Research Article

Dough-Stage Maize (*Zea mays* L.) Ear Recognition Based on Multiscale Hierarchical Features and Multifeature Fusion

Honglei Jia,¹ Minghao Qu,² Gang Wang ,^{2,3} Michael J. Walsh,⁴ Jurong Yao,⁵ Hui Guo,¹ and Huili Liu²

¹Key Laboratory of Bionic Engineering, Ministry of Education, Jilin University, Changchun 130022, China

²College of Biological and Agricultural Engineering, Jilin University, Changchun 130022, China

³Department of Geography and Regional Research, University of Vienna, Vienna 1090, Austria

⁴Department of Biological and Environmental Engineering, Cornell University, Ithaca, New York 14853, USA

⁵First Administrative Department of People's Canal, Dujiangyan Sichuan Province, Chengdu 611000, China

Correspondence should be addressed to Gang Wang; gw1988@jlu.edu.cn

Received 7 January 2020; Revised 24 April 2020; Accepted 5 May 2020; Published 18 May 2020

Academic Editor: Thomas Hanne

Copyright © 2020 Honglei Jia et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Crop-related object recognition is of great importance in realizing intelligent agricultural machinery. Maize (*Zea mays* L.) ear recognition could be a representative of crop-related object recognition, which is a critical technological premise for realizing automatic maize ear picking and maize yield prediction. In order to recognize maize ears in dough stage, this study combined deep learning and image processing, which have advantages of feature extraction and hardware flexibility, respectively. *LabelImage* was applied to mark and label maize plants, based on the deep learning framework *TensorFlow*, and this study developed multiscale hierarchical feature extraction together with quadruple-expanded convolutional kernels. To recognize the whole maize plant, 1250 images were acquired for training the recognition model, and its performance in a test set showed that the recognition accuracy was 99.47%. Subsequently, multifeatures of maize ear were determined, and the optimum binary threshold was obtained by fitting Gaussian distribution in the subblock image. Consequently, the maize ear was recognized by morphological process which was conducted by *Python* and *OpenCV*. Experiment was conducted in August 2018, and 10800 images were acquired for testing this algorithm. Experimental results showed that the average recognition accuracy was 97.02% and time consumption was 0.39 s for each image, which could meet a forward speed of 4.61 km/h for combine harvesters.

1. Introduction

Intelligent equipment with machine vision is a future trend of agricultural machinery [1]. With regard to intelligent agricultural machinery, harvesting robots together with yield predictive function could be representatives [2]. In maize fields, maize ear recognition is a premise in favoring robots to complete picking movements, and maize ear number is a critical parameter to realize yield prediction. Yu et al. [3] designed a recognition method to determine the flowering stage of maize, they combined support vector machine (SVM), low-rank matrix recovery technology, and color gradient histogram, and they stated that their method can meet the demand for practical observation. Zhang [4]

conducted the maize ear recognition based on deep learning, and their research purpose was to distinguish different maize cultivars through maize ear images; the experimental results showed that the distinguish rate could reach 94.6%. Zhang et al. [5] conducted research based on machine vision, which aimed at distinguishing abnormal maize ears. Their research applied the SVM and back propagation neural network (BPNN), and the experimental results showed that the accuracies were 96, 93.3, and 90% for mildew ears, worm ears, and mechanical damaged ears, respectively. Although the above research studies got their expected results, their research studies were completed in the laboratory. A lot of development should be completed before their research studies could be used for moving maize harvesters. Similar to

maize ear recognition, researchers applied substantial methods to realize maize plant recognition, such as machine vision, deep learning, and travel switch, and experimental results showed that the deep learning method is more accurate, while the travel switch method is more cost-effective [6, 7]. Maize has many growing stages, dough stage is an important stage from which maize can be used as edible products. Thus, recognizing maize ears in dough stage has practical significance. In terms of recognizing maize ears, machine vision technology will play a paramount role, and deep learning and image processing are key pathways in realizing machine vision [8].

Currently, morphological differences are discriminatory bases of machine vision technology, which needs cooperation of computer hardware and software [9]. In order to detect aiming objects, separating foreground and background is a necessary procedure. Colors, outlines, shapes, veins, and distributions are usually applied for foreground and background separation [10, 11]. Edge detections and character extractions are commonly used for aiming object recognition; in addition, the SVM and neural network (NN) perform excellently in object recognition [12]. Conventional machine vision technology needs several characters to extract aiming objects, but it does not show satisfied detection results in complicated background images. On the contrary, deep learning technology does not need to depict objective features. More conveniently, deep learning can map specified features to target space directly [13]. Deep learning technology can simulate the study process of human brains, and it can recognize features and labels of the prepared training dataset, then proceeding regression analysis or classification automatically [14]. Consequently, regression models or classifiers could be obtained. Before image training starts, acquired images will be separated into the training set and testing set according to a specific ratio, and the artificial neural network (ANN) and convolutional neural network (CNN) are commonly adopted in the deep learning field [15]. One obvious flaw is the recognition model obtained by deep learning usually shows perfect accuracy in the prepared testing dataset, but it decreases dramatically when applied to a brandnew scenario; beyond that, deep learning technology has a relative lower inter-pretability and generalization [16].

Compared with standard mechanical factories, agricultural fields are much more complicated, it is unrealistic to anticipate crop fields having factory-level environment, and crop-related objects will not have factory-level morphological features, either. So, auxiliaries were usually used in object recognition. Van Henten et al. [17] detected cucumber vines depending on recognizing vine-connected wires, and then they designed a robot which could recognize and locate cucumber vines; at last, this robot removed useless leaves from cucumber seedlings. Bac et al. [18] tied special auxiliary ropes to pepper vines in order to recognize pepper stems and locate them. Since auxiliary ropes could be detected easily, pepper vines could be recognized and located quickly and accurately. As for a huge maize field, it is hard to establish auxiliary marks owing to labor and financial reasons, but maybe the silk could act as a unique mark for maize

ear recognition [19]. Although there existed some research studies about orchard fruit detection and got relative satisfied results [15, 20], even there occurred some research studies relevant to fruit-picking robots [21, 22], maize field is a totally different scenario, such as maize leaves which can shelter ears sometimes, and maize ears can grow at any direction and any height. It is no doubt that maize ear recognition is very important for the development of harvesting machinery and yield prediction, but there still lacks sufficient research relevant to maize ear recognition, and even there lack optimal methods to realize maize ear recognition.

Concurrently, in order to reduce algorithm dependence on hardware performance and decrease time consumption in image processing, researchers have been working on developing simpler, faster, and more accurate algorithms [23]. Although deep learning shows excellent classification in many aspects, Athanasios et al. [24] indicated that agricultural practices maybe exceptions, and when applying deep learning in crop-related object recognition, specified-component recognition usually gets lower accuracy compared with whole crop plant recognition. Therefore, in order to remedy the deficiency of merely using image processing or merely using deep learning, their fusion has been studied by many researchers. Liu et al. [7] conducted a study to recognize the maize stem; firstly, they used the deep learning method to obtain the maize seedling plant; subsequently, they got the maize seedling stem within the plant area.

In order to provide a technical precondition for maize ear picking robots or yield prediction, this study was going to recognize maize ears in agricultural field. It is inevitable that some maize leaves shelter ears partially during dough stage, so it is a common phenomenon that only maize ears' main body or silk (even parts of them) exposed to lens; thus, this study would recognize maize ears' main body or silk separately. Owing to deep learning does not need to depict detail target features, image processing methodology does not rely on high-speed computer hardware, and this study would take both advantages to recognize maize ears. Firstly, one deep learning algorithm based on the deep learning network of VGG 16 would be developed to recognize whole maize plants, and then within the already known area, one image processing algorithm would be developed to recognize maize ears. Finally, the two algorithms would be combined to have a field experiment.

2. Materials and Methods

2.1. Image Acquisition and Labelling. In order to provide training images for deep learning, one charge-coupled device camera (Shanxi Vision Manufacturing Technology Co., Ltd., Model MV-EM200) and one fixed focus lens (Shanxi Vision Manufacturing Technology Co., Ltd., Model AFT-2514MP, and its focal distance is $2/3''$) were applied for image acquisition. Maize was in dough stage at that time. The distance from lens to central maize row was ~ 30 cm, and the main optical axis was set to the average height of maize ear nodes. Lens orientation was perpendicular to maize rows. A schematic diagram of image acquisition is shown in Figure 1.

It is well known that maize ears can grow in any direction around maize stalks, and Figure 2 shows four particular positions.

There are many leaves in maize plants during the dough stage, so it is a common phenomenon that the maize ear is sheltered by maize leaves, as shown in Figure 3.

These acquired images were 1280×800 pixels, and their formats were “.jpg.” Deep learning framework *Tensorflow* (Version 1.11.0) was adopted in this step. The hardware requirement should be high-level relatively in the stage of deep learning, and general information about the hardware and software was as follows: the memory is 16 GB, processor is Intel® Core™ i7-7 700 KCPU@ 4.00 GHz $\times$ 8, and digital image processor (GPU) is NVIDIA GTX 2080Ti. *Python* (Gudio van Rossum, Version 3.6.5) incorporated with *OpenCV* (computer vision repository, Version 3.4.2) was used as a programming language to complete image morphological process.

The outdoor study was initiated in Lishu County, Jilin Province, China (43.31°N, 124.62°E), on August 15th, 2018. The maize's inter-row and intrarow distances were 65 and 22 cm, respectively. Its species was *Jidan* 209, which is a prevailing maize species in northeast China. Five shooting periods were selected, and specific environmental conditions at that time are listed in Table 2.

Each shooting period acquired 250 images. Then, *LabelImage* (Google Brain, Version 1.7) was applied to mark the individual maize plant in these images. During labelling process, each marked object would generate a corresponding and unique label; furthermore, foreground maize plants were picked out in this procedure, as shown in Figure 4. Corresponding labels were going to be used in the training recognition model. Subsequently, these marked images were separated into the training set and testing set according to a ratio of 9:1 randomly.

2.2. Multiscale Hierarchical Feature Extraction. In order to enhance target-image details, a fast local Laplacian filter was applied to each image, then forming multiscale Gaussian pyramid images (MGPI) for each input image. As shown in Figure 5, the adjacent latter image's area was converted to $1/4$ of the former image, while its resolution was converted to only $1/2$ of the former image; thus, local details could be enhanced for target images [25].

Input images were increased several times by constructing the MGPI. From another view, each MGPI formed a convolutional neural network (CNN). In order to make all CNN models can share parameters, based on normalization transformation theory [26], we transformed local neighboring areas of the above MGPI with zero mean and unit standard deviation. In following studies, l_s was designated as a CNN model and β_s was designated as an inner parameter, as shown in the following equation:

$$\beta_s = \beta_0, \quad (1)$$

where β_0 is an initial parameter of a CNN model, s is a subscript, which designates the CNN model number, $s \in \{1, 2, \dots, N\}$, and N designates any integer.

FIGURE 1: Schematic diagram of image acquisition.

As for one specific s , l_s could be represented as

$$l_s(J_s; \beta_s) = W_N O_N - 1, \quad (2)$$

where W_N is a weight matrix of the N^{th} layer for a CNN model, O_{N-1} is the output of the $(N-1)^{\text{th}}$ layer of the same CNN model, J represents a specific image of the MGPI, and $O_0 = J_0$. Generally, the output of a hidden layer (y) could be represented as

$$O_y = \text{conv}[\tanh(W_N O_N - 1 + b_N)], \quad (3)$$

where y is a subscript that designates the layer number, $y \in \{1, 2, \dots, N-1\}$, “conv” represents convolutional operation, “tanh” represents activating function, and b_N represents a bias matrix for the N^{th} layer.

2.3. Convolutional Kernel Expanded. As for one specified image of the MGPI, in order to extract image features, convolutional operation will be the first step with regard to a classical CNN model. The second step is pooling operation, which can decrease image size and, meanwhile, increase the image's receptive field [27]. In classical CNN models, owing to pooling operation decreases image size, in order to retain the image size, upscaling should be taken after each pooling. Images loss useful information during each pooling and upscaling processes [27]. In order to compensate the above disadvantage, the “zero element” was placed in the middle of classical convolutional kernels in this study, as shown in Figure 6. That is to say, the classical convolutional kernel was expanded. This operation could also increase receptive field, so the pooling operation was abandoned [28].

Similar to the classical CNN, fully connected operation is necessary followed by expanded convolutional operation, and featured map was extracted within each CNN model simultaneously. Subsequently, upscaling operation would occur in each CNN model, so as to normalize image sizes. Consequently, all the feature maps were assembled together to form a multiscale hierarchical features, as shown in Figure 7.

The maize plant recognition model was trained by *TensorFlow* 1.11.0. Owing to the layer of the MGPI is the number of CNN models, there were three MGPI layers; thus, this study needed to construct three parallel CNN models. Each CNN model had 11 convolutional layers, and their weight matrixes, bias matrixes, activating functions, and convolutional kernels were mentioned above. In order to get

FIGURE 2: Positional relationships between maize ears and maize stalks. Maize ears were surrounded by red rectangles. (a) One maize ear is in front of the maize stalk, (b) one maize ear is behind the maize stalk, (c) one maize ear is on the left side of the maize stalk, and the other two maize ear's silks could be seen, and (d) one maize ear is on the right side of the maize stalk.

FIGURE 3: Maize ears were sheltered by maize leaves. Maize ears were surrounded by red rectangles.

rid of dependence on high-speed computer hardware, this study transformed *TensorFlow*-generated “cpkt” files to “pb” files and its corresponding “pbtxt” files. Thus, the *Python* language combined with *OpenCV* can call the recognition model by a moderate-speed computer.

2.4. Morphological Operation. The recognition model was generated through training by deep learning, as stated above, and the model usually has a relative low accuracy in real filed tests. So, this study adopted two steps to recognize maize ears: the first step was to recognize a whole maize plant and the second step was to recognize maize ears within an already known range. Specific procedures are presented below. Colorful images were converted into gray images by the function of *OpenCV* (see the program of `cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)`), and the transforming result is shown in Figure 8.

In order to conduct morphological operations, these gray images were transformed into binary images. Computing optimal threshold was a key link in binary transformation. Due to there are 256 gray levels for gray images, it can be assumed that n represents the total pixel number, and i represents a specified

gray level; thus, $n = n_0 + n_1 + \dots + n_{255}$; if p_i is used to represent occurrence probability of a specified gray level, then following equation can be deduced:

$$p_i = \frac{n_i}{n}. \quad (4)$$

Owing to $\sum_{i=0}^{255} p_i = 1$, the pixel distribution probability of a gray level that is lower than a certain L ($0 < L < 255$) should be presented as

$$\omega = \sum_{i=0}^L p_i. \quad (5)$$

Equations (6) and (7) represent pixel mean (μ) and variance (σ^2), respectively:

$$\mu = \sum_{i=0}^{255} i \frac{p_i}{\omega}, \quad (6)$$

$$\sigma^2 = \frac{1}{\omega} \sum_{i=0}^{255} (i - \mu)^2 \times p_i. \quad (7)$$

The algorithm proposed in this study was used to divide one image into four subblocks (R_e^g , $g \in [1, 2, 3, 4]$, where e represents the iteration number), which were the upper left, upper right, lower left, and lower right, respectively. Then, each subblock's subdivision was carried out, where μ and σ^2 were computed again in each subblock. Score (R_e^g) was proposed to represent each subblock's score, which can be calculated:

$$\text{Score}(R_e^g) = \mu - \sigma^2. \quad (8)$$

The subblock with the highest Score (R_e^g) was divided into four smaller subblocks in following iterations. The above steps were continued until stopping condition was satisfied. When optimizing the threshold in each final subblock, the Gaussian fitting method [29] was the basic theory, which required bimodal or obscure bimodal distribution for pixel histograms [30]. Consequently, pixel histogram of the final subblock occurred bimodal or obscure

FIGURE 4: Image marking. Only maize plants in foreground were marked by red rectangles, so background images could be avoided during training process.

FIGURE 5: Sketch map of multiscale Gaussian pyramid images (MGPIs). The area of J_3 is $1/16$ of J_1 , and the pixel of J_3 is $1/4$ of J_1 , so the resolution of J_3 is increased by four times.

bimodal distribution was regarded as stopping condition in this study. The other three subblocks which generated in the first division would be divided into the same amount of smaller subblocks as the highest score (R_c^g) one. FSB is short

for the final smaller subblocks in following paragraphs. In order to calculate initial threshold “ t ” for each FSB, one objective function [$J(t)$] was defined as

$$J(t) = 1 + e^{\frac{\sigma_0^2}{\omega(t)}} + \times e^{\frac{\sigma_0^2}{[1 - \omega(t)]}}, \quad (9)$$

where $\omega(t)$ denotes distribution probability for initial threshold “ t ”. Each initial threshold “ t ” could be obtained by

$$t = \arg \left\{ \min_{0 \leq t < 255} J(t) \right\}, \quad (10)$$

The final optimal threshold for each FSB can be obtained according to following four steps. Firstly, each FSB was transformed into binary images by the initial threshold “ t ”. Then, the FSB was divided into two parts, which were designated as C_0 and C_1 . Secondly, mean pixel values of μ_0 and μ_1 as well as pixel variances of σ_0 and σ_1 were calculated, respectively. In C_0 and C_1 , two Gaussian distribution functions, i.e., $N_0(\mu_0, \sigma_0^2)$ and $N_1(\mu_1, \sigma_1^2)$, were obtained by the fitting method according to equation (11), and then probability density functions were calculated within the range of $[\min(\mu_0, \mu_1), \max(\mu_0, \mu_1)]$:

$$p(x|q) = \frac{1}{\sqrt{2\pi}\sigma_q} e^{-\frac{(x-\mu_q)^2}{2\sigma_q^2}}, \quad (11)$$

where x denotes the pixel value of each FSB and $q=0, 1$.

Thirdly, μ_{begin} was designated to represent the mean value of the Gaussian distribution whose variance was relative smaller and μ_{end} was designated to represent the mean value of the other Gaussian distribution. The discrete μ value, ranging from μ_{begin} to μ_{end} , was substituted into equation (11), and the iteration operation was kept until $p(x|\mu_{\text{begin}}) \leq p(x|\mu_{\text{end}})$. The optimal threshold was designated as $\text{th}_{q+1} = x$ for this iteration. Fourthly, if following situation occurred between two adjacent iterations, i.e., $|\text{th}_q + 1 - \text{th}_q| \leq \Delta$ (Δ denotes given tolerance error, which

FIGURE 6: Schematic diagram of expanded convolutional kernels. (a) 3×3 unit convolutional kernel, (b) convolutional kernel with double expanding based on (a), and (c) convolutional kernel with quadruple expanding based on (a).

FIGURE 7: Network structural diagram of extracting multiscale hierarchical features. There are three MGPI layers in this study.

FIGURE 8: Original image and gray image. (a) Original image and (b) Gray image.

was selected via experimental verification), th_q was designated as final optimal threshold for each FSB. According to the above four steps, one binary image was obtained, as shown in Figure 9. Compared with transforming one gray image into binary image with one specific threshold, this method gave the best optimal threshold to each specific subblock; therefore, this method kept as much useful information as possible and lost least information.

It is very important to acquire gradients between different objects, so as to segment different objects. Specifically, useless objects should be discarded and scattered elements of maize ears should be jointed as shown in Figure 9. After experimental verification, one structural element $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ was subjected to four opening operations and one independent dilating operation. One morphological result is shown in Figure 10.

2.5. Maize Ear Recognition by Multifeature Fusion. During dough stage, there are many obvious differences between maize ears and other parts of a maize plant, such as the maize ear is similar to football shape, and colorful silk usually exposed at the top of a maize ear. In image morphological process, a whole maize plant would be separated into several parts, and this study selected the following features to recognize maize ears.

2.5.1. Aspect Ratio. Considering a maize ear's outline, the aspect ratio of its bounding rectangle should be regarded as a preferred indicator, which can be calculated as

$$AT = \frac{WIDTH}{HEIGHT} \times 100\%, \quad (12)$$

where AT means the aspect ratio and WIDTH and HEIGHT mean the width and the height of the bounding rectangle.

2.5.2. Convex-Concave Degree. The convex-concave degree is calculated according to the following equation:

$$CONV = \frac{CHORD}{PERI} \times 100\%, \quad (13)$$

where CONV means the convex-concave degree, CHORD means the maximum chord length, and PERI means the perimeter of the objects' outline.

Morphological operation provides basis for contour detection, at the same time, through analyzing 1250 images that mentioned in the section of "Image Acquisition and Labelling," and we found that maize ear's AT fell into a range of (0.33, 0.41) and its CONV fell into a range of (0.46, 0.48). Based on equations (12) and (13), this study regarded objects which have qualified AT and CONV as a maize ear's main body in this step. One recognition result is shown by marking blue on the original image and one maize ear's main body (excluding silk) is shown in Figure 11.

2.5.3. Color. Based on colorful images which have not been transformed into gray images, such as Figure 8(a), this study

FIGURE 9: Binary image.

FIGURE 10: Morphological image.

combined HSV (hue, saturation, and value) and RGB (red, green, and blue) color space to distinguish maize ears. It is obvious that maize ear's silk has different colors compared with the other parts. During dough stage, maize ear's silk will range [26–34, 43–255, 46–255] of HSV color space and it will range [184–240, 134–230, 11–140] of RGB color space by testing the 1250 images in the section of "Image Acquisition and Labelling." Firstly, this study would choose pixels that fell into the above colorful ranges, and then their colorful moments were calculated. Their first moments, second moments, and third moments were calculated according to equations (14)–(16), which represented the means, variances, and skewness, respectively:

$$M_{i1} = \frac{1}{N} \sum_{j=1}^N P_{ij}, \quad (14)$$

$$M_{i2} = \left[\frac{1}{N} \sum_{j=1}^N (P_{ij} - M_{i1})^2 \right]^{1/2}, \quad (15)$$

FIGURE 11: Contour detection of maize ear's main body.

FIGURE 12: Recognition of maize ear's silk.

$$M_{i3} = \left[\frac{1}{N} \sum_{j=1}^N (P_{ij} - M_{i1})^3 \right]^{1/3}, \quad (16)$$

where M_{i1} , M_{i2} , and M_{i3} represent first moments, second moments, and third moments of color features, N represents the total pixel number, and P_{ij} represents occurring probability of pixel whose gray level is “ j ” and possesses a color channel of “ i .”

If the above pixels with significantly ($P < 0.05$) different colorful moments (M_{i1} , M_{i2} , and M_{i3}) compared with their neighboring-area pixels, they would be filled with red to represent maize ear's silk, as shown in Figure 12.

At last, this study filled the maize ear's main body contour with red and then overlapped the main body and the silk, so one whole maize ear was recognized as shown in Figure 13.

FIGURE 13: Recognition of a whole maize ear.

2.6. Experiments. Five days after first image acquisition, maize-ear-recognition experiment was taken place in the same field but with brandnew maize rows. This experiment lasted for three days, and the maize was still at dough stage. In order to shoot maize images similar to picking robots, one camera operator held the camera and walked along the maize rows, and camera lens was set up to a mean height of maize ear-nodes and ~ 30 cm between maize rows. There were five shooting periods, which were the same as the first image acquisition day. Specific environmental conditions of the three experimental days (from August 20th to August 22nd) are shown in Table 1.

The operator's forward speed was ~ 0.5 m/s. Camera was set to automatic exposure mode, with an exposure time of 1/100 s and exposure frequency of 3 Hz. Twice shootings were taken for each shooting period, and each lasted for ~ 2 min. According to the parameters mentioned above, the operator walked for ~ 60 m, and ~ 360 images were taken during each shooting. There were six replicates for each shooting period, so 30 shootings occurred and ~ 10800 images were acquired totally.

The maize-ear-recognition accuracy was calculated according to the following equation:

$$\alpha = \frac{\gamma - \partial}{\varphi} \times 100\%, \quad (17)$$

where α is the maize-ear-recognition accuracy, γ is the recognized maize ear by this algorithm, ∂ is the misrecognized maize ear (e.g., maize leaves were recognized as maize ears), and φ is the real total maize ears. α , ∂ , and φ were obtained by counting manually.

It is not necessary to use high-level hardware used for calculating the above recognition model, and image processing hardware was described as follows: the central processing unit (CPU) was Intel(R) Core™ i5-6200 CPU @2.30 GHz, operational memory was 4.00 GB, and operational system was WIN 10. The programing tool *Python* 3.6.5 incorporated with *OpenCV* 3.4.2 was used to calculate the recognition model files. Software of *SPSS* 22.0 for Windows (IBM Inc., USA) was applied in this study to carry out statistical analyses.

TABLE 1: Mean environmental parameters in experimental field.

Shooting period	7:00–8:30	10:00–11:30	12:00–13:30	15:00–16:30	18:00–19:30
Mean solar altitude (°)	27.34	54.68	56.98	32.38	0.04
Mean solar azimuth (°)	99.92	156.45	194.13	250.66	307.19
Light color temperature/K	7634	10255	16309	11076	2096
PM 2.5 ($\mu\text{g m}^{-3}$)	32	16	17	12	46

Notes. PM 2.5 designates the fine particulate matter that is less than $2.5 \mu\text{m}$.

3. Results and Discussion

3.1. Maize Plant Recognition. The training results are shown in Figure 14, since 25th training epochs, the recognition accuracy could maintain a relative stable level, and its final recognition accuracy reached 99.47% after completing 51 training epochs. One study conducted by Grinblat et al. [31] who adopted a similar CNN model as this study, but lacked multiscale hierarchical features and expanded convolution kernels, resulted in a highest recognition accuracy of 96.9%. We tried to mark maize ears directly as shown in Figure 4 and then proceeded the same training processes as above-mentioned, and the results are also presented in Figure 14. It shows that the recognition accuracy had a relative larger amplitude during entire training process, and the final accuracy only reached ~94%. As stated by Voose [16], deep learning is a black box, there is still limited knowledge about why the results are so different through marking different objects. Owing to lower recognition accuracy by marking maize ears directly, it should be a wise determination of recognizing the whole maize plant first and then recognizing maize ears within already known area (the whole maize plant was in this area). In addition, interference caused by background objects could be avoided through the first step. Separating foreground and background is critical in object recognition, and this kind of work usually needs complicated algorithms and costs a lot computing time [10]; furthermore, if foreground-background separating is not a final goal, separating error will be accumulated and passed to the next step. This study also tried to increase training epochs to 101 and 151, but increased training epochs did not improve recognition accuracy. As approved by Günter et al. [32] who indicated that in computer vision fields, there is not positive correlation between recognition accuracy and training epochs, the optimal training epoch is usually obtained through practical experiment. In other words, this study tested three kinds of training epochs, we testified that the recognition model could get a stable recognition accuracy under 51 training epochs; thus, there was no need to increase training epochs in this study.

Deep learning needs a huge amount of calculating, which needs high-speed computational hardware [33]. However, it is hard to equip high-speed computers in real productions because of cost reasons; thus, the whole maize plant recognition model was transformed into the “pb” file and corresponding “pbtxt” file, which could be called by *OpenCV*, due to *OpenCV* could be operated by moderate-speed computers, other than relying on high-speed computers. The experiment got 10800 images in the section of “Experiments,” and as for whole maize plant recognition,

OpenCV testing results show that average recognition accuracy was 99.17%. This indicates recognition model has transportability and practical application value.

3.2. Maize Ear Recognition. Field experiment was lasted for three days, there were five shootings every day. Statistical analysis shows that the average maize-ear-recognition accuracy was 97.02% and they were 96.93, 96.56, 97.41, 97.02, and 97.18% for 7:00–8:30, 10:00–11:30, 12:00–13:30, 15:00–16:30, and 18:00–19:30, respectively. Although research studies about maize ear recognition are still rare, similar agriculture-related recognition had been conducted in orchard fruits. Mureşan and Oltean [15] utilized deep learning to detect fruits, and their detection accuracy was 96.3% in the testing set. Wei et al. [21] designed an automatic fruit object extraction method for fruit-picking robots, and their design was mainly about image processing, especially about binary transformation methods; their study got a final extraction accuracy of more than 95%. Zhang [4] got an accuracy of 94.6% based on deep learning, and their research purpose was to distinguish different maize cultivars through maize ear images. Zhang et al. [5] got the accuracies of 96, 93.3, and 90% for mildew ears, worm ears, and mechanical damaged ears, respectively. Compared with the above studies, this study obtained a relative higher accuracy. Further statistical analysis shows that shooting periods do not have significant influence ($P > 0.05$) on maize-ear-recognition accuracy.

Table 2 and Table 1 illustrate that there existed weather condition differences between the first image acquisition period (August 15th, 2018) and the field experiment period (from August 20th to August 22nd). First image acquisition was used for training the whole maize plant recognition model, and field experiment was aiming to test the whole algorithm proposed by this study. It is well known that solar altitude and solar azimuth are changing all day long. Student’s paired t tests show light color temperatures between two periods had significant ($P < 0.05$) difference, and two periods’ PM 2.5 s also had slightly difference but without significance ($P > 0.05$). Above statements presented clearly that although under different conditions, this maize-ear-recognition algorithm has extensive adaptability. In other words, under certain shooting angles and shooting distances, illumination conditions will not have severe impact on maize-ear-recognition accuracy. Before this study, maize yield prediction was mainly depended on maize plant population counting [34]; due to one specific maize plant can grow several maize ears, this maize-ear-recognition algorithm may do great contribution to improve the accuracy of crop yield prediction in the future.

FIGURE 14: Variable curves of image recognition accuracy. Direct maize ear recognition resulted in lower accuracy compared with the whole maize plant recognition.

As shown in Figure 15, positional relationships between maize ears and maize stalks belonged to extreme positions, and this algorithm can recognize maize ears under four representative positions; therefore, this algorithm can also recognize maize ears that grow in other positions, but this paper could not present all recognized images due to paragraph reasons. In terms of Figure 15(b), the maize ear's main body was not recognized by this algorithm, but the maize ear's silk was recognized successfully. Owing to the maize ear was behind the maize stalk sometimes, this algorithm set the following rules: it was based on the image size which was 1280×800 pixels, and the shooting distance which was ~ 30 cm; if centroid distance between any two recognized objects was smaller than 100 pixels, two recognized objects are regarded as one object. Further, if any two recognized objects were fused with each other, they are also regarded as one object, too. Otherwise, recognized objects are regarded as separate objects. According to the above rules, this study recognized one maize ear in Figure 15(b) and three maize ears in Figure 15(c). Owing to agricultural field is complex, it was a wise method to utilize auxiliaries in order to recognize main objects. Van Henten et al. [17] utilized an auxiliary wire to help a deleafing robot recognize cucumber vines. Bac et al. [18] utilized supporting wires to detect sweet peppers' stems. Similar to the above two studies, due to the maize was in dough stage, the maize ear's silk acted as an import auxiliary in this study.

It should be mentioned that this algorithm only detected the maize ear's silk as shown in Figure 15(b), and same situations occurred in Figure 15(c). Meanwhile, considering extreme situations that maize ears were sheltered by maize leaves, this algorithm could detect partial maize ears, as shown in Figure 16(a), which only recognized maize ear's main body. If maize ear's silk was exposed to lens, this algorithm could detect it successfully (Figure 16(b) only recognized maize ear's silk). In one word, although this algorithm could not recognize the whole maize ear under some exceptional circumstances, according to the above-mentioned rules, partial sheltering does not have severe influence on maize ear recognition or maize ear counting.

3.3. Algorithm Advantages and Recognizing Error Analyses. This study adopted *LabellImage*, combining deep learning and image morphological process. When marking maize plants by *LabellImage*, foreground objects would be emphasized and background objects can be avoided. Otherwise, background objects should be a big trouble for image morphological process.

Binary transformation is a critical link in image recognition, and threshold determination is a key parameter in binary transformation. Common threshold determination methods included maximum interclass variance Otsu [35], maximum entropy [36], and minimum error [37]. These methods caused severe information losses, as shown in Figure 17. Based on the above binary transformation methods, this study could not recognize maize ears or could only get lower accuracies. Based on this study's main framework, threshold determination methods were substituted by the above methods, and comparison results are shown in Table 3. The binary threshold determination proposed in this study was based on the Gaussian fitting [29], which can optimize threshold for each subimage block.

Based on the "time" function which supplied by the *Python* language, this study calculated time consumption from the whole maize plant recognition to the maize ear recognition. The hardware and software were described in the section of "Experiments." Calculated results show that the average time consumption was 0.39 s for each image process. Although this paper's algorithm costed longer executive time (Table 3), according to prevailing maize-cultivating agronomy, the average intrarow distance is 250 mm [38]. Based on the shooting distance and shooting angle adopted in this study, each image will contain about two maize plants, so maximum forward speed could reach to 1.28 m/s (~ 4.61 km/h). In other words, this speed could catch up prevailing maize combine harvesters [39].

It is well known that cropland circumstance could be as complicated as we cannot imagine. Although this study tried its best to detect unique characters of maize ears, some maize ears were totally sheltered by maize leaves, or only a small part was exposed to lens, so recognition errors were mainly

TABLE 2: Average environmental parameters during the first image acquisition.

Shooting period	7:00–8:30	10:00–11:30	12:00–13:30	15:00–16:30	18:00–19:30
Mean solar altitude (°)	28.71	56.53	58.93	33.78	1.46
Mean solar azimuth (°)	98.22	155.95	194.43	252.16	309.88
Light color temperature (K)	10293	20221	20149	9872	2593
PM 2.5 ($\mu\text{g}\cdot\text{m}^{-3}$)	14	12	12	11	23

Notes. PM 2.5 designates the fine particulate matter that is less than $2.5\ \mu\text{m}$.

(a)

(b)

(c)

(d)

FIGURE 15: Recognition results of four representative positions between maize ears and maize stalks. The recognized objects were filled with red.

(a)

(b)

FIGURE 16: Recognition results of maize ears that sheltered by maize leaves. The recognized objects were filled with red.

FIGURE 17: Binary images transformed by other threshold determination methods. (a) Image transformed by maximum interclass variance Otsu, (b) image transformed by maximum entropy, and (c) image transformed by minimum error.

TABLE 3: Comparative results of recognition accuracies and executive times.

	Mean recognition accuracy (%)	Mean executive time (s)
This paper's algorithm	97.02	0.39
Otsu	3.1	0.23
Maximum entropy	1.1	0.25
Minimum error	2.1	0.09

Notes. Otsu means maximum interclass variance.

due to not enough unique characters could be detected. Beyond that, some crossing maize leaves formed unexpected morphological characters and these characters led their aspect ratios or convex-concave degrees fell into the setting range during morphological operations. All of these reasons caused recognition errors in this study.

Crop-related objects are much more complicated than factory-level objects, so there will not be standard morphological or colorful features for maize ears. Pursuing a higher recognition accuracy is an endless work, and discovering more unique features and optimizing multifeature fusion shall be our future work.

4. Conclusions

- (1) This study combined deep learning and image processing to realize maize ear recognition. Since deep learning does not need to depict object details and image processing does not rely on high-performance hardware, thus this algorithm insured recognition accuracy and recognition speed.
- (2) In order to train the whole maize plant recognition model, this study developed multiscale hierarchical feature extraction and the expanded convolutional kernel. Subsequently, multifeatures of maize ear were fused, so as to recognize maize ears. Experimental

results showed that average maize ear recognition accuracy was 97.02% and average time consumption of each image was 0.39 s, and this recognition speed could catch up the forward speed of commonly used maize harvesters.

- (3) In spite of the whole maize ear could not be recognized under some special circumstances, either maize ear's main body or maize ear's silk could be recognized by this algorithm. Partial recognition still does the same contribution to maize ear recognition and maize ear counting.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (31901408), the Science and

Technology Development Plan of Jilin Province (20180414074GH), and the National Key Research and Development Program (2018YFD0300207). Special thanks are due to the Austrian Agency for International Cooperation in Education and Research (OeAD-GmbH).

References

- [1] A. Bechar and C. Vigneault, "Agricultural robots for field operations: concepts and components," *Biosystems Engineering*, vol. 149, pp. 94–111, 2016.
- [2] L. Xiong, S. Sun, and M. Xiao, "Agricultural machinery automation and intelligent research and application," *IOP Conference Series: Materials Science and Engineering*, vol. 452, Article ID 042077, 2018.
- [3] Z. H. Yu, H. B. Zhou, and C. N. Li, "An image-based automatic recognition method for the flowering stage of maize," in *Proceedings of the MIPPR 2017: Remote Sensing Image Processing, Geographic Information Systems, and Other Applications*, SPIE, Xiangyang, China, October 2017.
- [4] F. Zhang, *Classification and Identification of Corn Ear Based on Deep Learning*, Jiangxi Agricultural University, Nanchang, China, 2019.
- [5] F. Zhang, S. Li, Z. Liu, D. Zhu, Y. Wang, and Q. Ma, "Screening method of abnormal corn ears based on machine vision," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 46, no. 1, pp. 45–49, 2015.
- [6] H. Jia, S. Li, G. Wang, and H. Liu, "Design and experiment of seedling avoidable weeding control device for intertillage maize (*Zea mays* L.)," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 34, no. 7, pp. 15–22, 2018.
- [7] H. Liu, H. Jia, G. Wang, S. Glatzel, H. Yuan, and D. Huang, "Method and experimental of maize (*Zea mays* L.) stems recognition based on deep learning and image processing," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 51, no. 4, pp. 207–215, 2020.
- [8] L. Hu, X. Luo, S. Zeng, Z. Zhang, X. Chen, and C. Lin, "Plant recognition and localization for intra-row mechanical weeding device based on machine vision," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 29, no. 10, pp. 12–18, 2013.
- [9] L.-Y. Chang, S.-P. He, Q. Liu, J.-L. Xiang, and D.-F. Huang, "Quantifying muskmelon fruit attributes with a tep-based model and machine vision measurement," *Journal of Integrative Agriculture*, vol. 17, no. 6, pp. 1369–1379, 2018.
- [10] J. Tang, G. Yang, Y. Sun, J. Xin, and D. He, "Salient object detection of dairy goats in farm image based on background and foreground priors," *Neurocomputing*, vol. 332, pp. 270–282, 2019.
- [11] G. Szwoch, "Extraction of stable foreground image regions for unattended luggage detection," *Multimedia Tools And Applications*, vol. 75, no. 2, pp. 761–786, 2016.
- [12] C. Wang and Z. Li, "Weed recognition using SVM model with fusion height and monocular image features," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 32, no. 15, pp. 165–174, 2016.
- [13] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.
- [14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [15] H. Mureşan and M. Oltean, "Fruit recognition from images using deep learning," *Acta Universitatis Sapientiae, Informatica*, vol. 10, no. 1, pp. 26–42, 2018.
- [16] P. Voosen, "How ai detectives are cracking open the black box of deep learning," *Science*, 2017, <https://www.sciencemag.org/news/2017/07/how-ai-detectives-are-cracking-open-black-box-deep-learning#>.
- [17] E. J. Van Henten, B. A. J. Van Tuijl, G.-J. Hoogakker et al., "An autonomous robot for de-leafing cucumber plants grown in a high-wire cultivation system," *Biosystems Engineering*, vol. 94, no. 3, pp. 317–323, 2006.
- [18] C. W. Bac, J. Hemming, and E. J. Van Henten, "Stem localization of sweet-pepper plants using the support wire as a visual cue," *Computers and Electronics in Agriculture*, vol. 105, pp. 111–120, 2014.
- [19] N. Brichet, C. Fournier, O. Turc et al., "A robot-assisted imaging pipeline for tracking the growths of maize ear and silks in a high-throughput phenotyping platform," *Plant Methods*, vol. 13, no. 1, p. 96, 2017.
- [20] U.-O. Dorj, M. Lee, and S.-S. Yun, "An yield estimation in citrus orchards via fruit detection and counting using image processing," *Computers and Electronics in Agriculture*, vol. 140, pp. 103–112, 2017.
- [21] X. Wei, K. Jia, J. Lan, Y. Li, Y. Zeng, and C. Wang, "Automatic method of fruit object extraction under complex agricultural background for vision system of fruit picking robot," *Optik*, vol. 125, no. 19, pp. 5684–5689, 2014.
- [22] M. Guo, J. Wang, and S. Zhu, "Research on design of intelligent mobile fruit picking robot based on neural network algorithm and socp technology," *Journal of Advanced Oxidation Technologies*, vol. 21, no. 2, 2018.
- [23] M. Jhuria, A. Kumar, and R. Borse, "Image processing for smart farming: detection of disease and fruit grading," in *Proceedings of the IEEE Second International Conference on Image Information Processing (ICIIP-2013)*, IEEE, Shimla, India, December 2013.
- [24] V. Athanasios, D. Nikolaos, B. George, and S. Tania, "Recent developments in deep learning for engineering applications," *Computational Intelligence and Neuroscience*, vol. 2018, Article ID 8141259, 2 pages, 2018.
- [25] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *Readings in Computer Vision*, vol. 31, no. 4, pp. 671–679, 1983.
- [26] P. Soo-Chang and C.-N. Lin, "Image normalization for pattern recognition," *Image and Vision Computing*, vol. 13, no. 10, pp. 711–723, 1995.
- [27] R. Venkatesan, *Convolutional Neural Networks in Visual Computing: A Concise Guide*, CRC Press, Boca Raton, FL, USA, 2018.
- [28] J. Toft, "Continuity properties in non-commutative convolution algebras, with applications in pseudo-differential calculus," *Bulletin des Sciences Mathématiques*, vol. 126, no. 2, pp. 115–142, 2002.
- [29] P. A. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, Kauai, HI, USA, December 2001.
- [30] R. Lienhart, A. Kuranov, and V. Pisarevsky, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection," *Lecture Notes in Computer Science*, vol. 2781, pp. 297–304, 2003.
- [31] G. L. Grinblat, L. C. Uzal, M. G. Larese, and P. M. Granitto, "Deep learning for plant identification using vein morphological patterns," *Computers and Electronics in Agriculture*, vol. 127, pp. 418–424, 2016.
- [32] S. Günter and H. Bunke, "Hmm-based handwritten word recognition: on the optimization of the number of states,

- training iterations and gaussian components,” *Pattern Recognition*, vol. 37, no. 10, pp. 2069–2079, 2004.
- [33] M. Mohammed, *Machine Learning*, Taylor & Francis, Abingdon, UK, 2017.
 - [34] H. Jia, G. Wang, M. Guo, S. Dylan, X. Jiang, and J. Zhao, “Methods and experiments of obtaining corn population based on machine vision,” *Transactions of the Chinese Society of Agricultural Engineering*, vol. 31, no. 3, pp. 215–220, 2015.
 - [35] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
 - [36] T. Pun, “A new method for grey-level picture thresholding using the entropy of the histogram,” *Signal Processing*, vol. 2, no. 3, pp. 223–237, 1980.
 - [37] J. Kittler and J. Illingworth, “Minimum error thresholding,” *Pattern Recognition*, vol. 19, no. 1, pp. 41–47, 1986.
 - [38] J. Zhao, X. Yang, and S. Sun, “Constraints on maize yield and yield stability in the main cropping regions in China,” *European Journal of Agronomy*, vol. 99, pp. 106–115, 2018.
 - [39] Y. Li, C. Tao, Q. Zhe et al., “Development and application of mechanized maize harvesters,” *International Journal of Agricultural and Biological Engineering*, vol. 9, no. 3, pp. 15–28, 2016.