Maize (*Zea mays* L.) seedling detection based on the fusion of a modified deep learning model and a novel Lidar points projecting strategy

Gang Wang[1*], Dongyan Huang[2], Deyi Zhou[1], Huili Liu[1], Minghao Qu[1], Zhongyang Ma[1]

(1. *College of Biological and Agricultural Engineering, Jilin University*, *Changchun 130022, China*;
2. *Key Laboratory of Bionic Engineering, Ministry of Education, Jilin University, Changchun 130022, China*)

Abstract: Accurate crop detection is the prerequisite for the operation of intelligent agricultural machinery. Image recognition usually lacks accurate orientation information, and Lidar point clouds are not easy to distinguish different objects. Fortunately, the fusion of images and Lidar points can complement each other. This research aimed to detect maize (*Zea mays* L.) seedlings by fusing Lidar data with images. By applying coordinate transformation and time stamps, the images and Lidar points were realized homogeneous in spatial as well as temporal dimensions. Deep learning was used to develop a maize seedling recognition model, then the model recognized maize seedlings by labeling them with bounding boxes. Meanwhile, Lidar points were mapped to the bounding boxes. Only one-third of points that fell into the right middle of bounding boxes were selected for clustering operation, the calculated center of the cluster provided spatial information for target maize seedlings. This study modified the classical single shot multi-box detector (SSD) by merely linking the last feature map to the final output layer, owing to the higher feature maps having the unique advantages of detecting relatively larger objects. In images, maize seedlings were just the largest objects owing to be shot on purpose. This modification enabled the recognition model to finish recognizing an image by only consuming around 60 ms, which saved about 10 ms/image compared with the classical SSD model. The experiment was conducted in a maize field, and the maize was during the elongation stage. Experimental results demonstrated that the standard deviations for maximum distance error and maximum angle error were 1.4 cm and 1.1°, respectively, which can be tolerated under current technical requirements. Since agricultural fields are subject to staple crop-orientated and changeable ambient environment, the fusion of images and Lidar points can derive more precision information, and make agricultural machinery smarter. This study can act as an upstream technology for other researches on intelligent agricultural machinery.

Keywords: maize seedling, detection, fusion, deep learning, Lidar
DOI: 10.25165/j.ijabe.20221505.7830

Citation: Wang G, Huang D Y, Zhou D Y, Liu H L, Qu M H, Ma Z Y. Maize (*Zea mays* L.) seedling detection based on the fusion of a modified deep learning model and a novel Lidar points projecting strategy. Int J Agric & Biol Eng, 2022; 15(5): 172–180.

1 Introduction

Indisputably, intelligent agricultural machinery is the future development direction. More and more sensors are equipped for agricultural machinery to let them become smarter, such as cameras and Lidar. The cameras' advantage is that they can present visible images and let agricultural machines know what they see, the Lidar's advantage is that it can present exact spatial information by sparse points[1]. Therefore, the fusion of the camera and Lidar can extend reliability for each other[2].

All agricultural works serve crops. For example, in the weeding control stage, the crops are going to be avoided and weeds are going to be removed. In the harvesting stage, the crops are going to be touched. No matter which agricultural procedures, distinguishing crops and weeds, as well as obtaining crop locations are very important. As for intelligent agricultural machinery, recognizing crops and mastering their locations are pre-technologies. Nowadays, image processing and deep learning are used widely in all kinds of research fields. Image processing is the technology of using computers to analyze digital images to achieve the desired results[3]. Deep learning is to learn the inherent laws and presentation levels of sample data, and its ultimate goal is to enable machines to have the ability to analyze and learn like human beings[4,5]. As a preliminary technology, image processing plays an important role in crop recognition, but compared with image processing only based on morphological features, deep learning has stronger generalization and robustness for crop detection[4].

Up to now, scientists had developed many object detection models based on deep learning framework, including the serial model of a Region-based Convolutional Network (R-CNN)[6-8], the serial model of 'You Only Look Once' (YOLO)[9-11], as well as single shot multi-box detector (SSD), and so on. In terms of R-CNN, which belongs 'two-stage' object detection strategy[6], it needs about four steps to complete object detection. As for each picture, R-CNN applies selective search to get many proposal

Received date: 2022-08-02 **Accepted date:** 2022-09-29

Biographies: Dongyan Huang, PhD, Professor, research interest: design of intelligent agricultural machinery, Email: huangdy@jlu.edu.cn; **Deyi Zhou**, PhD, Associate Professor, research interest: agricultural mechanization engineering, Email: 2297687739@qq.com; **Huili Liu**, Bachelor, Senior Engineer, research interest: deep learning and artificial intelligence, Email: liuhuili@163.com; **Minghao Qu**, Master, Doctoral candidate, research interest: design of intelligent agricultural machinery and multi-sensor fusion, Email: qumh19@mails.jlu.edu.cn; **Zhongyang Ma**, Bachelor, Postgraduate, research interest: design of intelligent agricultural machinery and multi-sensor fusion, Email: zyma20@mails.jlu.edu.cn.

***Corresponding author: Gang Wang**, PhD, Associate Professor, research interest: design of intelligent agricultural machinery and multi-sensor fusion. College of Biological and Agricultural Engineering, Jilin University, Changchun 130022, China. Tel: +86-13166888540, Email: wangganggoodman@163.com.

regions[12], each region will be scaled and imported to a Convolutional Neural Network (CNN) to extract features. A Support Vector Machine (SVM) will be used to classify these features. Aiming at these classified proposal regions, training a linear regression model to predict real bounding boxes, which showing target object labels and positions. R-CNN applies a pre-trained neural network model to extract features, compared with artificial features, R-CNN increases target detection accuracy efficiently[6]. YOLO is a ‘one-state’ target detection frame, different from R-CNN serials, YOLO does not need to get proposal regions, it uses a single neural network to predict bounding boxes, as well as target probabilities[11]. The convolution output of YOLO is 7×7 points, rather than a single point, so each point can label a corresponding region in the original picture. The unique benefit of YOLO is that each point can converge based on the corresponding region target while training, thus avoiding background interference[10]. SSD is one representative of one-stage algorithms, its basic principle is to discretize the bounding box into the default bounding box in output space, then predict and adjust in the default bounding box[13]. During the predicting process, the network will give each object a score based on the object's existing situation, so as to well match the object's shape[13]. In addition, SSD has the unique character of multiscale feature mapping[13].

Although there already exist some vision recognition models that can distinguish background interference, camera-based recognition is extremely computation consuming compared with Lidar, which is owing to Lidar having abundant depth information. Instead of merely applying object-detection models based on a camera, Yandún Narváez et al.[14] got rid of ground interferences by the fusion of Lidar, hence they detected fruit trees more easily. Both 3D and 2D Lidar can be fused to cameras, but compared with 3D Lidar, 2D Lidar is more cost-effective. Generally, 3D Lidar is usually used in autonomous vehicles[15], and 2D Lidar may be enough in agricultural machinery according to specific requirements. Since mechanized planting is widely applied at present, the inter-row spacing and intra-row spacing are uniform, thus 2D Lidar can detect crop positions through appropriate mounting adjustment[16]. In the fusion process, time synchronization between two sensors is really important[17]. What’s more, since the camera and Lidar have their own specific coordinate systems, the coordinate transformation is also a necessary procedure in data fusion. Usually, one world coordinate system is needed, it will be used to act as a benchmark[1]. The world coordinate system can be either the camera or Lidar coordinate system, or a system independent from them. The prevailing operation of fusing camera and Lidar is to project points to an image, then detect whether these points belong to a specific target[18]. But how to find the pixels of target objects in images accurately, then fused these pixels and Lidar points is still a longstanding challenge.

Crop detection is a popular research topic in intelligent agricultural machinery[19,20], crop detection pursues high-quality recognition and position in a certain scenario. For example, intelligent agricultural machinery needs to acquire crop positions, so as to reach or avoid them. The intelligent agricultural machinery needs to recognize a target, then the target can be simplified to a mass particle or a vertical line. Therefore, the crop recognition model needs to provide a bounding box for the target crop, then only the Lidar points that belong to the bounding box will be taken into consideration. Although the bounding box is a limited area, there still covers numerous points, so point cluster operation is also a necessary procedure[21]. According to different application scenarios, researchers have developed a large number of clustering algorithms for point clouds[19]. It is a truth that complex does not mean good, if the application scenarios are agricultural fields, it always looks forward to a clustering algorithm with more accurate positions and better effect. No matter whether staple crops or any kind of vegetation, they extend their leaves while growing, this phenomenon amplifies the actual bounding box area. So how to define the center of clustering points is related to the accuracy of crop positioning.

It is no doubt that weeding control is an essential procedure in agricultural production. The coercion effect of weeds on maize is particularly obvious in the early stages, which contain the elongation stage[22]. Thus the weeding control studies at maize early growth stages are particularly important. Since crop positions are more uniform than weeds, hence only if weeding executing parts can avoid crop seedlings, then executing weeding operations in the other areas, weeding control operations can be implemented successfully[23-25]. So in real agricultural production scenarios, obtaining the positions of crop seedlings is wiser than obtaining the positions of weeds. The above literatures illustrated that the fusion of camera and Lidar has so many advantages, but the relative researches aimed at maize were still limited, most of which is mainly laboratory research. The maize planting area is about 4×10^7 hm^2 in China[26], detecting maize seedlings is a prerequisite for realizing mechanical weeding. So much chemical herbicide can be saved if mechanical weeding can replace herbicide weeding in the near future, so the goal of this research was to detect maize (*Zea mays* L.) seedling’s position through the fusion of a monocular camera and a 2D Lidar, a maize seedling recognition model was going to be achieved by deep learning technology, point clouds which within the target-bounding-box would provide accurate spatial information for the recognized maize seedling.

2 Materials and methods

2.1 Maize seedling recognition model

In order to provide training images for deep learning, training-used images were acquired in Lishu County, Jilin Province, China (43.31°N, 124.62°E) on May 15th, 2021. The maize’s inter-row distance was 65 cm and with different intra-row distances. Its species was Jidan 209, which is a prevailing maize species in Northeast China. This field was sown by a no-till planter, through adjusting the handle which controls the intra-row distance in the no-till planter, different planting densities could be obtained. Specifically, the intra-row distances of 30.7, 25.6, and 21.9 cm are corresponding to 50 000, 60 000, and 70 000 plants/hm^2, respectively. Finally, this field was divided into three planting densities equally. Five shooting periods were selected, and the specific environmental conditions at that time are listed in Table 1. One charge-coupled device camera (Model MV-EM200C, Shanxi Vision Manufacturing Technology Co., Ltd., China) and one fixed focus lens (Model AFT-2514MP, its focal distance is 2/3″, Shanxi Vision Manufacturing Technology Co., Ltd., China) were applied for image acquisition. Color temperature was marked by using color temperature equipment (Model HPCS-320, Hangzhou HONGPU light color technology Co., Ltd., China) at the same time. One pole of 1 m was inserted into the field vertically, and 0.7 m was above the field. In order to measure the solar altitude, a virtual line formed by the endpoint of the pole and the endpoint of the pole shadow was created, the angle between the line and the vertical direction was the solar altitude. By measuring the angle between the pole shadow and the north direction, the solar azimuth was obtained. Solar altitude and solar azimuth were measured every 15 min during the image acquisition periods. Maize seedling was during the elongation stage, the heights were about 40 cm, and six to eight leaves could be found during the image acquisition period. During shooting, in order to get maize characteristics as much as possible, and tried to not overlap the leaves in the image, the distance from the lens to central maize rows was set to about 15 cm, and the main optical axis was set −45° to the horizontal.

Table 1 Mean ambient conditions while acquiring training-used images

Time section	Mean solar altitude/(°)	Mean solar azimuth/(°)	Light color temperature/K
07:00-08:30	29	101	5500
10:00-11:30	56	154	6500
12:00-13:30	59	195	6500
15:00-16:30	33	252	6300
18:00-19:30	1	309	5500

The camera was set to manually triggered for image acquisition. Three planting densities (i.e. 50 000, 60 000, and 70 000 plants/hm^2) contributed 100 images, respectively. 300 images were acquired during each shooting period, in other words, 1500 images were acquired totally. These acquired images were 1280×800 pixels, their formats were '.jpg'. Then *LabelImage* (Version 1.7, Google Brain, US) was applied to mark individual maize seedlings in these images. During the labeling process, each marked object would generate a corresponding label. Furthermore, foreground maize seedlings nearest to the camera were picked out in this procedure, as shown in Figure 1. Corresponding labels were going to be used for training the recognition model. Subsequently, these marked images were separated into a training set, a testing set, and a verifying set according to the ratio of 8:1:1 randomly. Deep learning framework Tensorflow (Version 1.11.0) was adopted in this step. General information about the hardware and software were, the operating system was Ubuntu 20.04, the memory was 16 GB, the processor was Intel® Core™ i7-7 700 KCPU@ 4.00 GHz×8, the digital image processor (GPU) was NVIDIA GTX 2080Ti. Python (Version 3.6.5, Gudio van Rossum) incorporated with OpenCV (Version 3.4.2, computer vision repository,) was used as the programming language. Only target maize seedlings in the foreground were marked and labeled, so background objects would be avoided during the training process. In addition, the luxuriant leaves which were too stretched were excluded, too.

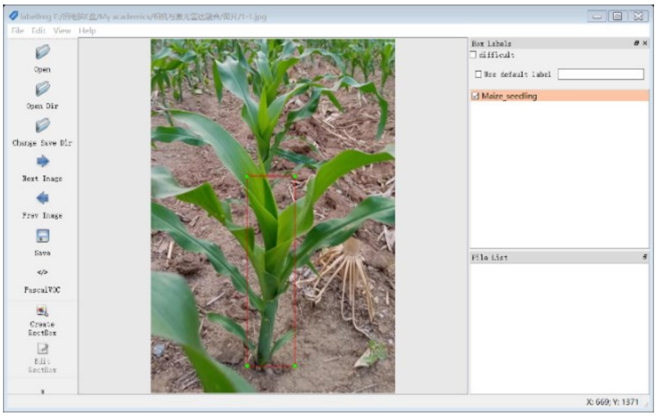
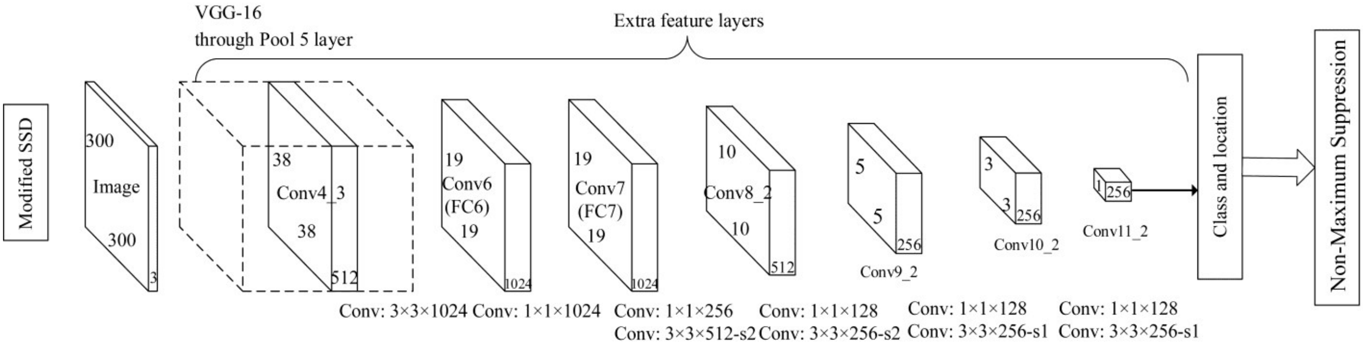


Figure 1 Image labeling aimed at the stem parts of maize seedlings

The classical SSD model was obtained from the feedforward neural network, which is Visual Geometry Group 16 (VGG16). Classical SSD added six extra feature layers, they were Conv4_3, Conv7, Conv8_2, Conv9_2, Conv10_2, Conv11_2, and their sizes decreased gradually. The reason for linking six feature maps to the final output layer was to realize target detection at different scales[27]. The top larger feature maps are more easily to detect small objects, while the rear smaller feature maps are more conducive to detecting larger objects. This study was detecting maize seedlings in real production, so the camera and Lidar would shoot maize seedlings on purpose, which resulted that maize seedlings would be the largest objects compared with the other objects (e.g., weeds) in the images. For this reason, our research modified the classical SSD model. Specifically, the top five links were abandoned, and only the last link to the final output layer was retained, as shown in Figure 2. The other configurations of the modified SSD network were the same as described in Liu et al.[27]



Note: SSD: Single shot multi-box detector; Conv: Convolution; VGG: Visual Geometry Group Net; FC: Fully connected layers.

Figure 2 Network of modified SSD used in this study

2.2 Fusion of camera and Lidar

The same camera and fixed focus lens mentioned above were applied in this procedure. One 360° laser scanning ranging Lidar (Model A2M8, Shanghai Silan Technology Co., Ltd., China) was fused to the camera. The Lidar's pitch angle ranges from −1.5° to 1.5°, its scanning distance ranges from 0.2 to 12 m, and its ranging resolution is less than 0.5 mm. The camera and Lidar were up and down installed, the Lidar was 40 cm above the field surface and 5 cm below the camera, and the camera's main optical axis was set −45° to the horizontal, as shown in Figure 3.

Synchronization of the camera and Lidar should be conducted so as to achieve successful fusion. The synchronization included both time dimension and spatial dimension. Owing to the camera frame rate is 40 fps, which is equal to 40 Hz, while the Lidar scanning rate is only 10 Hz. So this research regarded the Lidar scanning rate as a benchmark, the sampling rates of both the camera and the Lidar were set to 10 Hz. In order to implement one-to-one correspondence, time synchronization was implemented based on their time stamps. The main project was mapping Lidar data to image coordinates, so as to obtain spatial information for target objects in images. The Lidar coordinate and image coordinate were transformed according to Equation (1). In the transformation, the original image coordinate matrix and Lidar coordinate matrix should undergo homogeneous transformation.

$$q_i=K[RT]p_i \tag{1}$$

where, q_i represents the image coordinate of the ith Lidar data, $q_i=[u_i, v_i, 1]^T$; u_i and v_i represent the horizontal and vertical coordinate values, respectively; K represents the intrinsic matrix of the camera, which is given by the camera manual; R represents rotation matrix used for converting Lidar coordinate to image coordinate; T represents translation matrix used for converting Lidar coordinate to image coordinate; p_i represents the Lidar coordinate of the ith Lidar data, $p_i=[x_i, y_i, z_i, 1]^T$; x_i, y_i, and z_i represent the coordinate values in three dimensions, respectively.

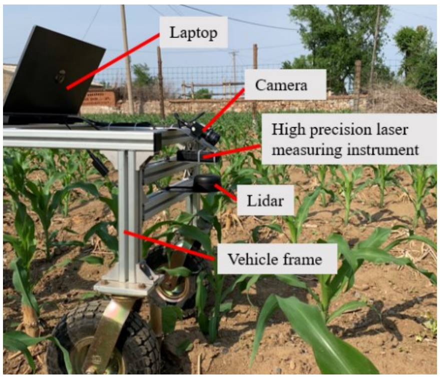


Figure 3 Positional relationships among the major components

In this study, the Lidar coordinate was also set as the world coordinate, which was a virtual right-hand Cartesian coordinate system, as shown in Figure 4. The origin was set to the geometry center of the Lidar, and the positive direction of the x-axis was set as the reverse direction of the Lidar junction cable. Such a setting was accordant with the definition of Lidar data scanning coordinate. According to Zhang[28], this research obtained the camera intrinsic matrix. The converting matrix from the Lidar coordinate to the camera coordinate was obtained with the aid of Lidar reflection cone. The specific calculation steps were as follows:

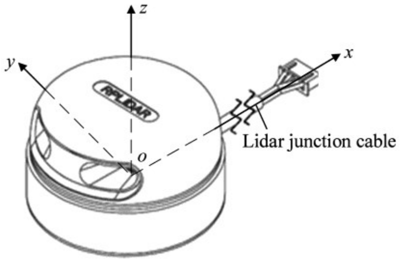


Figure 4 Schematic diagram of the virtual world coordinate system

Firstly, put the Lidar horizontally, and adjust the angle and height of the reflection cone, which aim was to maximize the radar cross section (RCS) with respect to a corresponding detection target. With the above premises, the center of the reflection cone was equivalent to the Lidar coordinate origin. That was to say, the vertical coordinate of z_i=0 in the matrix of $p_i=[x_i, y_i, z_i, 1]^T$, at this time, x_i, y_i were recorded.

Secondly, a corresponding image was shot manually, then the center coordinate of the reflection cone was labeled manually;

Thirdly, step one and step two were repeated until 50 group numbers were obtained.

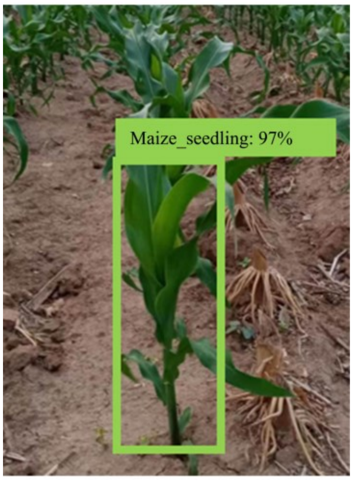
Fourthly, 25 groups of data were selected randomly, and Equation (2) was used as the objective function to solve a nonlinear optimization problem, which was to get the transformation matrix parameters from Lidar coordinate to the camera coordinate.

$$\arg\min \sum_{i=1}^{n} \| q_i - K[RT]p_i \|^2 \quad n \in [1, 25] \tag{2}$$

According to Budil et al.[29], the above solution was obtained.

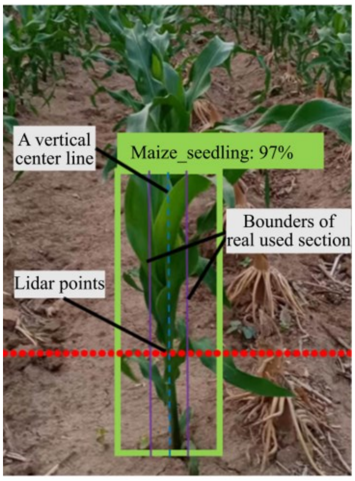
Fifthly, the remaining 25 groups of data were used to verify the effect of conversion matrix parameters, if average error requirements were within the threshold, the conversion matrix parameters would be retained. The threshold was described by Zhang et al.[30]

In light of the maize seedling recognition model, a bounding box would surround a target maize seedling, as shown in Figure 5. As for Lidar points, they would project to the image. This study was going to put the emphasis on the points located within the bounding box, rather than all the points. Moreover, since the bounding box exceeded the maize seedling's real outer contour, so points within a narrower area were going to be taken into account. These points were in the inner rectangle of Figure 5. Specifically,



Note: Only the maize seedling that was close to the data acquisition device was regarded as the target maize seedling, maize seedlings located in the other locations were regarded as background objects. The same as below.

Figure 5 Maize seedling which was recognized successfully



Note: The blue dotted line is a vertical center line of the bounding box, the two purple lines are used to show the bounders of the real used section, and the red points are Lidar scanning points.

Figure 6 Schematic diagram of extracting effective Lidar points for clustering

the corner coordinates of the bounding box could be got according to the coordinates of the top left corner and bottom right corner, and a vertical center line of the bounding box could be gotten, taking this center line as a benchmark, this research utilized one-third width of the original bounding box in the width dimension, then the points which located within this width were going to be clustered, as shown in Figure 6.

In the following step, a cluster center was going to be calculated. Specifically, these points' spatial coordinates would be averaged and assigned to a new center point. The new center point would be regarded as the cluster center, then the spatial information for a specific maize seedling can be derived based on the coordinate of this new center point. As for the fusion of the camera and Lidar at the implementation level, it was based on Robot Operating System (ROS, the specific version was Noetic).

2.3 Experiments

On May 20th, 2021 which was 5 d after training-used image acquisition, a maize seedling detection experiment was taken place in the same field but with brand new maize rows. This experiment lasted 3 d and corresponded to three different plant densities. The maize was still during the elongation stage. There were still five shooting periods each day, the specific environmental conditions of three experimental days (from May 20th to May 22nd) were listed in Table 2.

Table 2 Mean ambient conditions while data acquisition

Time section	Mean solar altitude/(°)	Mean solar azimuth/(°)	Light color temperature/K
07:00-08:30	28	100	5400
10:00-11:30	55	156	6500
12:00-13:30	57	194	6500
15:00-16:30	32	251	6200
18:00-19:30	0	307	5800

The camera-Lidar combined device (Figure 7) was manually propelling along the maize rows, and the operations were the same with training-used image acquisition if there was no more specific explanation. For the sake of getting real spatial information on maize seedlings, one high precision laser measuring instrument ('HLM' is short for 'high precision laser measuring instrument' in the following, Model MiLESEEY-X5, Shenzhen MAITEST Technology Co., Ltd., China, its measuring resolution reaches 1 mm) was mounted close to the camera, as shown in Figure 3, then this study could obtain real distances and angles by the HLM. In light of HLM's geometry dimension and basic parameters, the measuring benchmark could be converted to the world coordinate, which was also the Lidar coordinate. According to the configuration in Figure 3, and only considering the stem parts, the HLM detecting point was in the middle of maize seedlings, besides, the detecting point was also on the Lidar scanning plane.

According to the above interpretations, the camera and Lidar were set to a homogeneous data acquisition frequency of 10 Hz, and they had identical starting times by using time stamps. As for any specific maize seedling, a random approach position was selected as shown in Figure 7, and the data acquisition lasted about 1 s. During this period, about 10 images were acquired for each maize seedling. At the same time, the HLM recorded the corresponding distance and angle information. What was needed to emphasize is that after fine-tuning manually, the HLM measuring plane was parallel to the Lidar scanning plane, and its measuring points were in the right middle of the target maize stems. About 100 maize seedlings were measured during each acquisition. There were three replicates for each shooting period, so 15 acquisitions occurred and about 1500 maize seedlings were detected for 1 d experiment.

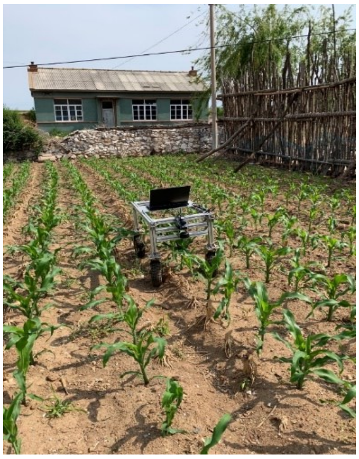


Figure 7 Experimental scenario with the fused device of camera and Lidar

As for the modified SSD model, for the sake of figuring out the real influence of abandoning the top five links from six extra feature layers, this research conducted a series of comparative experiments by substituting the modified SSD model with classical SSD, YOLOv1 and Faster R-CNN. This study was also going to evaluate the detecting performance from the following aspects, i.e., the maize seedling recognition accuracy, the distance detecting accuracy, and the angle detecting accuracy based on the world coordinate. The automatically obtained angle was measured by the fused method proposed in this study. There are three angles between a line and a plane, and the line was formed by the new cluster center and the origin of the world coordinate system. The automatically obtained angle based on '*xoy*' plane was the angle between the line and the '*xoy*' plane, and so did the automatically obtained angles based on the other two planes.

The HLM-obtained angle was regarded as a comparative standard. Specifically, the HLM obtained angle based on '*xoy*' plane was the angle between '*xoy*' plane and laser beam emitted from the HLM, through manual fine-tuning, the laser beam would reach the width center of a maize seedling. So did the HLM-obtained angles based on the other two planes. Individual maize seedling detecting accuracy was calculated according to Equation (3).

$$\alpha_j = \left|\frac{\gamma_j - \beta_j}{\beta_j}\right| \times 100\% \tag{3}$$

where, α_j represents the final data, %; γ_j represents the automatically obtained data; β_j represents the standard data; j represents different statistical items, i.e., maize seedling recognition accuracy, distance accuracy, angle accuracy based on '*xoy*' plane, angle accuracy based on '*yoz*' plane and angle accuracy based on '*xoz*' plane. As for maize seedling recognition accuracy, γ_j represents the maize seedling numbers that are recognized successfully by the camera; β_j represents the manually counted seedling numbers. As for distance accuracy as well as the angle accuracies based on '*xoy*' plane, '*yoz*' plane, and '*xoz*' plane, they were calculated within these successfully recognized maize seedlings.

2.4 Data analyses

SPSS 22.0 for Windows (SPSS Inc., US) was used for statistical analyses. The least significant difference (LSD) analysis, including Student's *t*-test and *F*-test, was used to compare whether there existed significant differences among different data acquisition periods or different plant densities, etc. (*p<0.05, **p<0.01)

3 Results

Figure 4 reflects that the modified SSD model reached a recognition accuracy of more than 92% in the training procedure. As for field experimental results, Table 3 demonstrates that the recognition accuracy was affected by planting densities, as for planting densities of 50 000, and 60 000 plants/hm^2, their recognition accuracies were all above 90%, no matter when were the data acquisition periods. However, with regards to the planting density of 70 000 plants/hm^2, four recognition accuracies were lower than 90%. Generally speaking, the recognition accuracy decreased along with the increase in planting density. But in the premise of identical planting densities, the recognition accuracies nearly did not exist significant (p>0.05) difference among different data acquisition periods.

Table 4 lists that the modified SSD model did not show a significant (p>0.05) advantage compared with the classical SSD model. But both the modified and classical SSD models exceeded the other two models significantly (p<0.05). The recognition accuracies of YOLOv1 decreased significantly (p<0.01) from 60 000 to 70 000 plants/hm^2. However, unlike the others, the recognition accuracies of Faster R-CNN kept stability regardless of planting density.

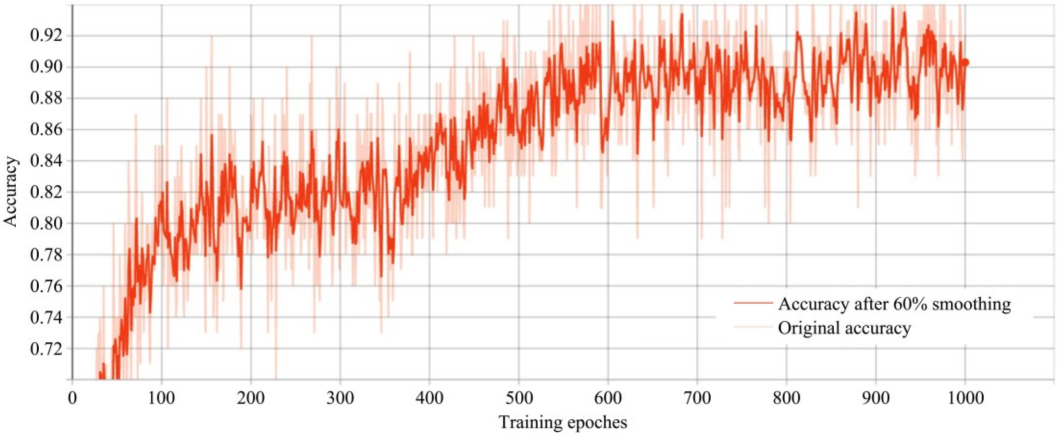


Figure 4 Recognition accuracy curve during training procedure for modified SSD model

Table 3 Mean maize seedling recognition accuracies in different time sections

Planting density /plants·hm^{-2}	Time section/%				
	07:00-08:30	10:00-11:30	12:00-13:30	15:00-16:30	18:00-19:30
50 000	91.3^a	92.3^a	91.5^a	92.6^a	92.2^a
60 000	90.4^b	90.3^b	91.2^b	91.1^b	90.2^b
70 000	89.2^c	89.9^c	89.7^c	89.4^c	90.5^c

Note: Different lowercase letters representing the designated parameters had significant differences (p<0.05) compared with the others. The same as below.

Table 4 Mean maize seedling recognition accuracies among different deep learning models

Planting density /plants·hm^{-2}	Modified SSD/%	SSD/%	YOLOv1/%	Faster R-CNN/%
50 000	91.98^a	91.30^a	81.50^b	89.60^c
60 000	90.64^d	90.63^d	81.20^b	89.10^c
70 000	89.75^e	89.59^e	75.70^f	89.40^c

Table 5 exhibits different time consumptions with regard to different models, the modified SSD shows significant (p<0.05) time advantage compared with the others. Compared with three commonly used models which are classical SSD, YOLOv1, and Faster R-CNN, this modified SSD saved about 10, 30, and 130 ms/image, respectively. As for any specific model, there did not show any time consumption difference (p>0.05) among different planting densities.

Table 5 Mean time consumptions among different deep learning models (ms/image)

Planting density/plants·hm^{-2}	Modified SSD	SSD	YOLOv1	Faster R-CNN
50 000	61.3^a	72.3^b	82.5^c	192.6^d
60 000	60.4^a	70.3^b	82.2^c	191.1^d
70 000	59.2^a	69.9^b	83.7^c	189.4^d

In order to evaluate the measured distance accuracy and angle accuracy, these standard data were obtained by using HLM mentioned above. Table 6 reveals that averaged across data acquisition periods, and with regards to the planting density of 50 000 plants/hm^2, the average distance error (d), average angle errors based on '*xoy*' plane (x), '*yoz*' plane (y), and '*xoz*' plane (z) were 0.88%, 0.76%, 0.80%, and 0.78% respectively. They were 1.38%, 1.44%, 0.88%, and 1.18% with respect to the planting density of 60 000 plants/hm^2. And in terms of the planting density of 70 000 plants/ hm^2, they were 1.36%, 0.62%, 1.02%, and 1.08%, respectively. Focused on planting densities and data acquisition periods, further statistical analyses showed that they did

Table 6 Mean positional errors of different data acquisition periods and different planting densities

Planting density /plants·hm^{-2}		Data acquisition period/%					Mean standard deviation
		07:00-08:30	10:00-11:30	12:00-13:30	15:00-16:30	18:00-19:30	
50 000	d	1.1	0.9	0.8	0.7	0.9	1.1 cm
	x	1.1	0.5	0.6	1.2	0.4	0.71°
	y	0.5	0.8	0.9	1.1	0.7	0.45°
	z	0.2	0.1	1.9	1.0	0.7	0.9°
60 000	d	1.8	1.6	0.1	1.2	2.2	0.9 cm
	x	1.0	1.0	1.2	2.1	1.9	0.68°
	y	0.8	0.5	0.9	0.7	1.5	0.78°
	z	1.5	1.1	1.1	1.3	0.9	0.88°
70 000	d	1.0	2.0	0.9	1.1	1.8	1.4 cm
	x	0.2	0.6	0.6	0.8	0.9	0.91°
	y	1.1	1.2	1.3	1.1	0.4	1.1°
	z	2.0	1.5	0.6	0.4	0.9	0.31°

Note: lower letter 'd' represents average distance error, lower letter 'x' represents average angle error based on '*xoy*' plane, lower letter 'y' represents average angle error based on '*yoz*' plane, the lower letter 'z' represents the average angle error based on '*xoz*' plane.

not cause significant (p>0.05) differences in the above-mentioned errors. Furthermore, statistical analyses showed that the errors were in a normal distribution, and the mathematical expectation was close to 0. Besides, the real spatial parameters and detecting spatial parameters did not have significant differences at a significance level of 0.05.

4 Discussion

4.1 Maize seedling recognition accuracy

As listed in Table 3, the recognition accuracy of maize seedlings was around 90%, and the accuracy decreased along with the increase in planting density. Object classification and object recognition have different levels of complexity. Object classification always operates images that have only one kind of object, and object recognition always operates images that have several kinds of objects, besides, it always needs to label different objects in the operation of object recognition, so object recognition is more complex than object classification[31]. The fewer objects in the image, the easier for successful detection. It is a truth that images acquired from high-density fields are more complex than those acquired from low-density fields. For example, there were more overlapping leaves in high-density field images, as shown in Figure 9. Substantial overlapping leaves caused more trouble for the maize-seedling recognition model.

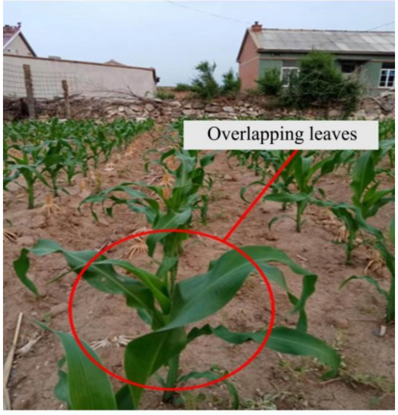


Figure 9 Overlapping leaves of elongation maize seedlings under 70 000 plants/hm^2

In the research field of computer vision, light does have a close relationship with the image process. But to our surprise, statistical analyses showed that image acquisition periods did not have significant influence (p>0.05) on the recognition accuracy. Table 2 reveals that different image acquisition periods had a close relationship with light conditions, such as solar altitude, solar azimuth, and light color temperature. The little light influence should ascribe to the modified SSD model in this research since different extracting feature maps have different unique functions, i.e., the relative lower maps have relative smaller receptive fields, which have unique advantages of detecting relative small objects[32]. The relative higher feature maps have relative larger receptive fields, which have the unique advantage of detecting relative larger objects[32]. This research put emphasis on detecting maize seedlings, and we would shoot maize seedlings on purpose while working, so maize seedlings would be the largest objects in acquired images. This is why we modified the classical SSD model, so as to let only the last feature map (the highest feature map) be linked to the final output layer. All in all, the application-orientated purpose of this study, together with the modified linkage made the light influence no longer obvious while aiming at specific research objects. Similar research results can be found in the study of Jia et al.[33], their purpose was to detect maize ears, and maize ears were shot on purpose while working, so even though illumination conditions changed all the time, their detecting accuracies did not have obvious variations.

Although in terms of recognition accuracy, this modified SSD model did not have obvious advantages over the classical SSD model, this modified SSD model had obvious advantages compared with YOLOv1 and Faster R-CNN (Table 4). This should ascribe to the unique advantage of SSD model, which added six extra feature layers and predicted by convolutional kernels[27]. The unique characteristics of SSD model were also reflected in time consumption, Table 5 illustrates that the time consumption of the modified SSD model was about 60 ms/image, which saved a lot of time compared with the others. On the one hand, this was because the SSD model is a one-stage strategy model, while Faster R-CNN is a two-stage strategy model[7], so it is no surprise that SSD model has time consumption advantage over Faster R-CNN. Some other researches also showed that SSD model was faster than YOLOv1[11]. On the other hand, this research abandoned the top five links compared with the original network architecture, owing to only one link being retained in the final output layer, this should save substantial time compared with the classical SSD model. Usually, one image contains about one effective maize seedling, namely, one maize seedling is exposed appropriately and can be detected successfully. Assuming that the maize intra-row distance is 25 cm on average, thus a time consumption of 60 ms/image could guarantee a non-stop operating speed of 15 km/h. Such speed could catch up with any kind of modern agricultural machinery.

4.2 Maize seedling detecting accuracy

Experimental results showed that camera and Lidar fused well enough to give a detecting result with a maximum error of less than 1.38%. As for the maximum mean standard deviations for distance and angle, they were 1.4 cm and 1.1°, respectively. Furthermore, the errors were in a normal distribution, and the mathematical expectation was close to 0. According to the real growth scenario and agronomy in northeast China, the average diameter of maize seedlings is about 1.5 cm, the commonly used inter-row distance is 65 cm[34]. As for the distance-detecting accuracy in this research, the variation range was equivalent to a maize seedling's diameter. As for the angle detecting accuracy, the standard deviation was 1.1°. Assuming that the weeding execution part is 65 cm long and starts from the origin of the world coordinate system, besides, the weeding execution part can reach the target point without significant deviation. The final result is that such deviation will merely cause about 1.25 cm offset for the weeding execution part. Under the current technical requirements, the above errors can be tolerated, owing to mechanical weeding will retain a circular safety zone, this zone is centered on maize seedlings, which is about one-fifth of the intra-row distance, namely, a circular zone with a diameter of about 5 cm in maize field[35]. More specifically, with regard to the current distance standard deviation and angle standard deviation, our research results will not cause seedling damage in real operation.

Unaffected by light conditions is the unique advantage of Lidar. Table 6 demonstrates that the detecting accuracy did not change significantly (p>0.05) under any illumination circumstance. Owing to being unaffected by light, Lidar is usually used for outdoor object detection, such as agricultural crop detection and fruit tree detection[1,14,16]. If taken all the Lidar points within the bounding boxes into consideration, this research could not reach such detecting accuracy. The reason is that owing to two-dimension laser Lidar only emits one beam in height dimension, but many beams in the width dimension. As for image recognition, it was a bounding box that surrounded a maize seedling. The bounding box is larger than the real object, especially in the width dimension in this research. Despite some points did fall into the bounding boxes, actually, they did not belong to the real target object, but belonged to background objects. In order to avoid such phenomenon, this research narrowed the real used areas of the bounding boxes, which retained only one-third in the width dimension. And one-third of Lidar points in bounding boxes were clustered, then a new center coordinate that represented the cluster was used for spatial analyses.

As for the actual composition of laser points, some of them were reflected by maize stems, while some of them were reflected by maize leaves, and even some laser points were reflected by background maize seedlings. Although we narrowed the bounding box to one-third, this area still exceeded the actual width of the maize stems. Actually, the maize stems only occupied a relative small area, as shown in Figure 6. So it is inevitable that some laser beams reached background, and background reflected points were a major reason for error occurrence. From another aspect, some maize seedlings were not in the right middle of the bounding boxes. Under these circumstances, a majority of effective points were reflected by maize leaves. If stem reflected points occupied a relative small proportion, while leave reflected points occupied a relative larger proportion. However, the real positional detection was based on maize stems in this study, such leave reflected points made the calculated position deviate from the real position, too. Besides, although there did not exist visible positional changes between the camera and the Lidar, their positional variations were inevitable, some studies hold the viewpoint that vibration-caused positional changes between the camera and the Lidar is also a non-negligible reason for error occurrence[36].

5 Conclusions

This research achieved maize seedling detections by the fusion of camera and Lidar, which took full advantage of camera and Lidar. Deep learning played an important role in maize seedling recognition based on the hardware of camera, and Lidar points provided accurate spatial information regardless of ambient light conditions. The experiment was conducted in a maize field, and the maize was during the elongation stage. Experimental results revealed that maize seedling recognition accuracies had a relationship with planting densities, but with a premise of identical planting densities, the recognition accuracy was stable. No matter what were the planting densities and data acquisition periods, our detected positional errors did not have significant differences at a significance level of 0.05. Furthermore, the errors were in normal distribution, and the mathematical expectation was close to 0. Therefore, although errors exist, this technology can guide an executing part to reach or avoid maize seedlings without damaging them. We are sure this study does have some practical application value in intelligent agricultural machinery.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (Grant No. 31901408) and the Science and Technology Development Plan of Jilin Province (Grant No. 20200201206JC).

[References]

[1] Wang Y X, Xu S S, Li W B, Kang F F, Zheng Y J. Identification and location of grapevine sucker based on information fusion of 2D laser scanner and machine vision. Int J Agric & Biol Eng, 2017; 10(2): 84–93.

[2] Zhang R Y, Cao S Y. Extending reliability of mmWave radar tracking and detection via fusion with camera. IEEE Access, 2019; 7: 137065–137079.

[3] Gonzalez R C, Woods R E. Digital image processing. 3rd. New Jersey: Prentice Hall, 2008; 976p.

[4] Kamilaris A, Prenafeta-Boldu F X. Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 2018; 147: 70–90.

[5] Lecun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015; 521(7553): 436–444.

[6] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH: IEEE, 2014; pp.580–587.

[7] Girshick R. Fast R-CNN. 2015 IEEE International Conference on Computer Vision and Pattern Reconition (CVPR), Santiago: IEEE, 2015; pp.1440–1448.

[8] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and machine intelligence, 2016; 39(6): 1137–1149.

[9] Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv, 2018; arXiv 1804.02767.

[10] Redmon J, Farhadi A, YOLO9000: Better, Faster, Stronger. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu: IEEE, 2017; pp.6517–6525. doi: 10.1109/CVPR.2017.690.

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

[12] Uijlings J R R, Van De Sande K E A, Gevers T, Smeulders A W M. Selective search for object recognition. International Journal of Computer Vision, 2013; 104(2): 154–171.

[13] Zhao H, Li Z, Zhang T. Attention based single shot multibox detector. Journal of Electronics & Information Technology, 2021; 43(7): 2096–2104.

[14] Yandún Narváez F J, Salvo Del Pedregal J, Prieto P A, Torres-Torriti M, Auat Cheein F A. Lidar and thermal images fusion for ground-based 3D characterisation of fruit trees. Biosystems Engineering, 2016; 151: 479–494.

[15] Zhao X, Sun P, Xu Z, Min H, Yu H. Fusion of 3D Lidar and camera data for object detection in autonomous vehicle applications. IEEE Sensors Journal, 2020; 20(9): 4901–4913.

[16] Xue J, Fan B, Yan J, Dong S, Ding Q. Trunk detection based on laser radar and vision data fusion. Int J Agric & Biol Eng, 2018, 11(6): 20–26.

[17] Ji Y H, Li S C, Peng C, Xu H Z, Cao R Y, Zhang M. Obstacle detection and recognition in farmland based on fusion point cloud data. Computers and Electronics in Agriculture, 2021, 189: 106409. doi: 10.1016/j.compag.2021.106409.

[18] Underwood J P, Hung C, Whelan B, Sukkarieh S. Mapping almond orchard canopy volume, flowers, fruit and yield using lidar and vision sensors. Computers and Electronics in Agriculture, 2016; 130: 83–96.

[19] Ma Z H, Tao Z Y, Du X Q, Yu Y X, Wu C Y. Automatic detection of crop root rows in paddy fields based on straight-line clustering algorithm and supervised learning method. Biosystems Engineering, 2021; 211: 63–76.

[20] Raja R, Nguyen T T, Slaughter D C, Fennimore S A. Real-time robotic weed knife control system for tomato and lettuce based on geometric appearance of plant labels. Biosystems Engineering, 2020; 194: 152–164.

[21] Schirrmann M, Hamdorf A, Garz A, Ustyuzhanin A, Dammer K-H. Estimating wheat biomass by combining image clustering with crop height. Computers and Electronics in Agriculture, 2016; 121: 374–384.

[22] Tharp B E, Kells J J, Bauman T T, Harvey R G, Johnson W G, Loux M M, et al. Assessment of weed control strategies for corn in the north-central united states. Weed Technology, 2004; 18(2): 203–210.

[23] Cordill C, Grift T E. Design and testing of an intra-row mechanical weeding machine for corn. Biosystems Engineering, 2011; 110(3): 247–252.

[24] Raja R, Thuy T N, Vuong V L, Slaughter D C, Fennimore S A. Rtd-seps: Real-time detection of stem emerging points and classification of crop-weed for robotic weed control in producing tomato. Biosystems Engineering, 2020; 195: 152–171.

[25] Jia H, Gu B, Ma Z, Liu H, Wang G, Li M, et al. Optimized design and

experiment of spiral-type intra-row weeding actuator for maize (*Zea mays* L.) planting. Int J Agric & Biol Eng, 2021; 14(6): 54–60.

[26] National Bureau of Statistics of China. China statistical yearbook. Beijing: China Statistics Press, 2021; 945p. (in Chinese)

[27] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, et al. SSD: Single shot multibox detector. Computer Vision – ECCV 2016, Springer, 2016; 9905: 21–37.

[28] Zhang Z Y. Flexible camera calibration by viewing a plane from unknown orientations. Proceedings of the Seventh IEEE International Conference on Computer Vision, Redmond: IEEE, 1999; pp.666–673. doi: 10.1109/ICCV.1999.791289.

[29] Budil D E, Lee S, Saxena S, Freed J H. Nonlinear-least-squares analysis of slow-motion EPR spectra in one and two dimensions using a modified levenberg–marquardt algorithm. Journal of Magnetic Resonance Series A, 1996; 120(2): 155–189.

[30] Zhang Y, Pan S Q, Xie Y S, Chen K, Mo J Q. Detection of ridge in front of agricultural machinery by fusion of camera and millimeter wave radar. Transactions of the CSAE, 2021; 37(15): 169–178. (in Chinese)

[31] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 2015; 115(3): 211–252.

[32] Shen C, Zhao X M, Fan X, Lian X, Liu Z, Zhang F, Kreidieh A R, et al. Multi-receptive field graph convolutional neural networks for pedestrian detection. IET Intelligent Transport Systems, 2019; 13(9): 1319–1328.

[33] Jia H L, Qu M H, Wang G, Walsh M J, Yao J R, Guo H, et al. Dough-stage maize (*Zea mays* L.) ear recognition based on multiscale hierarchical features and multifeature fusion. Mathematical Problems in Engineering, 2020; 2020: 9825472. doi: 10.1155/2020/9825472.

[34] Wang G, Jia H L, Zhao J L, Li C Y, Wang Y, Guo H. Design of corn high-stubble cutter and experiments of stubble retaining effects. Transactions of the CSAE, 2014, 30(23): 43–49. (in Chinese)

[35] Jia H L, Li S S, Wang G, Liu H L. Design and experiment of seedling avoidable weeding control device for intertillage maize (*Zea mays* L.). Transactions of the CSAE, 2018, 34(7): 15–22. (in Chinese)

[36] Fu Y J, Tian D X, Duan X T, Zhou J S, Lang P, Lin C M, et al. A camera–radar fusion method based on edge computing. In: 2020 IEEE International Conference on Edge Computing (EDGE), Beijing: IEEE, 2020; pp.9–14. doi: 10.1109/EDGE50951.2020.00009.