Accurate crop row recognition of maize at the seedling stage using lightweight network

Jian Wei[1,2†], Mengfan Zhang[1,2†], Caicong Wu[1,2], Qin Ma[1,2*], Weitao Wang[1,2], Chuanfeng Wan[1,2]

(1. *College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China*;
2. *Key Laboratory of Agricultural Machinery Monitoring and Big Data Applications, Ministry of Agriculture and Rural Affairs, Beijing 100083, China*)

Abstract: Accurate extraction of crop row is very important for automation of agricultural production. Crop rows are required for accurate machine guidance in agricultural production such as fertilization, plant protection, weeding and harvesting. In this study, an efficient crop row detection algorithm called Crop-BiSeNet V2 was proposed, which combined BiSeNet V2 with a spatial convolutional neural network. The proposed Crop-BiSeNet V2 detected crop rows in color images without the use of threshold and other pre-information such as number of rows. A data set had 2697 maize crop images was constructed in challenging field trial conditions such as variable light, shadows, presence of weeds, and irregular crop shape. The proposed system was experimentally determined to overcome the interference of different complex scenes. And it can be applied to crop rows of different numbers, straight lines and curves. Different analyses were performed to check the robustness of the algorithm. Comparing this algorithm with the Fully Convolutional Networks (FCN) algorithm, it exhibited superior performance and saved 84.85 ms. The accuracy rate reached 0.9811, and the detection speed reached 65.54 ms/frame. The Crop-BiSeNet V2 algorithm proposed in this study show strong generalization performance for seedling crop row recognition. It provides high-reliability technical support for crop row detection research and assists in the study of intelligent field operation machinery navigation.

Keywords: computer vision, crop row detection, precision agriculture, semantic segmentation
DOI: 10.25165/j.ijabe.20241701.7051

Citation: Wei J, Zhang M F, Wu C C, Ma Q, Wang W T, Wan C F. Accurate crop row recognition of maize at the seedling stage using lightweight network. Int J Agric & Biol Eng, 2024; 17(1): 189–198.

1 Introduction

Due to the development of precision agriculture, crop row detection has become an important part of agricultural intelligent equipment[1]. It is widely used in agricultural production, such as assisted navigation, precise fertilization, weeding, spraying and harvesting[2-5]. With the help of crop row detection, people can apply herbicides and fertilizers in a targeted manner, which can further save fertilizer and pesticide, reduce labor intensity and environmental pollution, improve economic benefits.

Crop row detection process usually involves three main steps: image acquisition, image segmentation and crop row detection, and lots of work has been carried out by the researchers in these steps.

In image acquisition and preprocess step, Zheng et al.[2] used unmanned aerial vehicles to obtain RGB, NIR-GB and MS images of rice crops. The overall accuracy of these data sources reached 0.9125, 0.9288, and 0.9353. Another problem is that in the image taken by the front-view camera, due to the perspective effect, things that were originally parallel crossed in the image. In order to further highlight the distant target area, Rabab et al.[6] used Inverse Perspective Mapping (IPM) to realize rice row detection and improve the recognition accuracy. Also, the Region of Interests (RoI) of every image is needed before crop line detection, and there are multiple ways to choose it[7-10]. Chen et al.[11] determined the ROI by using the prior knowledge that the adjacent images of the video will not have mutations.

In terms of crop row feature extraction, to avoid the interference of complex lighting, shadows, weeds and other factors in the field, Rabab et al.[6] used the green features of crops to extract green factors from RGB images. The recognition accuracy of maize, celery, potato, onion, sunflower and soybean rows was 0.84 in the open data set[12]. Su et al.[13] obtained the number of maize row by extracting ultra-green features and Hough transform. However, these models require manual input of threshold parameters according to different weed densities. In order to overcome the influence of different weed densities, plant growth stages and weather conditions, Kanagasingham et al.[14] converted the image into HSV color space and manually selected the threshold to extract green rice plants. At low weed density, the accuracy of heading compensation was less than 2.5°. To achieve detection, the results of image segmentation were used to fit crop rows.

As mentioned before, existing machine vision methods are relatively mature[15-20], but there are also some challenges:

1) Under outdoor conditions, variable lighting conditions, shadows, insufficient lighting and other factors will affect the image quality.

2) Crops may be confused with weeds that are similar in shape, texture and color.

Received date: 2022-09-06 **Accepted date:** 2023-03-05

Biographies: Jian Wei, Postgraduate student, research interest: computer vision and artificial intelligence in agriculture, Email: 764643510@qq.com; **Mengfan Zhang**, Postgraduate student, research interest: computer vision and artificial intelligence in agriculture, Email: 2683972496@qq.com; **Caicong Wu**, PhD, Professor, research interest: agricultural machinery navigation and location service, Email: wucc@cau.edu.cn; **Weitao Wang**, Postgraduate student, research interest: computer vision and artificial intelligence in agriculture, Email: wangweitao1998@cau.edu.cn; **Chuanfeng Wan**, Postgraduate student, research interest: computer vision and artificial intelligence in agriculture, Email: 2473024252@qq.com.

†The authors contributed equally to this work.

***Corresponding author: Qin Ma**, PhD, Professor, research interest: computer vision and artificial intelligence in agriculture. College of Information and Electrical Engineering, China Agricultural University, Beijing 10083, China. Email: sockline@163.com.

3) Different shapes and rules of crops at different growth stages lead to wrong detection result.

Deep learning model can effectively solve problems such as complex lighting and weed interference. Liu et al.[21] established a multi-scale layered feature deep learning network model to identify maize stalks. And the average recognition accuracy was 0.9893 under six weed densities and different light conditions. Zhang et al.[22] extracted the row and center lines of rice seedlings based on the YOLOv3 target detection algorithm. Under different growth stages of seedlings, strong wind, reflection and other special scenes, the average accuracy reached 0.9147. And the processing time of a single image was 82.6 ms. However, models often sacrifice time performance to obtain higher recognition accuracy. Therefore, it is more reasonable to adopt an end-to-end crop row segmentation network. Adhikari et al.[5] constructed a deep convolutional codec network (ESnet) for crop row detection of rice seedlings. Crop rows were extracted from the input image, with an average pixel deviation of 2.89 pixels and detection speed of 10.97 fps.

Common convolutional neural models do not fully explore the spatial location relationship of crop rows. Spatial information is very important for targets with strong priori shapes but discontinuous appearance. By learning spatial information, the deep neural network can predict crop row position and solve the problems of the seedling belt fracture. In order to make full use of spatial structure information of images, Liu et al.[23] and Luo et al.[24] combined Convolutional Neural Network (CNN) and Conditional Random Field (CRF), and used large convolution kernels to transmit spatial information. Pan et al.[25] proposed a new convolution method (Spatial CNN, SCNN), which converted the traditional CNN layer-to-layer connection into a slice-to-slice form. Pixels could transmit information between rows and columns, and the results showed that it had a good recognition effect on long objects, such as lane lines and telephone poles.

To sum up, due to the complex and changeable environment, traditional methods cannot adapt to detection tasks in various scenarios. Therefore, this study proposed a crop row detection method based on convolutional neural network. Different data labeling methods are explored in order to improve the detection effect. In order to achieve a faster detection speed, a network with better real-time performance was selected as the backbone network. The multi-branch structure was designed to overcome the interference between rows of different crops. Images of crop rows containing different numbers, straight lines and curves can be detected correctly. Aiming at the situation of crop rows breakage and few visualization features, the model fully explored the image spatial information features.

2 Materials and methods

As illustrated in Figure 1, crop row detection is a complicated task with multiple steps. The general convolutional neural network had complex structure and many model parameters, which cannot meet the real-time requirements of crop row detection. Therefore, the lightweight network was selected as the backbone network. The backbone network was optimized for a variety of complex scenes and strong spatial information of crop rows. The main research contents are shown in Figure 2.

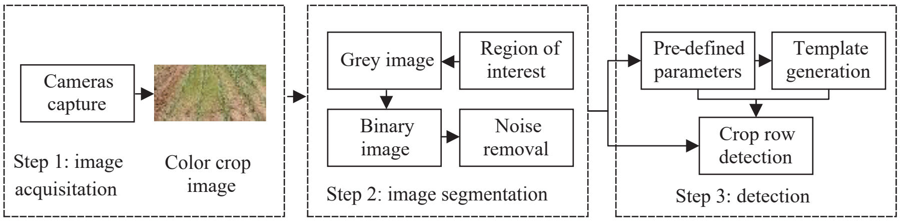


Figure 1 General crop row detection process

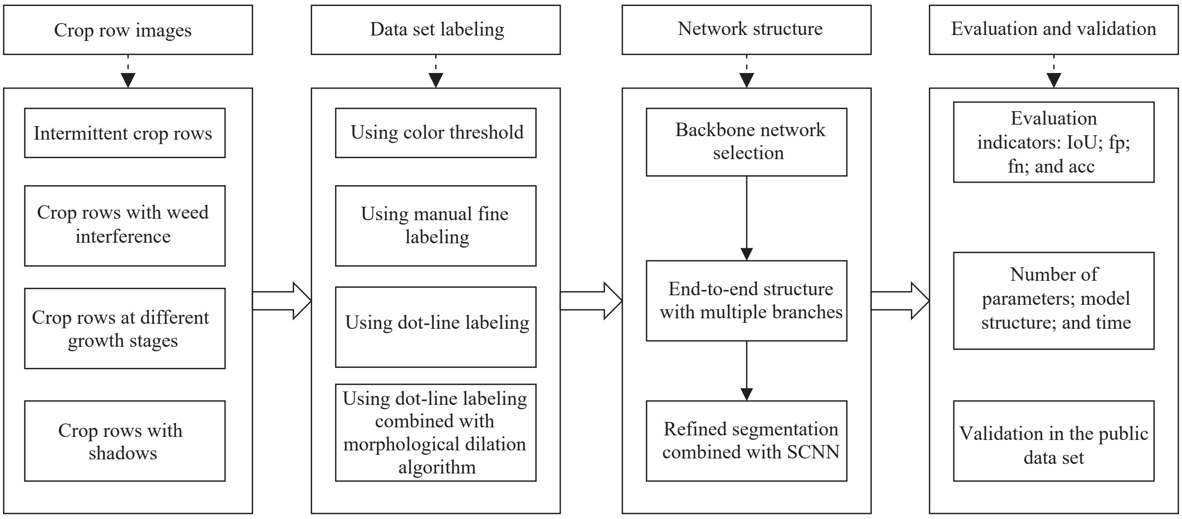

Crop row images
Intermittent crop rows
Crop rows with weed interference
Crop rows at different growth stages
Crop rows with shadows
Data set labeling
Using color threshold
Using manual fine labeling
Using dot-line labeling
Using dot-line labeling combined with morphological dilation algorithm
Network structure
Backbone network selection
End-to-end structure with multiple branches
Refined segmentation combined with SCNN
Evaluation and validation
Evaluation indicators: IoU; fp; fn; and acc
Number of parameters; model structure; and time
Validation in the public data set

Figure 2 Overall process of proposed method

1) Crop row images. The crop row data sets under different growth stages of maize seedlings were obtained, including different weed densities, different light, seedling belt fracture and other complex scenarios.

2) Data set Labeling. In order to simplify the labeling, a method using dot-line labeling combined with morphological dilation was proposed.

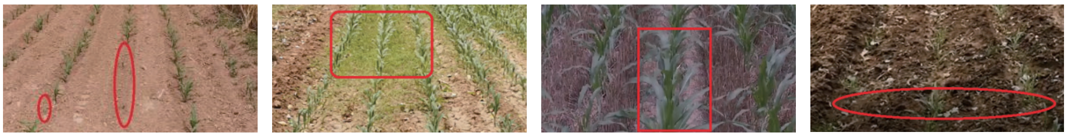
3) Networks structure. A crop row segmentation model based on convolutional neural network was constructed. After comparing the real-time performance and accuracy of different semantic segementation networks, BiSeNet V2 was selected as the backbone network. And Crop-BiSeNet V2 model combined with Spatial CNN

was proposed to improve the robustness of crop row segmentation model while meeting the real-time requirements.

4) Evaluation and Validation. The algorithm was evaluated experimentally in terms of detection accuracy and time. According to Intersection of Union (IoU), false positive rate (fp), false negative rate (fn) and accuracy rate (acc) indexes, Crop-BiSeNet V2 model was compared with BiSeNet V2 to show its superiority. Meanwhile, the sizes of training parameter files (DATA) and network structure files (META) of different models were compared to verify the real-time performance of the algorithm.

2.1　Data acquisition and preprocessing

The seedling stage of maize refers to the period from sowing to jointing. The seedling period is generally 30 d. The number of leaves contains 3-6 pieces. The growth of leaf volume is relatively slow. This stage is convenient for field management of agricultural machinery. At the same time, this stage determines the number of plants, and lays the foundation for the key period of large ears, more grains, and high yield of maize. Taking maize seedling as the research object, the data set was constructed in this paper. These images were taken with a Sony IMX380 digital camera. Furthermore, these images were taken in different scenes with moderate changes in yaw, pitch and roll angles to simulate field machinery operations. The data set contained 2697 images of different scenarios, such as different weed densities, seedling belt breakage, different growth stages, and shadow interference. The different complex scenarios are shown in Figure 3.



a. Seedling belt fracture　b. Weed interference　c. Different canopy sizes　d. Shadow

Figure 3　Four complex scenarios

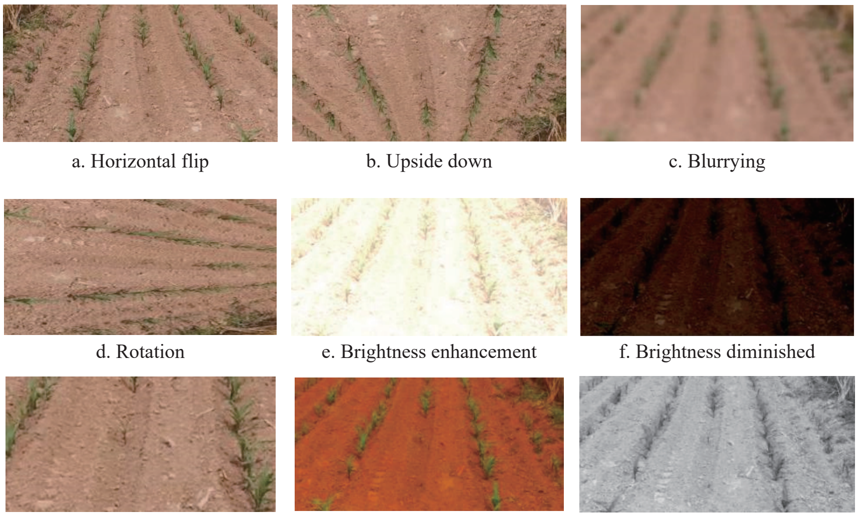
In Figure 3a, the seedling belt may be broken. The recognizable crop row pixels are missing, so segmentation is difficult. The model needs to explore the crop row context information to infer possible locations.

In Figure 3b, since the color and texture of weeds are similar to maize crops, different weed densities will cause great interference in crop row identification.

In Figure 3c, the canopy width is different at different growth stages of crops, which brings the challenge to curve fitting.

In Figure 3d, there are some differences in crop row under different light conditions. The area of shadow occlusion greatly increases the difficulty of segmentation.

The original image resolution was 1280×720 pixels, and the model input was adjusted to 512×1024 pixels to further reduce the calculation time and memory requirements. The data set of 2697 images was randomly divided into training set, validation set and test set at the 3:1:1 ratio. In order to adapt to the complex environment lighting and different angles, data augmentation technology was applied to the raw data. The data augmentation method converts the original image data into a variety of transformed outputs to increase the number of samples. The preprocessing steps included flipping, blurring, cropping, rotating, and adjusting the brightness, contrast, and saturation of the image. Data augmentation can effectively improve the generalization ability of the model. Based on Figure 3a, the data enhancement effects are shown in Figure 4.



a. Horizontal flip　b. Upside down　c. Blurrying

d. Rotation　e. Brightness enhancement　f. Brightness diminished

g. Cropping　h. Saturation enhancement　i. Saturation decreased

Figure 4　Examples of data enhancement

Large amounts of training data can improve the prediction performance of deep learning model. However, for DNN supervised training, data must be manually annotated with ground facts. Labeling large amounts of data is expensive and time-consuming. In addition, crop rows do not have well-defined boundaries, and fuzzy boundaries make labeling difficult.

In order to reduce the reliance on large-scale detailed labelings, the weak supervision technique was applied in this paper. Under the condition of weak supervision, the training images were only labeled at the image level or sparsely labeled at the pixel level. And the labeling required less time and effort. In this paper, a method based on dot-line labeling combined with morphological dilation was proposed. After the following test verification, this method had better performance.

Figure 5 simply shows the process of the proposed labeling method. The intersection of the stem and leaf of maize crop usually represents the center of the crop. A line is connected based on the center point of the same row of crop rows. A single line segment cannot represent more cropped pixels. Morphological dilation operation is applied to obtain more information. Compared with general color threshold-based segmentation and manual fine segmentation method, the labeling method proposed in this paper has better performance. The binarization results of different labeling methods are shown in Figure 6.

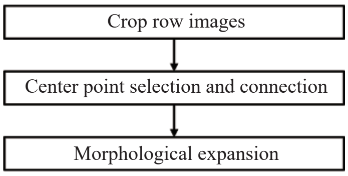
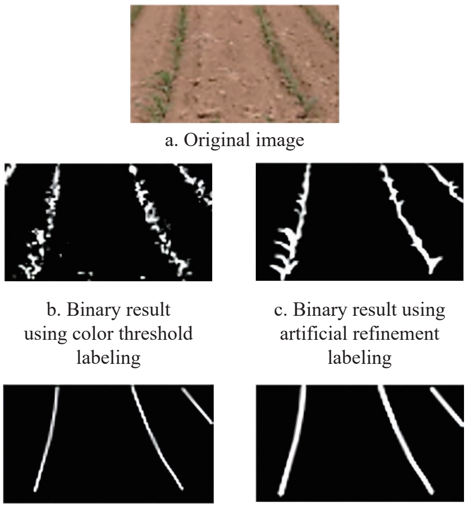


Figure 5 Process of dot-line labeling



a. Original image

b. Binary result using color threshold labeling

c. Binary result using artificial refinement labeling

d. Binary result using dot-line labeling

e. Binary result using our labeling method

Figure 6 Crop row labeling results

As shown in Figure 6, the effect of different methods varies greatly. Color threshold labeling method requires manual setting of the threshold, which is time-consuming and labor-intensive. This method also cannot meet a variety of complex scenarios. Artificial refinement labeling method is time-consuming. Dot-line labeling method can greatly reduce the complexity of labeling. But at different growth stages, crop rows have different widths, which means a single dot-line label cannot cover the actual situation. The proposed dot-line labeling method combined with morphological dilation algorithm can obtain more useful crop row pixel information, help the network to learn and converge quicker.

2.2 Lightweight backbone network selection

Image semantic segmentation is to classify images pixel by pixel and identify semantic information such as entity category. Classical semantic segmentation models include Fully Convolutional Networks (FCN)[26], U-Net[27], SegNet[28] and Deeplab V2[29]. The early FCN and SegNet have simple structures, low accuracy and poor real-time performance. Deeplab, RefineNet and PSPNet have made a lot of improvements to FCN, but they cannot meet the real-time requirements.

2.2.1 Network structure of BiSeNet V2

BiSeNet V2 network was a lightweight semantic segmentation network with fast segmentation speed. It proposed a two-sided segmentation structure. In order to directly compare the accuracy and real-time performance of various models, Yu et al.[30] conducted a test on the Cityscapes data set, the result is shown in Figure 7. From the perspective of detection accuracy (Mean IoU) and inference speed, BiSeNet V2 has better performance than other networks.

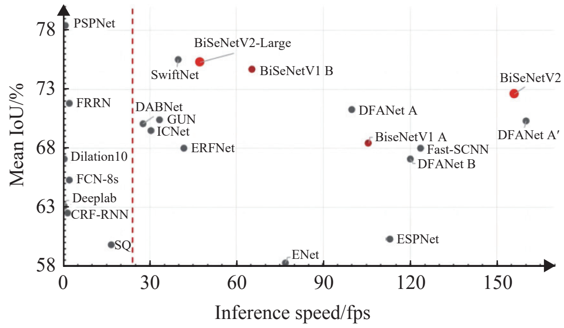


Figure 7 Comparison of common semantic segmentation network models

Based on the basic FCN network and referring to the structure of BiSeNet V2, a dual-branch crop row recognition backbone network is constructed. Low-level detail information and high-level semantic information are very important for semantic segmentation tasks. However, in order to speed up the inference, current methods always reduce high-level information, resulting in lower accuracy. In recent years, the BiSeNet V2 network has been used to process spatial details and semantic information respectively for the high precision and efficiency of real-time semantic segmentation. The model achieved 0.726 Mean IoU on the Cityscapes test set with a speed of 156 fps[30]. The structure of BiSeNet V2 network is shown in Figure 8, which mainly includes detail branch, semantic branch, guidance aggregation layer and enhancement training module.

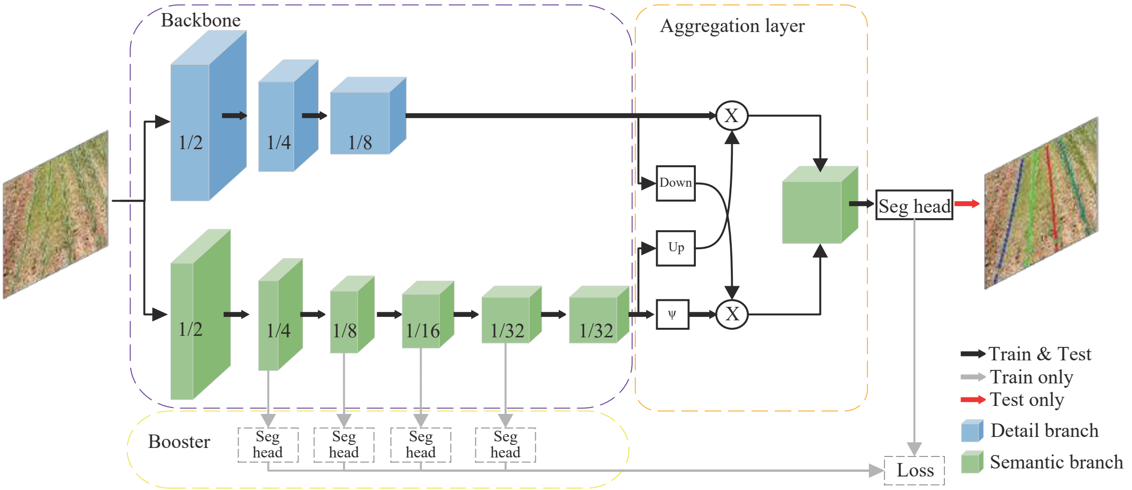


Figure 8 BiSeNet V2 network structure

The detail branch mainly focuses on the low-level details and extracts the high-resolution feature representation of the image. It has a wide channel and a shallow structure, so it has a rich channel capacity and can encode more spatial details.

Semantic branching is a lightweight convolutional structure with narrow channels and deep networks. The channel capacity is 1/4 of the detail branch. Fast sub-sampling strategies, such as convolution operation with a step size of 2 and maximum pooling, are used to combine outputs to improve feature representation and computational efficiency. At the same time, global pooling and residual link are used to embed global context information to increase receptive field to obtain high-level semantic information.

The aggregation layer is used for detail branching and feature fusion with semantic branching. Down and Up indicate image down-sampling and up-sampling, respectively. The features of different branches complement each other, and different scale feature representations can be obtained through different scale guidance. Feature fusion is performed by bilateral aggregation. Different sizes of output feature maps have different sizes, and simple fusion and stacking will only lose more image features. In the bilateral aggregated feature layer structure, the semantic branch obtains more detailed crop row context information, but the feature map size is smaller and requires 4 times the upsampling operation. The detail branch obtains more effective detail information and guides the semantic branch through a 4-fold downsampling operation. Finally, the results of the two branches are superimposed, and the same feature map is output to achieve the best segmentation effect. In order to further improve the segmentation accuracy, an enhanced training strategy is proposed, which guided the final segmentation results through multi-layer semantic feature branches. The feature representation is enhanced in the training stage and discarded in the inference stage without increasing the cost of the inference process.

2.3 End-to-end detection

In this study, an end-to-end crop row detection algorithm is proposed. A multi-branch structure is designed as shown in Figure 9. The network includes semantic segmentation and instance segmentation. The multi-branch structure realizes different segmentation based on the backbone network. And it provides the basis for subsequent crop row fitting.

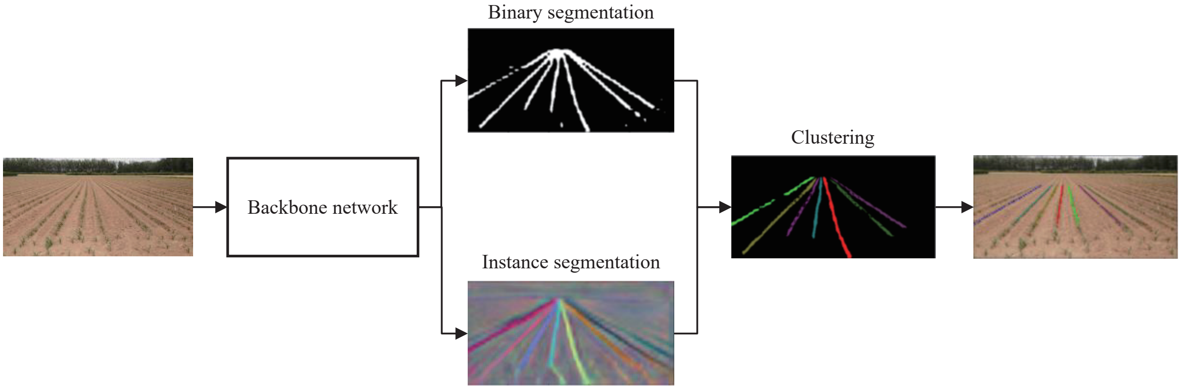


Figure 9 End-to-end structure with multiple branches

Semantic segmentation realizes the secondary classification of input images. The pixel is judged to be a background or crop. Each crop row forms a connecting line. When drawing the actual crop line, some cases where there is no obvious visual crop are also included. In this way, the network will learn to predict the crop position.

Since the two classes (crop and background) are highly imbalanced, the bounded inverse class weighting (W_{class}) is applied, as shown in Equation (1). In this equation, p_{class} is the probability of the corresponding category in the entire sample, and c is a hyperparameter (set to 1.02). The problem of unbalanced sample distribution is solved.

$$W_{\text{class}} = \frac{1}{\ln(c + p_{\text{class}})} \tag{1}$$

In order to achieve different number of crop rows, instance segmentation initializes each pixel as an embedding vector. The loss function is designed. The distance of the pixel vectors belonging to the same crop row is very small, and the pixel distance of different crop rows is very large. The loss function is mainly composed of two parts: variance loss (L_{var}) and distance loss (L_{dist}), as shown in Equation (2).

$$\begin{cases} L_{\text{var}} = \dfrac{1}{C}\sum\limits_{c=1}^{C}\dfrac{1}{N_c}\sum\limits_{i=1}^{N_c}\left[\|\mu_c - x_i\| - \delta_v\right]_+^2 \\ L_{\text{dist}} = \dfrac{1}{C(C-1)}\sum\limits_{cA=1}^{C}\sum\limits_{cB=1}^{C}\left[\delta d - \|\mu_{cA} - \mu_{cB}\|\right]_+^2, \ cA \neq cB \end{cases} \tag{2}$$

where, C is the number of crop rows; N_c is the number of pixels belonging to the same crop; μ_c is the average vector of crop rows; x_i is the pixel vector; $[x]_+$=max(0, x); δ_v represents the distance from the pixel vector to the mean vector; δ_d represents the distance between the mean vectors.

When the distance between the pixel vector of the same crop line and the average vector μ_c is greater than δ_v, L_{var} effectively makes x_i close to δ_d. When the distance between the average vectors μ_{cA} and μ_{cB} of different crop rows is less than δ_d, L_{dist} works to keep μ_{cA} and μ_{cB} away from each other. The trained feature vector is used in the clustering algorithm to achieve the purpose of instance segmentation.

2.4 Feature extraction of spatial information

2.4.1 Crop row spatial information analysis

In addition to the good scenes with clear crop rows, there are also a large number of complex scenes. Four kinds of complex farmland scenes are analyzed, including crop row breaking, weed interference, different canopy widths and shadow interference scenes, as shown in Figure 3.

In this study, the improvement method was proposed by analyzing the crop row in complex scenes. In the farmland scene, different crop rows have a certain position relationship. At the same time, the crop rows also have a certain priori shape, and the same crop is collinear. The different crop rows are approximately parallel in the real world. This kind of spatial location relationship is called spatial context information. Pan et al.[25] used Spatial CNN (SCNN) to extract Spatial context information. Through comparative experiments, SCNN is superior to CRF/MRF (Probability Graph Model), which has a large amount of computation and poor real-time performance. It performs well on targets with strong spatial relationships but poor visual cues, such as lane lines, wires and walls. Therefore, the Spatial CNN was combines with original BiSeNet V2 to improve the accuracy of crop row segmentation.

2.4.2 Network structure of Spatial CNN

Different from the traditional convolution method, the spatial CNN extends the layer-by-layer convolution to slice-by-slice convolution, in which the rows and columns of feature graphs are regarded as 'layers' and convolved sequentially. Assuming that the height, width, and channel number of the input 3D feature map are *H*, *W*, and *C*, the structure of Spatial CNN is shown in Figure 10. The model effectively overcomes crop row breakage and missed detection.

Spatial CNN firstly slices the 3D feature map from top to bottom, consisting of *H* slices. Then the first piece of vector is convolved, and the size of the convolution kernel is $C\times w$ (w=9). After the convolution result is non-linearly activated, the second slice vector is updated, and so on until the end. The whole process is called SCNN_D. A bottom-up convolution operation, SCNN_U, is then performed, which is similar to the SCNN_D process but with a change of direction. Similarly, slice and convolution operations are carried out from left to right and right to left to obtain the final 3D feature map. This piece-by-piece updating method is similar to residual network and can reduce the difficulty of training. Spatial CNN's unique convolution mode enables pixel information to be transmitted between different neurons on the same layer. It improves the ability of extracting spatial information from images. Therefore, it was applied to BiSeNet V2 network to improve the accuracy of crop row extraction in complex scenarios. In this paper, the improved BiSeNet V2 network is named Crop-BiSeNet V2. The structure is shown in Figure 11.

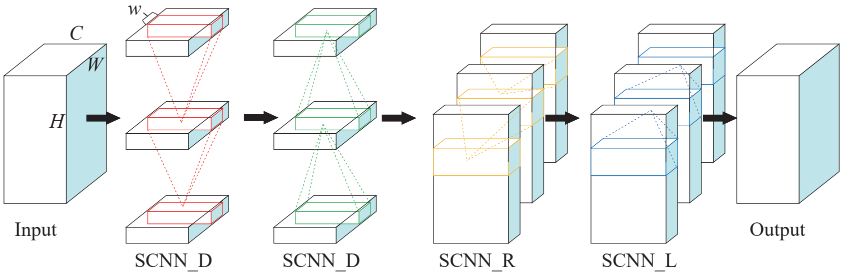


Figure 10 Spatial structure of CNN

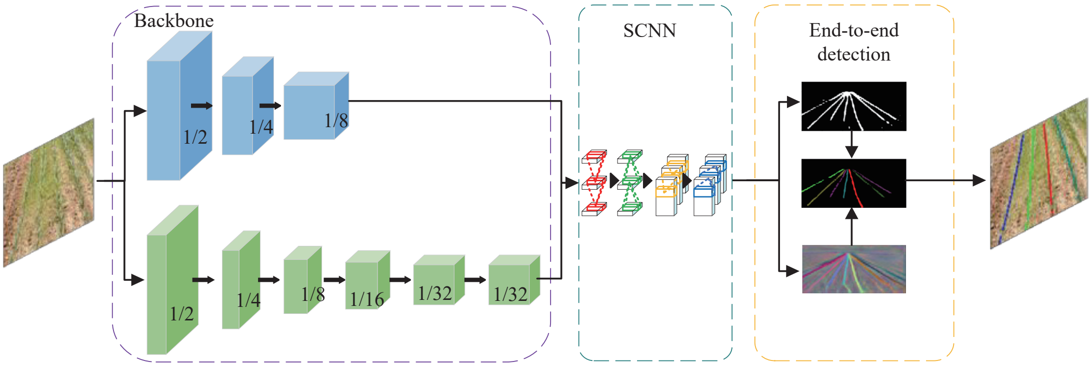


Figure 11 Crop-BiSeNet V2 network architecture

Although Spatial CNN structure can be added in any position of the network structure, the size of the input feature graph of Spatial CNN structure should not be too large in order to ensure the real-time performance of the network. In this study, it is placed between the encoder and the decoder. The feature map is 32×64 pixels.

Although the output crop row segmentation results of convolutional neural network are relatively fine, specific crop row parameters are still needed for the convenience of autonomous navigation, weeding and spraying applications. In this study, the quadratic polynomial curves are used to complete the fitting. Before crop row fitting, crop row feature points need to be obtained. The network segmentation model obtains the pixel points of each crop row as the fitting feature points. By inputting a series of feature points, the least squares method fits a curve with the smallest deviation from the feature points. Since crop row recognition needs to take into account both straight lines and curves, the polynomial function is used as the objective function of fitting. Finally, the results are remapped back to the original image.

3 Results and discussion

3.1 Evaluation indicators

The model performance was evaluated in terms of detection accuracy and time. Detection accuracy was an important index of model quality. The detected crop row was compared with the crop row of the ground truth image. The detection speed was directly related to the practical application value of the model. The output of the model was mainly divided into two categories: crop row pixels and background pixels. The confusion matrix of classification results is listed in Table 1. TP represents the positive sample predicted by the model to be positive. TN represents a negative sample predicted by the model to be negative. FP represents a negative sample that the model predicts to be positive. FN represents the positive samples that the model predicts to be negative.

Table 1 Classification results confusion matrix

Truth value	Predicted	
	Crop row pixel	Background pixel
Crop row pixel	TP	FN
Background pixel	FP	TN

The segmentation performance of the model was measured by IoU, and the false positive detection rate (fp), false negative detection rate (fn) and accuracy rate (acc) were calculated. The details are shown in Equations (3)-(6).

$$\mathrm{IoU}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \tag{3}$$

$$\mathrm{fp}=\frac{\mathrm{FP}}{\mathrm{TP}+\mathrm{FP}} \tag{4}$$

$$\mathrm{fn}=\frac{\mathrm{FN}}{\mathrm{TP}+\mathrm{FN}} \tag{5}$$

$$\mathrm{acc}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{6}$$

3.2 Model training and visualization

In this study, the TensorFlow framework was used to build the crop row segmentation model, and the model was trained and tested on the PC side. The computer configuration details are listed in Table 2.

Table 2 Computer configuration details

Name	Configuration
Operating system	Ubuntu18.04
CPU	Intel(R) Xeon(R) Silver 4210 CPU @ 2.20 GHz
GPU	NVIDIA Tesla T4
memory	16 G
CUDA	10.0
TensorFlow	1.13.1
Python	3.7

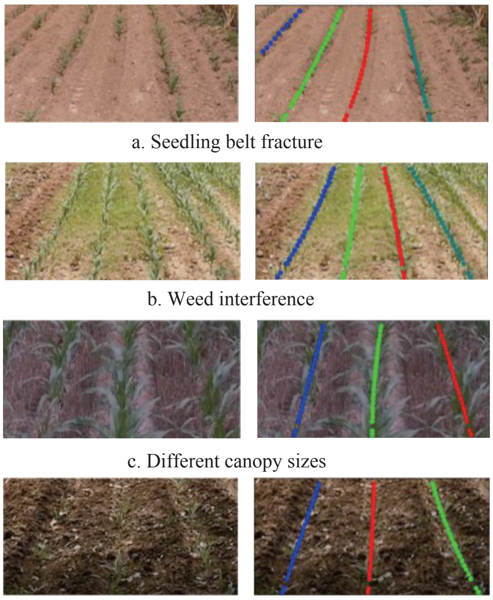
The size of the model input image was 512×1024, and the data set was divided into training, validation and testing sets at a 3:1:1 ratio. The model was trained for 500 epochs in total. The loss function was Softmax function and the optimizer was Adam. Specific training parameters are listed in Table 3.

Table 3 Training parameter setting

Parameter names	Parameter value
Training batches	32
Validation batches	4
Training epochs	500
Weight decay	0.0005
Learning rate attenuation factor	0.1
Initial learning rate	0.01

After the training of crop row model was completed, experiments were carried out in the test set, and the segmentation results are shown in Figure 12.

The accuracy rates in different scenarios were counted. The accuracy of the seedling belt fracture, weed interference, different canopy sizes and shadow data sets reached 0.9845, 0.9762, 0.9891 and 0.9746, respectively. As can be seen from Figures 12a-12d, there were some interference of weeds and seedling belt fracture or discontinuity in the image. The Crop-BiSeNet V2 proposed in this paper can successfully identify crop rows, and the segmentation results were relatively fine, which indicated that the model had good robustness. Since weeds have similar colors and textures, crop row pixels are easily confused. Pixels were missing due to the local seedling belt fracture. But the model had good spatial information extraction capabilities, and the interference of complex scenes had little effect on the overall results.



a. Seedling belt fracture

b. Weed interference

c. Different canopy sizes

d. Shadow

Figure 12 Data set test results

3.3 Comparison of different labeling methods on performance

In order to explore the impact of different labeling methods on performance, the experimental comparisons were conducted. The network model used BiSeNet V2 as the backbone network for training. The training parameters are listed in Table 3. Finally, IoU is used to measure the performance of the model. The results are listed in Table 4. The labeling method proposed in this paper has better performance.

Table 4 Comparison of different labeling methods

Labeling methods	IoU
Color threshold labeling	0.6700
Artificial refinement labeling	0.7250
Dot-line labeling	0.6200
This study	0.8138

Most of the focus in the literature is on detecting or segmenting “objects” with well-defined shapes, appearances, and boundaries. Due to the similar appearance and blurred boundaries, less attention is paid to complex scenes that are difficult to understand or even difficult to annotate correctly. A labeling method using dot-line labeling combined with morphological dilation algorithm is proposed. The labeling of complex scenes in crop rows is simplified. The proposed labeling method is particularly useful for pixel-based crop row segmentation.

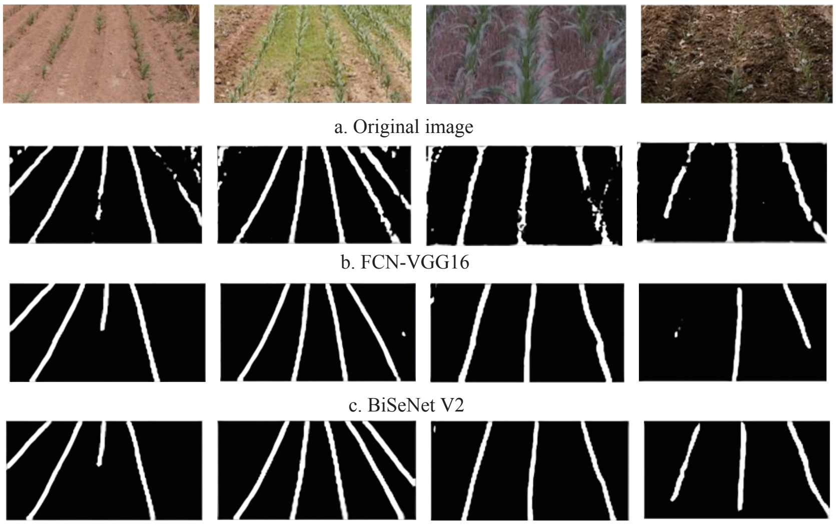
3.4 Model evaluation

According to IoU, fp, fn and acc indexes, FCN, BiSeNet V2 and Crop-BiSeNet V2 models were compared, and the basic network of FCN was VGG16. Different network models used the same training parameters for training, as listed in Table 3. The evaluation indexes were calculated on the test set, and the results are listed in Table 5.

Table 5 Evaluation index calculation results on the test set

Model	IoU	fp	fn	acc
FCN-VGG16	0.7928	0.2408	0.0631	0.9369
BiSeNet V2	0.8648	0.1903	0.0333	0.9667
Crop-BiSeNet V2	0.8980	0.1501	0.0189	0.9811

It can be seen from the above table that the segmentation accuracy of BiSeNet V2 model was far better than that of FCN model. The Crop-BiSeNet V2 model had the highest IOU. IOU was 0.0332 higher than original BiSeNet V2. Actual segmentation effects of different models are shown in Figure 13.


a. Original image

b. FCN-VGG16

c. BiSeNet V2

d. Crop-BiSeNet V2

Figure 13 Segmentation results of different models

Figure 13b shows the segmentation results of FCN-VGG16 model. On the whole, the crop row segmentation accuracy of this model was low. The crop row recognition was inconsistent and smooth, and there were missed detection under different background interference. And the anti-interference ability of the model was poor. Figure 13c shows the segmentation results of BiSeNet V2 model. Compared with FCN-VGG16, the results of the crop rows tested by this model are more coherent. Figure 13d shows the segmentation results of Crop-BiSeNet V2 model. Compared with BiSeNet V2, segmentation results of this model were more refined. SCNN module was used to enhance spatial context awareness and overcome the missed detection in test images 2 and 4.

To sum up, Crop-BiSeNet V2 had the highest segmentation accuracy and the best ability to resist complex environment in the data set used in this study.

3.5 Time performance comparison

On the crop row data set, the speed of processing single image was compared, and the sizes of training parameter files (DATA) and network structure files (META) of different models were also compared. The results are listed in Table 6. Considering the practical application cost in the future, this model used NVIDIA GeForce GTX 1050 low-cost graphics card for reasoning. The segmentation time of BiSeNet V2 model was the fastest, and the processing time of each image was 27.3 ms. The detection speed of FCN model was the slowest, and Crop-BiSeNet V2 was about 38.24 ms slower than the original BiSeNet V2 network, and compared with FCN, the detection speed of Crop-BiSeNet V2 was saved by 84.85 ms. FCN model had a large number of parameters and structure due to its many layers. BiSeNet V2 was a lightweight network, and Crop-BiSeNet V2 added Spatial CNN module on the basis of the former, which increased the number of parameters and model structure. Crop-BiSeNet V2 took about 65 ms to detect a single image, that is, about 15 fps.

Table 6 Performance comparison of different models

Model	Time/ms	Number of parameters/MB	Model structure/MB
FCN-VGG16	150.39	408.0	7.75
BiSeNet V2	27.30	26.6	14.7
Crop-BiSeNet V2	65.54	33.3	20.9

3.6 Performance on public data sets

The proposed algorithm was applied to the public data set[12]. The data set consists of 281 images. The image contains images of maize, celery, potatoes, onions, sunflowers, and soybeans. At the same time, the images were taken at moderately varying yaw and pitch angles. The accuracy rate reaches 0.9712. Some image examples are shown in Figure 14. The images a-d are examples of crops with weeds. The images e-h are examples of different types of crops. The images i-l are examples of crops at different growth stages. The images m-p are examples of crop rows with curvature. The data set has not been trained, directly input for testing. The result of binarization segmentation is shown in Figure 15. The test result is shown in Figure 16. It can be seen from the figure that the proposed algorithm is robust. Any number of crop rows can be detected. Straight and curved crop rows can be accurately detected.

4 Conclusions

In this study, a novel method for crop row detection was proposed, which combine light weight network BiSeNet V2 and Spatial Convolutional Neural Network (SCNN). Compared with the color threshold and manual fine labeling, the dot-line labeling combined with the morphological dilation method was simple and efficient. The backbone network of BiSeNet V2 had fewer network parameters to ensure the recognition speed. The Crop-BiSeNet V2 method was applied to the new and challenging crop row data set. It was able to effectively overcome the weed density, seedling with fracture, crop canopy width and shade. The results showed that continuous and long thin structures can be learned effectively, and performance can be greatly improved. This method was end-to-end recognition and can be used for unknown crop rows and straight or curved geometry. It was highly insensitive to different weed densities and shadow interference, and can accurately detect rows of maize crops at different growth stages. Compared with the classic

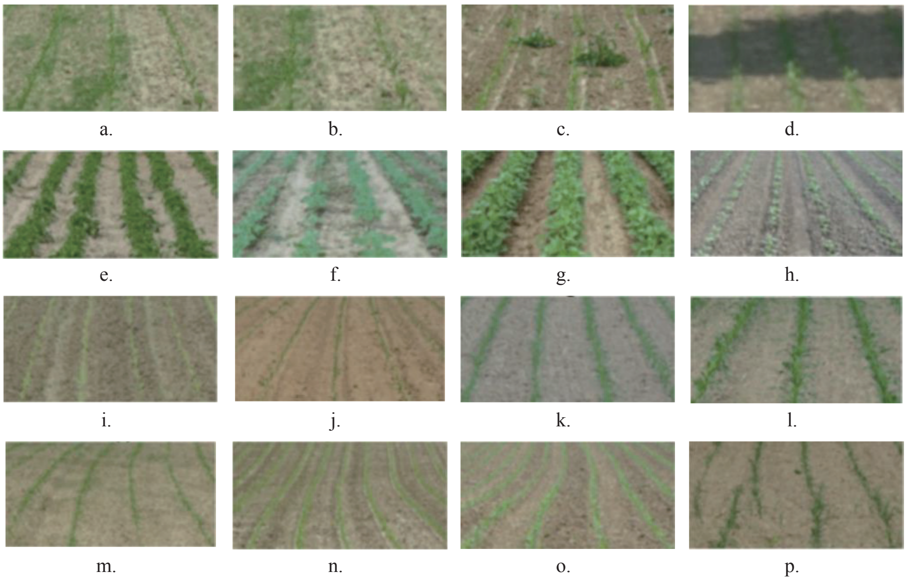


Figure 14 Sample images of public datasets

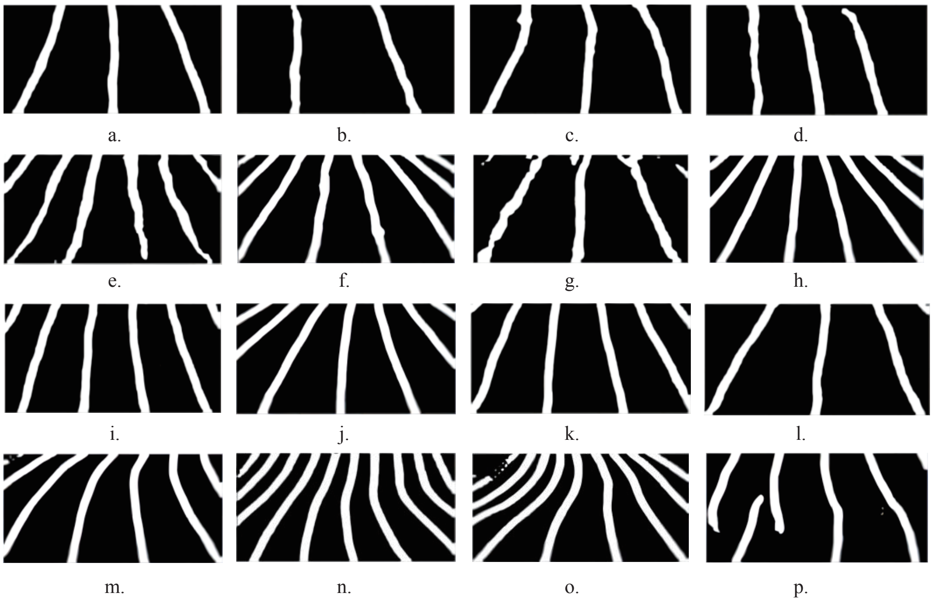


Figure 15 Binary segmentation of sample images

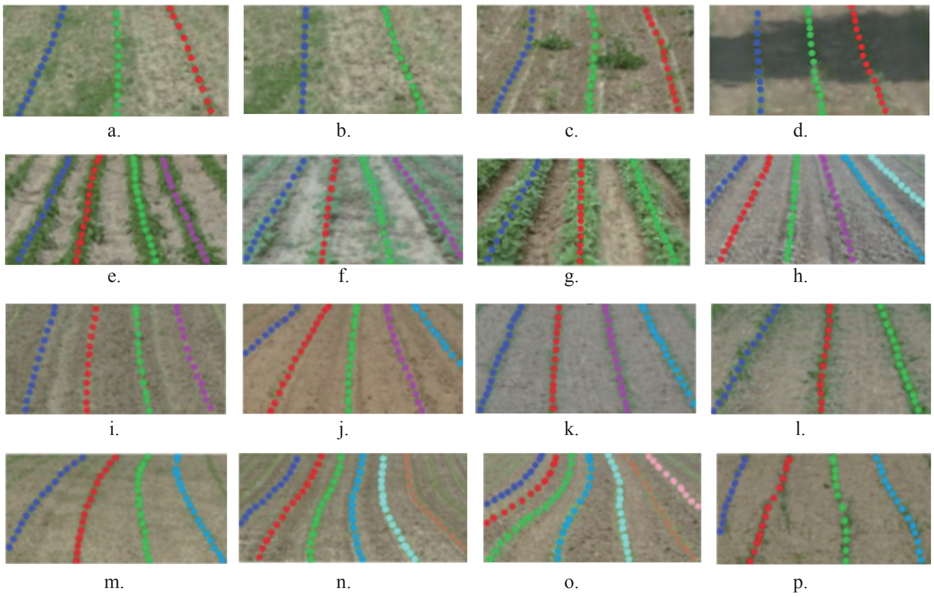


Figure 16 Public data set test results

Fully Convolutional Networks (FCN) algorithm, the proposed model had better accuracy and detection speed. The accuracy rate reached 0.9811, and the detection speed was saved by 84.85 ms.

The future work will apply the crop row detection algorithm to the embedded system for testing and deployment, and serve in the fields of unmanned spraying operation, weeding operation and assisted navigation, etc. Finally, the Crop-BiSeNet V2 method should be tested in practical application as a component of a prototype intelligent agricultural machinery system.

Acknowledgements

This work was financially supported by National Key R&D Program of China(Grant No. 2021YFB3901302) and Shandong Province, China (Grant No. 2021YFB3901300).

[References]

[1] Liu C, Lin H, Li Y, Gong L, Miao Z. Analysis on status and development trend of intelligent control technology for agricultural equipment. Transactions of the CSAM, 2020; 51(1): 1–18.

[2] Zheng H, Zhou X, He J, Yao X, Cheng T, Zhu Y, et al. Early season detection of rice plants using RGB, NIR-G-B and multispectral images from unmanned aerial vehicle (UAV). Computers and Electronics in Agriculture, 2020; 169: 105223.

[3] Shkanaev A Y, Krokhina D A, Polevoy D V, Panchenko A V, Sholomov D L, Sadekov R N. Analysis of straw row in the image to control the trajectory of the agricultural combine harvester (Erratum). Tenth International Conference on Machine Vision (ICMV 2017), SPIE, 2018; pp.19–28. doi: 10.1117/12.2310143.

[4] Ronchetti G, Mayer A, Facchi A, Ortuani B, Sona G. Crop row detection through uav surveys to optimize on-farm irrigation management. Remote Sensing, 2020; 12(12): 1967.

[5] Adhikari S P, Kim G, Kim H. Deep neural network-based system for autonomous navigation in paddy field. IEEE Access, 2020; 8: 71272–71278.

[6] Rabab S, Badenhorst P, Chen Y P P, Daetwyler H D. A template-free machine vision-based crop row detection algorithm. Precision Agric, 2021; 22(1): 124–53.

[7] Yang Y, Zhou Y, Yue X, Zhang G, Wen X, Ma B, et al. Real-time detection of crop rows in maize fields based on autonomous extraction of ROI. Expert Systems with Applications, 2023; 213: 118826.

[8] Li X, Su J H, Yue Z C, Wang S C, Duan F T, Hua J W. Vision-based navigation line extraction by combining crop row detection and RANSAC algorithm. In: 2022 IEEE International Conference on Mechatronics and Automation (ICMA), 2022; pp.1097–1002. doi: 10.1109/ICMA54519.2022.9856296.

[9] He J, Zang Y, Luo X W, Zhao R M, He J, Jiao J K. Visual detection of rice rows based on Bayesian decision theory and robust regression least squares method. Int J Agric & Biol Eng, 2021; 14(1): 199–206.

[10] Cao M Y, Tang F F, Ji P, Ma F Y. Improved real-time semantic segmentation network model for crop vision navigation line detection. Frontiers in Plant Science, 2022; 13: 898131.

[11] Chen J Q, Qiang H, Wu J H, Xu G W, Wang Z K, Liu X. Extracting the navigation path of atomato-cucumbergreenhouse robot based on a median point Hough transform. Computers and Electronics in Agriculture, 2020; 174: 105472.

[12] Vidović I, Cupec R, Hocenski Ž. Crop row detection by global energy minimization. Pattern Recognition, 2016; 55: 68–86.

[13] Su W, Jiang K P, Yan A, Liu Z, Zhang M Z, Wang W. Monitoring of planted lines for breeding corn using UAV remote sensing image. Transactions of the CSAE, 2018; 34(10): 92–98. (in Chinese)

[14] Kanagasingham S, Ekpanyapong M, Chaihan R. Integrating machine vision-based row guidance with GPS and compass-based routing to achieve autonomous navigation for a rice field weeding robot. Precision Agric, 2020; 21(4): 831–855.

[15] Wang S, Zhang W, Wang X, Yu S. Recognition of rice seedling rows based on row vector grid classification. Computers and Electronics in Agriculture, 2021; 190: 106454.

[16] Tu C, van Wyk B J, Djouani K, Hamam Y, Du S. An efficient crop row detection method for agriculture robots. In: 2014 7th International Congress on Image and Signal Processing, 2014; pp.655–659. doi: 10.1109/CISP.2014.7003860.

[17] Jiang G Q, Zhao C J, Si Y S. A machine vision based crop rows detection for agricultural robots. In: 2010 International Conference on Wavelet Analysis and Pattern Recognition, 2010; pp.114–118. doi: 10.1109/ICWAPR.2010.5576422.

[18] Li Y H, Wang C F, Wang C Y, Deng X L, Zhao Z X, Chen S D, et al. Detection of the foreign object positions in agricultural soils using Mask-RCNN. Int J Agric & Biol Eng, 2023; 16(1): 220–231.

[19] Chen P F, Ma X. Research status and trends of automatic detection of crop planting rows. Scientia Agricultura Sinica, 2021; 54(13): 2737–2745.

[20] Wang A, Zhang M, Liu Q, Wang L, Wei X. Seedling crop row extraction method based on regional growth and mean shift clustering. Transactions of the CSAE, 2021; 37(19): 202–10. (in Chinese)

[21] Liu H, Jia H, Wang G, Glatzel S, Yuan H, Huang D. Method and experiment of maize (*Zea mays* L.) stems recognition based on deep learning and image processing. Transactions of the CSAM, 2020; 51(4): 207–215. (in Chinese)

[22] Zhang Q, Wang J H, Li B. Extraction method for centerlines of rice seedings based on YOLOv3 target detection. Transactions of the CSAM, 2020; 51(8): 34–43. (in Chinese)

[23] Liu F Y, Lin G S, Shen C H. CRF learning with CNN features for image segmentation. Pattern Recognition, 2015; 48(10): 2983–2992.

[24] Luo Y S, Yang L, Wang L, Cheng H. Efficient CNN-CRF network for retinal image segmentation. Cognitive Systems and Signal Processing, Singapore: Springer, 2017; pp.157–65. doi: 10.1007/978-981-10-5230-9_17.

[25] Pan X G, Shi J P, Luo P, Wang X G, Tang X O. Spatial as deep: Spatial CNN for traffic scene understanding. Proceedings of the AAAI Conference on Artificial Intelligence, 2018; 32(1): 12301.

[26] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017; 39(4): 640–651.

[27] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W M, Frangi A F, editors. Medical image computing and computer-assisted intervention–MICCAI 2015, Cham: Springer International Publishing; 2015; pp.234–241. doi: 10.1007/978-3-319-24574-4_28.

[28] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017; 39(12): 2481–2495.

[29] Chen L C, Papandreou G, Kokkinos I, Murphy K, Yuille A L. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018; 40(4): 834–848.

[30] Yu C Q, Gao C X, Wang J B, Yu G, Shen C H, Sang N. BiSeNet V2: Bilateral network with guided aggregation for real-time semantic segmentation. Int J Comput Vis, 2021; 129(11): 3051–3068.