F. J. Huang and Y. Lecun, *Large-scale learning with svm and convolutional nets for generic object categorization*, vol. 1, pp. 284–291, 2006.

[40] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[41] I. Arel, D. C. Rose, and T. P. Karnowski, "Deep machine learning - a new frontier in artificial intelligence research [research frontier]," *IEEE Computational Intelligence Magazine*, vol. 5, no. 4, pp. 13–18, 2010.

[42] F. Cady, "Machine learning overview," *The Data Science Handbook*, John Wiley & Sons, Hoboken, NJ, USA, 2017.

[43] F. H. C. Tivive and A. Bouzerdoum, "A new class of convolutional neural networks (SICoNNets) and their application of face detection. International Joint Conference on Neural Networks," *IEEE*, vol. 3, pp. 2157–2162, 2003.

[44] Y. N. Chen, C. C. Han, C. T. Wang, B. S. Jeng, and K. C. Fan, "The application of a convolution neural network on face and license plate detection. International conference on pattern recognition," *IEEE*, vol. 3, pp. 552–555, 2006.

[45] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Convolutional neural network committees for handwritten character classification," in *Proceedings of the 2011 International Conference on Document Analysis and Recognition*, pp. 1135–1139, Beijing, China, 2011.

[46] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document Analysis.International conference on document analysis and recognition," *IEEE Computer Society*, vol. 958, 2003.

[47] O. Abdelhamid, D. Li, and Y. Dong, "Exploring convolutional neural network structures and optimization techniques for speech recognition," in *Proceedings of the INTERSPEECH 2013*, Lyon, France, 2013.

[48] S. Sukittanon, A. C. Surendran, J. C. Platt, and C. J. C. Burges, "Convolutional networks for speech detection," *Convolutional Networks for Speech Detection*, vol. 22, no. 10, 2004.

[49] H. Pratt, F. Coenen, D. M. Broadbent, S. P. Harding, and Y. Zheng, "Convolutional neural networks for diabetic retinopathy," *Procedia Computer Science*, vol. 90, pp. 200–205, 2016.

[50] J. Antony, K. Mcguinness, K. Moran, and N. E. O'Connor, "Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks," *Machine Learning and Data Mining in Pattern Recognition*, vol. 15, 2017.

[51] C. Yao, Y. Zhang, Y. Zhang, and H. Liu, "Application of convolutional neural network in classification of high resolution agricultural remote sensing images," *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 12, pp. 989–992, 2017.

[52] Firdaus, Y. Arkeman, A. Buono, and I. Hermadi, "Satellite image processing for precision agriculture and agroindustry using convolutional neural network and genetic algorithm," *Earth and Environmental*, vol. 54, Article ID 012102, 2017.

[53] H. S. Abdullahi, R. E. Sheriff, and F. Mahieddine, "Advances of image processing in precision agriculture: using deep learning convolution neural network for soil nutrient classification," *World Academy of Science, Engineering and Technology, International Science Index, Agricultural and Biosystems Engineering*, vol. 4, no. 7, p. 1436, 2017.

[54] NY/T 1066-2006, *Chinese Official Grades and Specifications of Potatoes*, 2006.

[55] PrimeSense, http://en.wikipedia.org/wiki/PrimeSense.

[56] B. Michael, D. Tom, C. Grzegorz, S. Graeme, and H. Glyn, "Visual detection of blemishes in potatoes using minimalist boosted classifiers," *Journal of Food Engineering*, vol. 98, no. 3, pp. 339–346, 2010.