



# Hybrid rice row detection at the pollination stage based on vanishing point and line-scanning method

Dongfang Li<sup>1</sup> · Chengye Dong<sup>1</sup> · Boliao Li<sup>1</sup> · Yongwei Wang<sup>1</sup> · Jun Wang<sup>1</sup>

Accepted: 5 December 2022 / Published online: 22 December 2022  
© Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Hybrid rice row detection at the pollination stage is critical for the automation of in-field pollination agricultural vehicles. The parental crops of hybrid rice are planted at intervals in seed production fields with narrow inter-row spacing. During the advance of the pollination vehicle, in addition to the centerline of the crop row, information on the crop region boundaries is required to guide the vehicle and prevent it from running over the crop. For complete crop row detection, a novel machine vision-based method was presented to identify each of the individual regions of the crop rows, more than the centerlines, by line-shaped mask scanning combined with the vanishing point of the crop rows. The approach consisted of grayscale transformation, vanishing point detection, crop region identification, boundary position fine-tuning and crop region segmentation. Its region detection performance outperformed the convolutional neural network-based (CNN-based) methods with an intersection over union (IoU) of 0.832, an accuracy of 90.48%, a recall of 86.36%, a precision of 98.96% and an f1-Score of 92.23%. Its centerline extraction ability was compared with Hough Transform-based and SegNet-based methods on the basis of average lateral distance (ALD) between the ground truth line and the detected line. The proposed method resulted in an ALD of 1.943 pixels in a 640\*360 resolution image, which was superior to the Hough Transform-based (5.704) and the SegNet-based (3.555) methods.

**Keywords** Autonomous navigation · Machine vision · Crop detection · Vanishing point · Region detection · Line-scanning method

## Introduction

High-yield and high-quality seed production of hybrid rice can only be achieved by artificially assisted cross-pollination due to rice's meager natural hybridization rate (Li et al., 2017). At present, the pollination of hybrid rice is mainly carried out manually or mechanically and has not yet been automated (Jiang et al., 2021). Pollination requires high precision and long-time repetitive work, exhausting and demanding for humans. Automated

---

✉ Jun Wang  
jwang@zju.edu.cn

<sup>1</sup> Department of Biosystems Engineering, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China

pollination is urgently required due to growing rice consumption and an ever-declining agricultural labor force (Kakani et al., 2020). As a foundational task to realize automated pollination, in-field autonomous navigation of agriculture vehicles is required in the first place.

In-field autonomous navigation has been widely researched over recent years. The potential advantages of in-field autonomous navigation generally encompass human labor reduction, vehicle positioning accuracy improvement and operation safety enhancement (Gerrish et al., 1997; Han et al., 2004). Commonly used in-field autonomous navigation methods include global navigation satellite systems (GNSS), machine vision navigation and laser navigation (Bonadies and Gadsden, 2018). Among these, a vision sensor outperforms the others considering its cost-effectiveness, flexibility and technical characteristics of simulating the human eye, making machine vision a research priority (Mousazadeh, 2013). Accurate crop row detection is crucial to provide reliable guiding information for agricultural vehicles in machine vision navigation. Extensive research has been conducted on detecting crop rows with image processing approaches, and they can be divided into two main categories: methods based on traditional image processing techniques and methods based on convolutional neural networks (CNN).

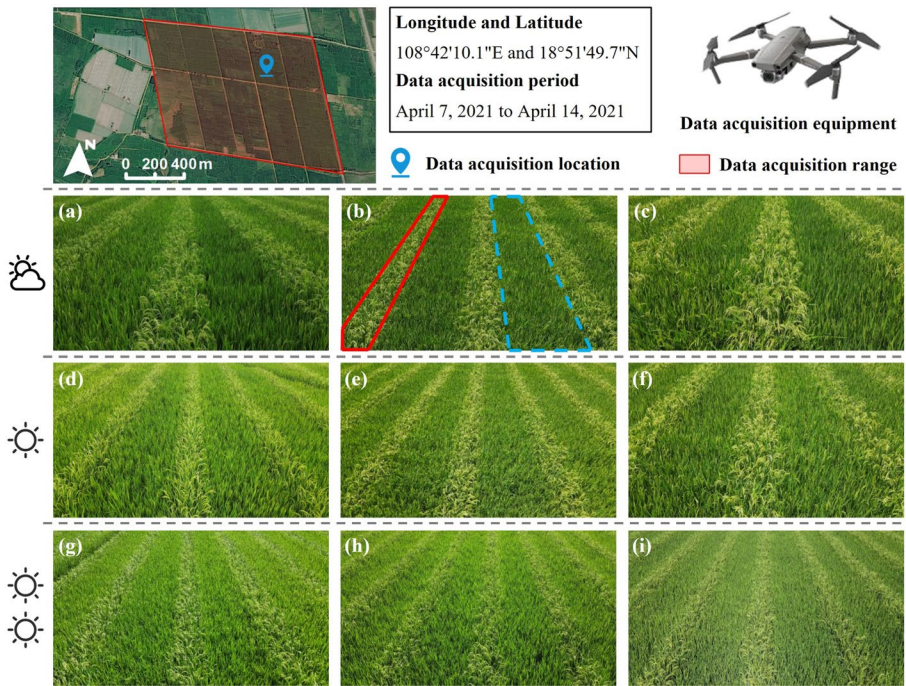
Hough transform-based (HT-based), linear regression-based and vanishing point-based methods are the three most widely studied methods based on traditional image processing techniques for crop row detection. HT converts the straight line detection in an image into a peak point detection in parameter space. Zhang et al. (2017) carried out HT on SUSAN corner points representing the rice crop rows to obtain reference guidelines. By limiting the detection angle and setting a known point in advance, Chen et al. (2021a) alleviated the computational burden of HT. Though HT is convenient to apply, it's limited to the detection angle and the detection number of crop rows, and the algorithm cannot work in complex scenarios such as the presence of crop gaps and weed noise. Linear regression is used to fit the centerline of the crop row, so it is necessary to complete the extraction and clustering of crop feature points before. Zhang et al. (2018) utilized a modified vegetation index and double thresholding technique to obtain the binary image of maize crops. A position clustering algorithm and the shortest path method were applied to confirm the final clustered feature points, and the linear regression method was employed to fit the crop rows. Yu et al. (2021) accomplished paddy crop row detection based on an improved Otsu (Otsu, 1979) algorithm and a double-dimensional clustering method. Then linear regression was performed to obtain the navigation lines. A major drawback of linear regression is its sensitivity to additional visual noise from weeds or other outliers, even with robust regression methods. However, such noise is typically unavoidable in field scenes, which restricts the fitting accuracy of linear regression methods. The basic principle of the vanishing point-based approach comes from the image perspective property. Crop rows parallel to each other in the real world will converge at one point in the image, which is the vanishing point. Then the crop rows arrangement can be inferred according to the vanishing point position. Pla et al. (1997) extracted the skeleton information of the vegetation area to obtain several candidate crop lines and selected those lines that could converge to the vanishing point as the real crop row. Jiang et al. (2016) determined the location of the vanishing point by clustering the voting points in the Hough space and eliminated false detection lines according to the positional relationship between the vanishing point and candidate crop rows. However, the above vanishing point-based methods rely heavily on complex image preprocessing to obtain feature points representing crop rows, which result in image information loss and increases image processing time. Besides the three types of studies, Vidović et al. (2016) matched a crop row distribution template with row width

and inter-row spacing as the main parameters for every input image. A global optimization function was established to determine the coefficients of the two parameters. This method performed well both in the straight line and curve detection. But the algorithm contained a lot of user-defined hyper-parameters, and its time cost needed to be further optimized for the actual field operation process. Ma et al. (2021) first determined the straight lines representing the crop rows approximately using the horizontal strip method and the prior knowledge of crop row distribution patterns. Then the crop lines were refined through iteratively clustering. However, this method relied on strong prior knowledge, leading to declining applicability.

In addition to the shortcomings mentioned above, these crop row detection algorithms based on traditional image processing techniques have a common defect. That is, they only focus on extracting the crop row centerline while ignoring the crop region detection. The detected centerline is suitable for providing navigation information for agricultural vehicles operating in crop rows with wider inter-row gaps because only the center tracking accuracy is required in such fields. However, the planting pattern of crop rows in hybrid rice seed production fields is quite different. The paternal and maternal hybrid rice rows in seed production fields are planted at intervals and close to each other to enable uniform and adequate pollination, so the inter-row gap for pollination vehicles to pass through is small. Crop region detection extends the work of centerline extraction to obtain complete crop rows with more helpful information like crop area and boundary so that the navigation accuracy would be complimented. The centerline of the crop row and the crop region need to be identified simultaneously during the pollination period to properly formulate the navigation strategy to guarantee that the pollination vehicle won't drive over the crops while in use.

CNN is undergoing a period of rapid development. Scholars have attempted to apply CNN-based target detection and semantic segmentation models in the crop recognition process due to their solid ability of data expression. Ponnambalam et al. (2020) adopted the SegNet to segment strawberry field images into crop and not-crop regions, and the navigation path planning was completed according to the prediction masks of the crop regions. Kim et al. (2020) divided an orchard image into rectangular patches, then determined the type of each patch through a CNN-based classification model, and finally completed the navigation centerline fitting according to the location of the image patches belonging to the road part. CNN-based methods are capable of solving the crop region detection problem. However, as a strongly supervised learning algorithm, CNN models require substantial non-repetitive samples for training and a mobile device with powerful computing (Wang et al., 2020). The time-consuming and laborious manual annotation work has become a bottleneck hindering the successful implementation of CNN-based methods. Meanwhile, when deploying CNN models on mobile computing terminals, crop region detection accuracy has to be compromised in order to ensure their real-time performance, which will cause sub-optimal navigation performance in hybrid rice fields at the pollination stage.

To address the shortcomings of previous research, this study proposed a real-time and accurate image processing-based crop row detection method to accomplish the region identification as well as centerline extraction in the seed production field of hybrid rice during the pollination period. By providing agricultural machinery with additional information on the location of crop region boundaries, their navigation accuracy is expected to be further improved. Taking full advantage of the perspective imaging characteristics, the method presented located the crop area according to the vanishing point and the arrangement pattern of hybrid rice rows. A line-shaped mask scanning algorithm was constructed for complete crop row detection according to the overall distribution of pixel values in the image.



**Fig. 1** Crop row images of hybrid rice at the pollination stage under different lighting conditions, **a–c** are images in low light condition, **d–f** are captured under moderate illumination, **g–i** are taken under high light

The proposed method did not call for elaborate graphic preprocessing steps, so the image features could keep to a maximum extent, which reduced the time required and guaranteed the rice row detection accuracy.

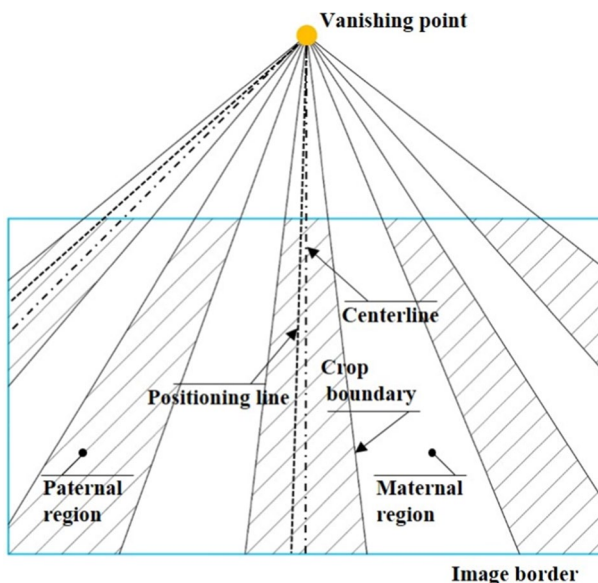
## Materials and methods

### Image acquisition

The camera, mounted on a Mavic Air 2 drone produced by Da-Jiang Innovations (DJI), Shenzhen, China, was used for hybrid rice image acquisition. The focal length of the camera lens was 24 mm and the aperture setting was  $f/2.8$ . When performing image acquisition, the drone flying height was kept 1 to 3 m above the crop, and the inclination angle between the camera and the horizontal line was maintained at  $10^{\circ}$ – $45^{\circ}$ . Sampling operation was performed in Gancheng Township, Dongfang City, Hainan Province, China, at  $18^{\circ}51'49.7''$  N and  $108^{\circ}42'10.1''$  E, in April 2021. Images were taken and stored as 24-bit RGB color images in jpeg format with a resolution of  $1920 \times 1080$ . Before image processing, the images were downsampled to  $640 \times 360$  through bilinear interpolation to reduce the computational load.

Some images of hybrid rice taken under different light conditions during pollination are shown in Fig. 1. In Fig. 1(b), the part enclosed by the solid red frame is one paternal region and the dotted blue frame surrounds one maternal region. The paternal and maternal

**Fig. 2** The region distribution model of hybrid rice rows

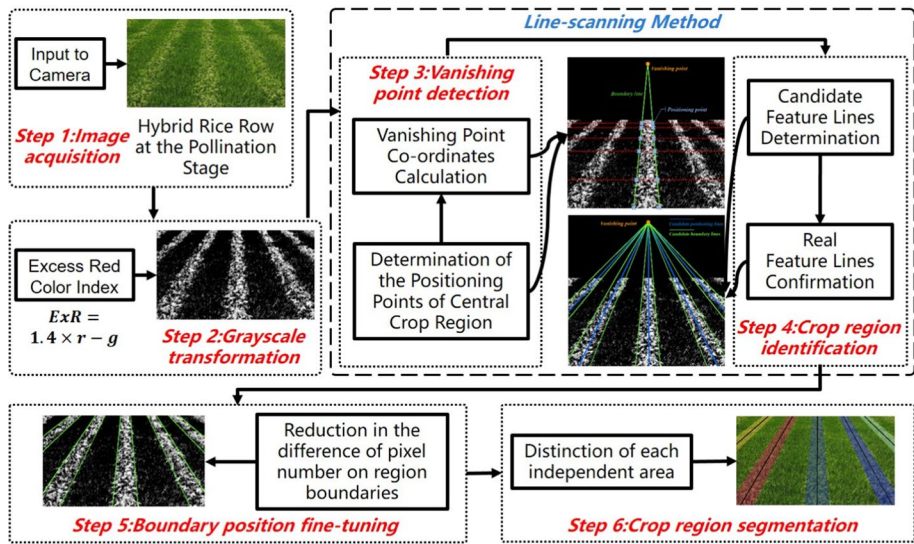


regions are planted interlaced to facilitate pollen dispersal. The pollination period generally lasts 7 days. The pollination process is completed during rice flowering time, which is 1.5–2 h at noon each day. During the flowering period, the color of the paternal region turns from green to yellow due to pollen production, while the maternal region remains basically green. When the agricultural vehicle performs auxiliary pollination operations, the vehicle straddles the paternal row, and its wheels are located within the inter-row gaps between the paternal and maternal areas. With the pollination machine marching forward, the crop regions are continuously detected according to the currently captured image frame, and the navigation information is updated in real-time to adjust the vehicle's motion state. The constructed image dataset of hybrid rice rows in seed production fields during the pollination stage has been made publicly available on this web page.<sup>1</sup>

### Crop region detection strategy

A geometric model constructed to describe the hybrid rice region distribution is shown in Fig. 2. Since the mechanized planting mode makes the crop row a more regular arrangement, the boundaries of the crop rows can be regarded as straight lines starting from the vanishing point within the working field of view (Pla et al., 1997). Each of the required crop regions is enclosed by image borders and two boundaries of crop row. The centerline is the angle bisector of the two region boundaries. The positioning line is the auxiliary line generated in the intermediate process to determine the approximate location of the paternal area. Once the boundary lines and the positioning lines are determined, the crop regions and centerlines can be obtained accordingly.

<sup>1</sup> <https://github.com/ZeroHeading/Image-dataset-of-hybrid-rice-crop-rows-during-pollination-in-seed-production-fields.git>.



**Fig. 3** Flowchart of the proposed method

The flowchart of the proposed method is demonstrated in Fig. 3. The strategy is to grayscale the image first and determine the vanishing point directly from the gray image using the line-scanning method. After that, the line scanning method was performed again to find the lines representing the crop line boundaries from the rays with the vanishing point as the endpoint. Meanwhile, the rays that represent the positioning lines are identified to distinguish the paternal regions from the maternal regions. The line-scanning method is in the form of a line-shaped mask scanning through the image towards a specific direction. Several pixel value statistics of different image parts are recorded to generate a set of feature curves during the scanning process. The waveforms and inflection points of the curves corresponded strictly to the image content, so the position of the boundary lines and positioning lines can be inferred from their distribution. The vanishing point detection results may be inaccurate because the crop row distribution in the field does not perfectly follow the ideal model, so the identified crop boundaries are fine-tuned in the fourth step. Finally, the crop area is segmented based on the relationship between the detected boundary and positioning lines. Next, the role of each step and its implementation way is described in detail.

### Grayscale transformation

Grayscale transformation was the only image preprocessing step required in this method. The primary purpose of grayscale transformation was to emphasize the pixels belonging to the paternal region and restrain the rest. Redundant information was removed to relieve the computational burden. The three-channel grayscale histogram of the color image in Fig. 4a is depicted in Fig. 4b. The overall color image is greenish, so the green channel dominates the high grayscale response. The red channel has the middle response level due to the yellow rice panicle in the paternal regions. Therefore, the excess red color index (ExR) (Meyer et al., 2004), a grayscale method that emphasizes the content of the red channel, was selected in this study to accentuate the paternal areas. ExR was calculated by Eq. (1), where  $r$  and  $g$  represented the components of the red and green channels of the

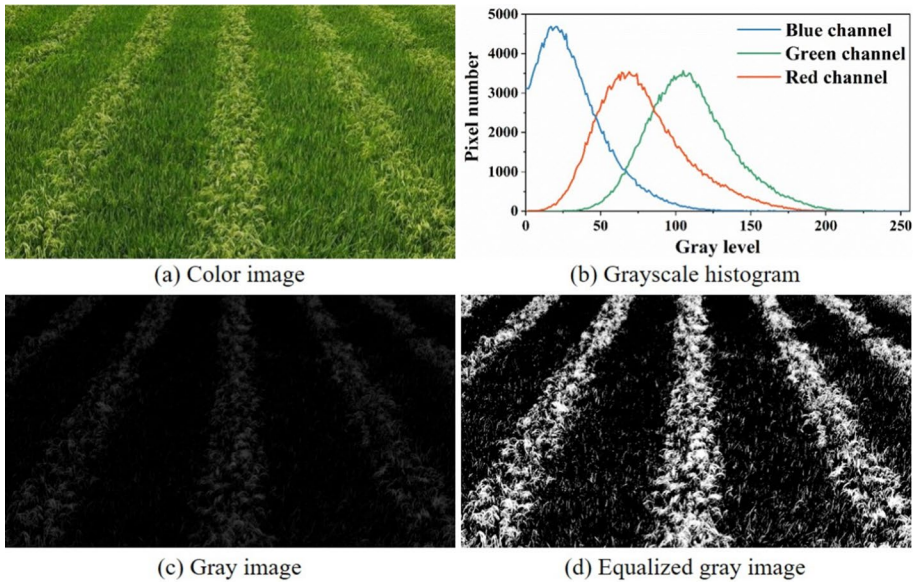


Fig. 4 Grayscale transformation

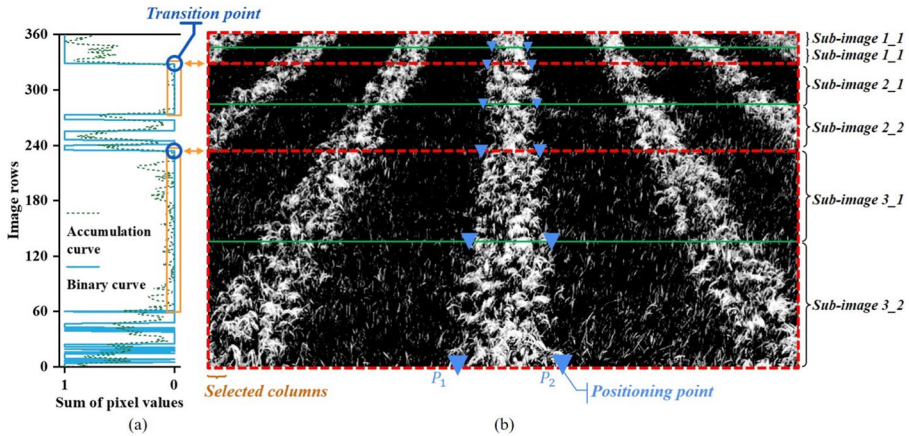
color image, respectively. The grayscale image was histogram equalized to promote the display effect. Equation (2) was used for equalizing the histogram of the gray image, where  $H$  and  $W$  were the height and width of the gray image, respectively,  $k$  was the number of gray levels in the image, and  $n_j$  denoted the number of pixels belonging to the gray level. The result of the grayscale transformation is shown in Fig. 4d.

$$ExR = 1.4r - g \tag{1}$$

$$E_k = \frac{(L-1)}{HW} \sum_{j=0}^k n_j, k = 0, 1, 2, \dots, L - 1 \tag{2}$$

### Vanishing point detection

The vanishing point is the intersection of the extension lines of the boundary lines of all crop rows in the image. The vanishing point detection was carried out on the gray image to set a reference for crop region inference. The vanishing point can be obtained by calculating the intersection of any two crop region boundary lines. The paternal region lying near the middle of the image was selected for the crop boundary calculation. Compared with other regions, the central paternal region was superior to symmetry, which facilitated its boundary calculation. The vanishing point detection was completed in two major steps. First, determine the positioning points on the left and right boundaries of the central paternal region using the proposed line-scanning method. Second, fit the left and right points into two straight lines, respectively, to get the intersection point.



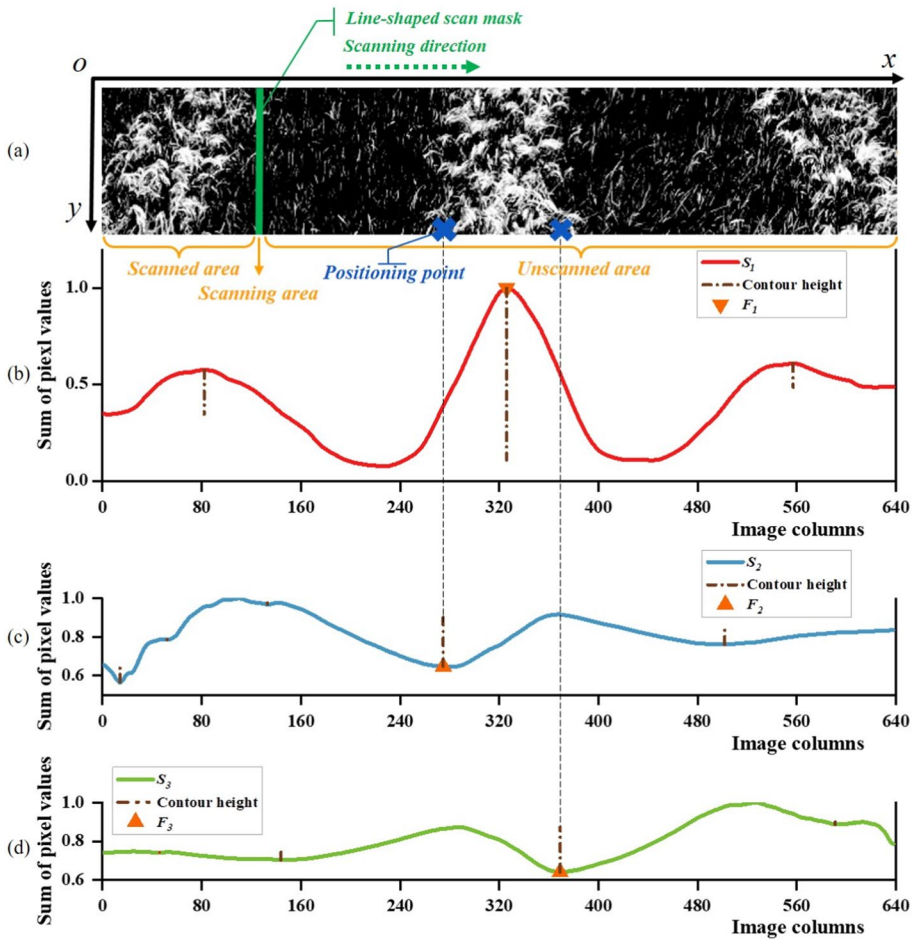
**Fig. 5** Image cutting: **a** normalized pixel value accumulation curves, **b** sub-images

### Determination of the positioning points of the central crop region

**(a) Sub-images cutting** Taking the crop row image in Fig. 5b as an example, the central crop area was determined by the positioning points located at the region boundary. As end-points of the central crop area,  $P_1$  and  $P_2$  have fewer pixels on one side and denser pixels on the other. This pixel distribution pattern was used by the line-scanning method to determine the abscissa of the positioning points. Two positioning points on the left and right sides of the central region could be resolved in one image at a time, but a boundary line needed to be fitted with multiple positioning points, so the grayscale image was cut horizontally into multiple sub-images. The cutting process was done in two steps. The red dotted line in Fig. 5b outlined the three sub-images obtained by the first cutting procedure. In this cutting way, basically, all foreground pixels in the same column of the sub-images belonged to one paternal row. The same pixel distribution pattern in all sub-images facilitated the line-scanning method to find the positioning points. The solid green line divided each sub-image equally into two parts to avoid the situation where no sub-image was produced in the first cutting procedure and increased the number of positioning points.

The first cutting procedure consisted of three steps. (1) Add the pixel values of each row horizontally in the selected columns to generate an accumulation curve. (2) Classify the accumulation curve into a binary curve with  $T_v$ , where  $T_v$  is the classification threshold of the curve height value. (3) Find those curve segments whose length was longer than  $T_l$  in the zero-value part of the binary curve, where  $T_l$  is the classification threshold for the curve length value. The curve segments found were marked with orange rectangle frames in Fig. 5a, and their transition points to the one-value part of the binary curve were recorded as the cutting points. The second cutting was to bisect the sub-images along the horizontal midline. Suppose both row boundaries were not visible on the image for all crop rows (e.g., the left or right boundary was not visible on the image for the left or right edge rows), then no sub-images would be generated according to the above two-step cutting method. In this case, the grayscale image would be horizontally cut into five equal parts to ensure sufficient positioning points for boundary line fitting. After the image cutting, the bottom endpoints of the truncated central paternal region in each sub-image were to be located and used as positioning points.





**Fig. 6** The determination process of the positioning points in the sub-image: **a** the bottom sub-image, **b** the curve of the normalized sum of pixel values in the current scanning area, **c** the curve of the normalized sum of pixel values in the scanned area, **d** the curve of the normalized sum of pixel values in the unscanned area

**(b) Positioning point identification** The crop area in the center of the sub-image has the densest distribution of pixels, and the number of pixels on the left and right sides in the horizontal direction at the bottom endpoint of the area changes drastically. The line-scanning method was used to quantify the above pixel distribution to obtain the exact location of the positioning points. Take the bottom sub-image as an example to illustrate the specific implementation steps. The process diagram is illustrated in Fig. 6a. In this part, the scanning line used was a vertical line-shaped mask with a width of one pixel and a length equal to the height of the current sub-image. The sub-image was scanned horizontally from left to right. At each scanning position, the specific gravity of the pixel values in the scanning, scanned and unscanned areas were calculated by Eq. (3) and denoted as  $W_1$ ,  $W_2$  and  $W_3$ , respectively.

$$W_i = \sum_{m=0, n=0}^{cols, rows} v_{m,n} / (cols * rows) \quad (3)$$

where  $i = 1, 2, 3$ .  $cols$  and  $rows$  were the width and height of the corresponding area and  $v_{m,n}$  was the pixel value at the image position  $(m, n)$ .

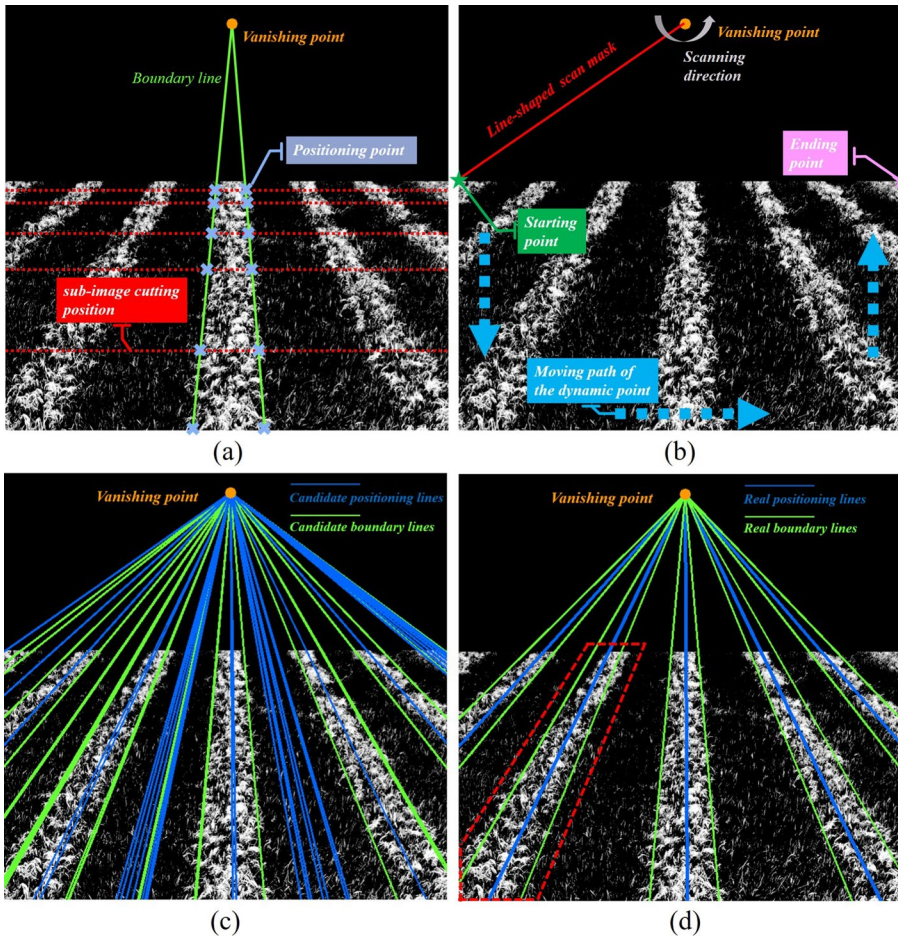
Three curves were generated according to the recorded  $W_1$ ,  $W_2$  and  $W_3$ . The waveform of each curve and the position of the inflection point on it could reflect the specific law of pixel distribution in the sub-image. The Savitzky-Golay filter (Luo et al., 2005) was applied to smooth the curves so that the interference of the noise pixels that remained in the grayscale transformation process could be eliminated while retaining the curve's primary inflection points. The smoothed curves  $S_1$ ,  $S_2$  and  $S_3$  are shown in Fig. 6b-d. The prominent local peak points of  $S_1$ , which should appear within the paternal region, were used to determine the approximate location of the paternal regions. The appearance of local valley points on  $S_2$  and  $S_3$  indicated that the scanning mask was entering or leaving a region with a high density of pixels, so the endpoints of the central crop region were determined by the abscissa of the local valley points on  $S_2$  and  $S_3$ . When identifying the local peak points on  $S_1$  and the local valley points on  $S_2$  and  $S_3$ , the curve was taken as a one-dimensional array and all local extreme points were found by comparison of neighboring values.

Multiple local extreme points existed on a curve, some of them were desired feature points, and the rest were noise points. The contour height of the curve at the local extreme points was calculated to distinguish the two. The contour height of a curve peak measures how much a peak stands out from the surrounding baseline of the curve and is defined as the vertical distance between the peak and its lowest contour line. The strategy to compute a peak's contour height was as follows: (1) Extend a horizontal line from the current peak to the left and right until the line either reaches the window border or intersects the signal again at the slope of a higher peak. An intersection with a peak of the same height is ignored. (2) On each side, find the minimal curve value within the interval defined above. These points are the peak's bases. (3) The higher one of the two bases marks the peak's lowest contour line. The contour height can then be calculated as the vertical difference between the peak height itself and its lowest contour line.

The co-ordinates of the positioning points were determined by follow-up steps. (1) Calculate the contour height of  $S_1$  at all local peak points and record the peak point with maximum contour height as the feature point  $F_1$ . The image column at  $F_1$  has the greatest pixel density, which should lie within the central paternal area. (2) Record the first local valley point to the left of  $F_1$  on  $S_2$  and the first local valley point to the right of  $F_1$  on  $S_3$  as the positioning point  $F_2$  and  $F_3$ .  $F_2$  and  $F_3$  were the local valley points with the largest contour height on  $S_2$  and  $S_3$ , and they were adjacent to the central paternal region, indicating that they were the endpoints of the central crop region. The positioning points on the other sub-images could also be obtained in the same way.

### Vanishing point co-ordinates calculation

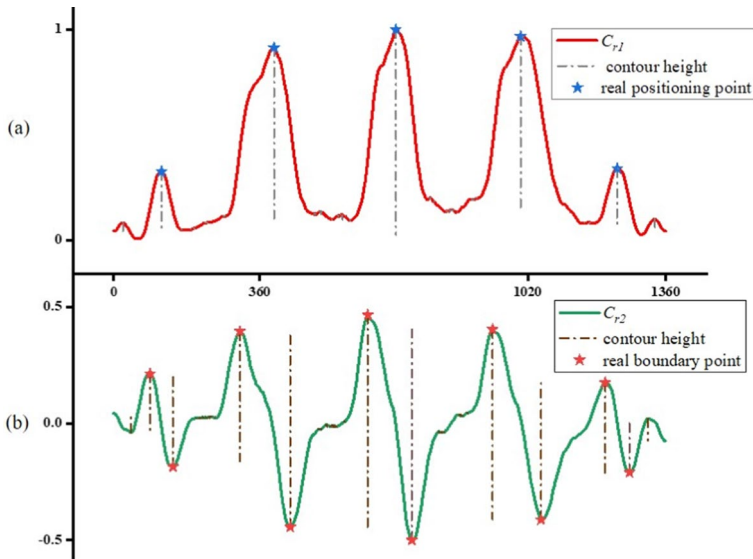
The detected positioning points were divided into two categories representing the left and right boundaries of the central paternal region, respectively. The random sample consensus (RANSAC) (Fischler and Bolles, 1981) method was used to perform robust linear regression for each category. The intersection colored in orange in Fig. 7a of the two fitting lines is the vanishing point of the crop rows.



**Fig. 7** **a** Schematic diagram of the vanishing point detection process, **b** Schematic diagram of the rotary scan process, **c** Candidate feature lines, **d** Real feature lines

### Crop region identification

The paternal regions could be determined by their left and right boundaries and distinguished by the positioning lines. Boundary and positioning lines were feature lines that were hidden in rays ending at the vanishing point. An infinite number of rays can be emitted from the vanishing point, and the line-scanning method was used to narrow the range from all rays to the candidate boundary lines and the candidate positioning lines. Afterwards, the candidate feature lines were divided into two categories, with the noise lines removed and the real feature lines retained. The candidate feature lines and the real feature lines are drawn in Fig. 7c and d.



**Fig. 8** Feature curves generated in the rotary scan process: **a** normalized distribution curve of Vline, **b** normalized gradient curve of Vline

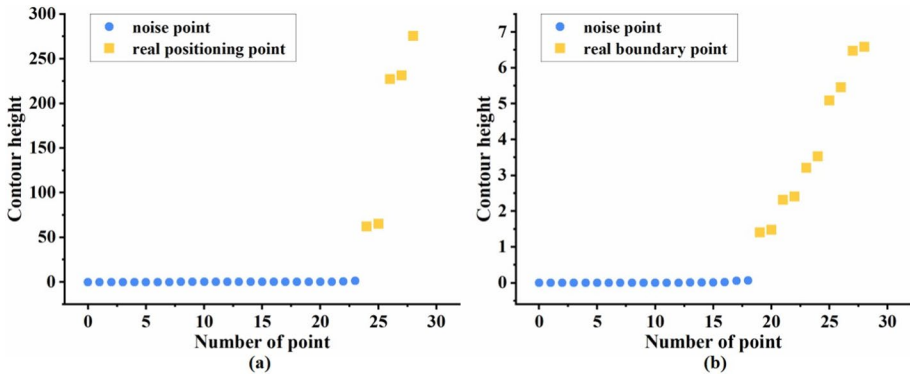
### Candidate feature lines determination

The specific implementation of the second line-scanning process was as follows. A rotary scan was performed on the grayscale image to obtain the candidate feature lines, as shown in Fig. 7b. The scan line used here was a line-shaped mask of one-pixel width. One endpoint of the line was the vanishing point, and the other was a dynamic point, moving on the image borders. The starting point, ending point and moving trajectory of the dynamic point are given in Fig. 7b. The moving step was set to one pixel, so the co-ordinates of the dynamic point were obtained by sequentially taking points from the left, bottom and right borders of the image. At each scanning position, the sum of pixel values in the current scanning area was calculated by Eq. (4) and recorded as  $V_{line}$ .

$$V_{line} = \sum_{t=1}^l v_t \tag{4}$$

where  $l$  was the number of pixels in the scanning area and  $v_t$  was the value for each pixel.

As shown in Fig. 8,  $C_{r1}$  was generated according to the recorded  $V_{line}$ , and  $C_{r2}$  was the gradient curve produced with one derivation of  $V_{line}$ .  $C_{r1}$  and  $C_{r2}$  were also smoothed headmost by the Savitzky-Golay filter to reduce noise interference.  $C_{r1}$  represented the pixel distribution in the scanning area. The local peak points of  $C_{r1}$  appeared when the scanning mask passed through the paternal regions so that the candidate positioning lines could be determined by the abscissa of local peak points on  $C_{r1}$ .  $C_{r2}$  corresponded to the pixel value changing rate along the scan direction. Therefore, the abscissa of local extreme points on  $C_{r2}$  could be used to identify the candidate boundary lines containing the region boundaries. However, there were some noise points in the local extreme points of  $C_{r1}$  and  $C_{r2}$  that must be eliminated since they were produced when the mask



**Fig. 9** Classification results of the **a** real positioning points, **b** real boundary points. Note: The points in **a** and **b** are the candidate positioning points and the candidate boundary points arranged in ascending order of contour height, respectively

was scanned across the remaining pixel blocks in the maternal area and the holes in the paternal area.

**Real feature lines confirmation**

Although both noise points and real feature points appeared as local extreme points on  $C_{r1}$  and  $C_{r2}$ , the noise points were produced by small pixel blocks, so they could only cause slight curve fluctuations. On  $C_{r1}$  and  $C_{r2}$ , the contour heights of the noise points, as seen in Fig. 9, are roughly equal to and substantially lower than those of the actual feature points. Accordingly, the real feature lines could be picked out from candidate feature lines. The following are the precise steps for implementation.

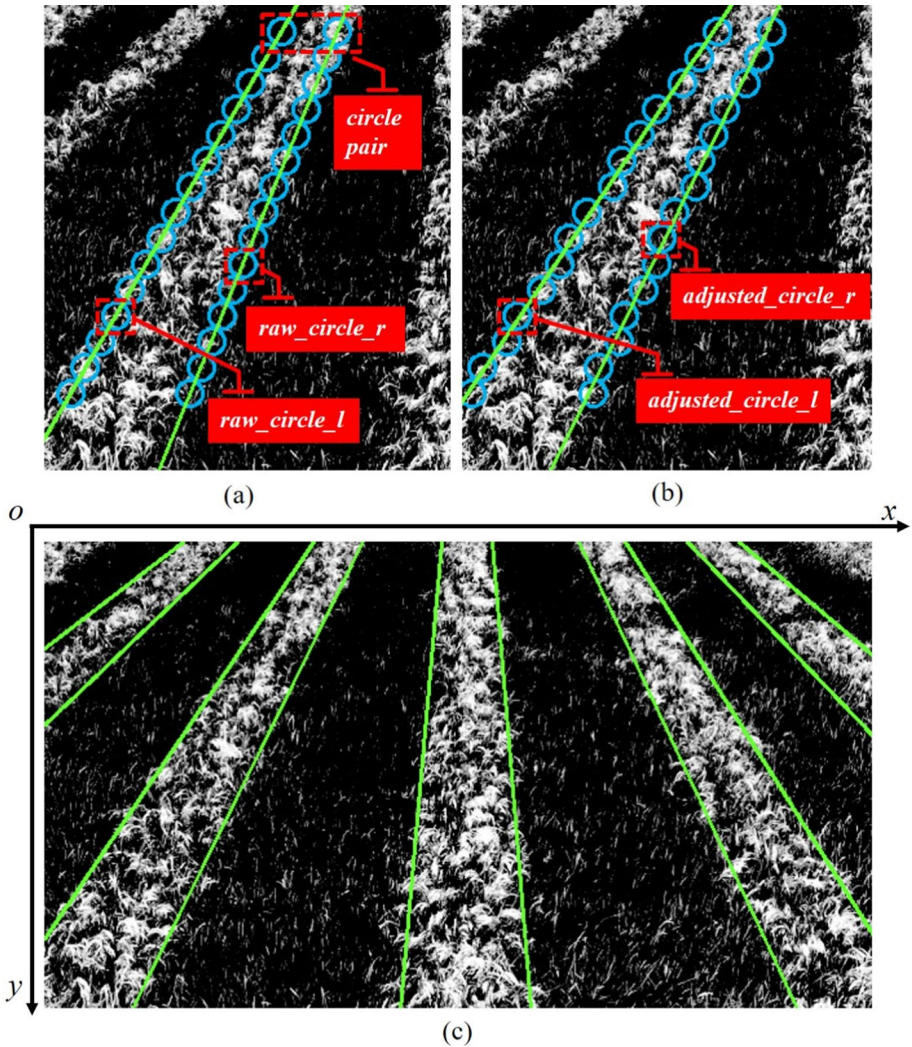
The contour height was denoted as  $H_{cnt}$ . The relative distance of  $H_{cnt}$  between two adjacent points, denoted as  $d_h$ , was used to classify the point set.  $d_h$  was calculated by Eq. (5). Arrange  $H_{cnt}$  in ascending order first, calculate  $d_h$  between two adjacent points in turn, and the point with the largest  $d_h$  was the demarcation point.

$$d_h = (H_{cntj+1} - H_{cntj})/H_{cntj+1} \tag{5}$$

where  $j=0, 1, \dots, \text{number of points}$ . The real feature points are marked with stars in Fig. 8 and the real feature lines are shown in Fig. 7d.

**Boundary position fine-tuning**

As shown in Fig. 7d, the identified boundaries of crop regions are sometimes inaccurate when the vanishing point detection is biased or when the planting of the crop rows does not quite fit the ideal distribution model. When the two boundary lines of a crop area are accurately detected, they should locate at the junction of the paternal and maternal areas, and the number of pixels distributed around them should be roughly equal. Based on this, an algorithm for fine-tuning the boundary position was proposed. As depicted in Fig. 10, the number of foreground pixels in the two circles belonging to one circle pair would be



**Fig. 10** Boundary position fine-tuning: **a** the initial boundary detection result of the crop region surrounded by the dotted red frame in Fig. 7(d), **b** fine-tuned result of the single region, **c** fine-tuned results of the image. Note: raw\_circle\_l and raw\_circle\_r were the circles generated along the initially identified left and right crop region boundary, respectively. adjusted\_circle\_l and adjusted\_circle\_r were the corresponding fine-tuned circles. The radius of the blue circles was set to 10 pixels, the vertical distance between the centers of two adjacent circles on the same side was set to 20 pixels. The positive horizontal direction of the image is specified as right

approximately the same when one circle pair is located at the correct boundary position. Otherwise, the center position of one circle pair should be adjusted horizontally.

When the difference between the number of pixels contained in the two boundary circles at the same horizontal position was bigger than the preset proportional threshold  $T_1$ , the position of the circle pair was considered to be inaccurate. In order to make the algorithm maintain a certain tolerance for noise interference, the boundary line

fine-tuning algorithm was only implemented when the proportion of circle pairs in the correct position was less than the preset tolerance threshold  $T_2$ . During the fine-tuning process, the circle pair was moved left or right horizontally by a fixed step  $S_e$  each time, and the moving direction was always toward the side of the circle that contained more pixels. The fine-tuning process was looped until the proportion of circle pairs with accurate center positions of all circles exceeded  $T_2$ , and the algorithm would be terminated.

The pseudo-code for fine-tuning the two boundaries of each region is given in *Algorithm 1*. The fine-tuned boundaries were fitted by the least-square method. In *Algorithm 1*,  $C_{pair}$  was the abscissa of the center of the circle pair.  $k$  was the number of the circle pairs.  $circle\_area$  was the area for a single circle.  $T_1$ ,  $T_2$ ,  $S_e$  were set to 0.15, 0.8, 3, respectively.  $Abs()$  was a function to find the absolute value.

---

**Inputs:**  $raw\_circle\_l, raw\_circle\_r, C_{pair}, circle\_area, k, T_1, T_2, S_e$

**Outputs:**  $adjusted\_circle\_l, adjusted\_circle\_r$

---

```

# Calculate the initial boundary position score
1:  $F_{score} = 0$ 
2: for  $i = 1:k$ 
3:    $g\_left_i = \text{number of the foreground pixels in } raw\_circle\_l_i / circle\_area$ 
4:    $g\_right_i = \text{number of the foreground pixels ins } raw\_circle\_r_i / circle\_area$ 
5:   if  $Abs(g\_left_i - g\_right_i) \leq T_1$ 
6:      $F_{score} += 1$ 
# Determine whether the initial boundary position is good. If not, adjust it
7: while  $(F_{score}/k) < T_2$ 
8:   for  $t = 1:k$ 
9:     if  $Abs(g\_left_t - g\_right_t) > T_1$  and  $g\_left_t > g\_right_t$ 
10:       $C_{pair}_t -= S_e$ 
11:     if  $Abs(g\_left_t - g\_right_t) > T_1$  and  $g\_left_t < g\_right_t$ 
12:       $C_{pair}_t += S_e$ 
# get the adjusted circles
13:    $adjusted\_circle\_l_t, adjusted\_circle\_r_t = raw\_circle\_l_i, raw\_circle\_r_i$ 
# Calculate the adjusted boundary position score
14:    $F_{score} = 0$ 
15:   for  $m = 1:k$ 
16:      $g\_left_m = \text{number of the foreground pixels in } adjusted\_circle\_l_m / circle\_area$ 
17:      $g\_right_m = \text{number of the foreground pixels in } adjusted\_circle\_r_m / circle\_area$ 
18:     if  $Abs(g\_left_m - g\_right_m) \leq T_1$ 
19:        $F_{score} += 1$ 
20: return  $adjusted\_circle\_l, adjusted\_circle\_r$ 

```

---



**Fig. 11** Ground truth image of **a** crop regions, **b** centerlines

### Crop region segmentation

The crop regions were segmented according to the relative position of the fine-tuned boundary lines and positioning lines. When there was a positioning line between the two boundary lines, it would be considered an individual paternal region. Then the centerline of each crop row was extracted as the angular bisector of the detected region's left and right boundaries.

### Evaluation metrics

#### Ground truth image creation

The results obtained by the proposed methods were evaluated for their similarity to ground truth data. As shown in Fig. 11, two types of ground truth images are manually created by experts. The ground truth of the paternal region was generated by enclosing the region corners. The ground truth of the paternal centerline was created by defining both the midpoint of the upper and lower border of the crop row.

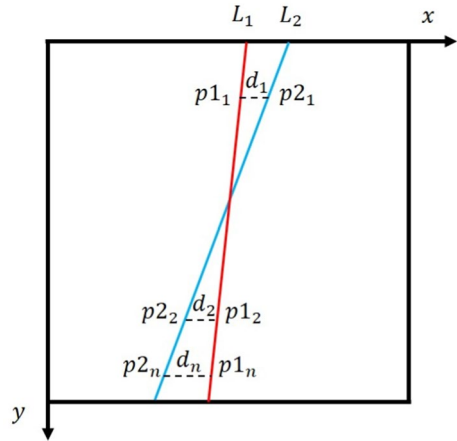
#### Evaluation index of crop region detection

Intersection over union (IoU) was adopted in the evaluation process of region detection. IoU was used as a measurement index to describe the degree of coincidence between the detection results and the ground truth. The detection frame of the paternal region was regarded as a prediction frame and labeled as positive data. In contrast, the maternal region was viewed as the background and labeled as negative data. It was considered a valid detection when the IoU between the data frame and the corresponding ground truth frame was more significant than 0.8. Otherwise, it would be treated as an invalid match. Taking into account the statistics of true positives (TP), false negatives (PN), true negatives (TN), and false positives (FP), multiple metrics like accuracy, recall, precision, and f1-score as written in Eqs. (6–9) were used to measure the performance of paternal region identification (Chen et al., 2021b).

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$



**Fig. 12** Schematic diagram of accuracy evaluation of the extracted centerline



$$recall = \frac{TP}{TP+FN} \tag{7}$$

$$precision = \frac{TP}{TP+FP} \tag{8}$$

$$f1 - Score = \frac{2TP}{2TP+FN+FP} \tag{9}$$

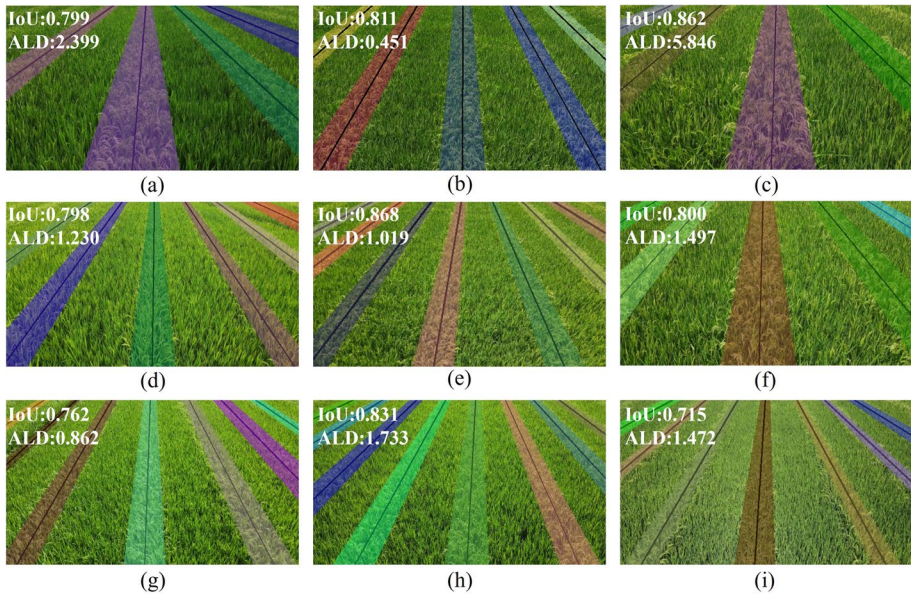
where TP represented the positive data that was correctly identified; TN was the negative data that was correctly identified; FP was the negative data that was incorrectly identified as positive; FN was the positive data that was incorrectly identified as negative.

**Evaluation index of crop row extraction**

A method was constructed to calculate the centerline extraction accuracy. Its schematic diagram of the mechanism is given in Fig. 12. \$L\_1\$ was the manually labeled ground truth line representing the centerline of the crop row, and \$L\_2\$ was the detected centerline. \$n\$ point pairs were picked at identical horizontal positions on the two straight lines. The lateral distance \$d\_n\$ between \$p1\_n\$ and \$p2\_n\$ was calculated, and the average lateral distance (ALD) of \$d\_n\$ was used as the evaluation criterion.

**Results and discussion**

Experimental evaluations were carried out from three aspects to assess the proposed method comprehensively. They were crop region detection performance, crop centerline extraction accuracy and execution time cost of image processing.



**Fig. 13** Crop row detection results obtained by the proposed method. **a–i** are test images taken under different lighting conditions. Note: The score of the detection result of the proposed method in terms of IoU and ALD is given in numbers

## Evaluation of the performance on crop region detection

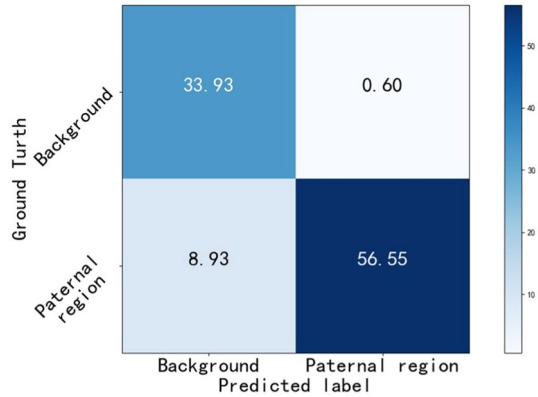
### Overall performance of crop region detection

The paternal region detection results obtained with the proposed method are shown in Fig. 13. The approach performed excellently in detecting paternal regions in hybrid rice seed production fields. The detected region boundaries were accurate, regular and complete in shape, which confirmed the applicability of the method presented.

Given the similarity between the approach presented and the CNN-based target detection algorithms in terms of region detection effect, the confusion matrix analysis was first introduced to gain a comprehensive understanding of the region detection capabilities of the proposed method. The established confusion matrix is depicted in Fig. 14. It can be seen intuitively from the confusion matrix that the correct detected crop areas accounted for a considerable proportion, and a few wrongly detected areas were concentrated in the paternal area. This was because the paternal regions located in the upper corners of the image were challenging to detect, while the maternal regions were treated as background, thus avoiding such problems.

Table 1 demonstrates the paternal region detection results accessed through Eqs. (6–9). The accuracy achieved a value of 90.48%, which showed that the proposed method possessed considerable crop region detection performance. The recall was at a relatively low percentage of 86.36%, indicating that the proposed method had omissions in identifying positive data. The mis-detected paternal regions were basically those near the upper left or right corners of the image, as shown in Fig. 13. The boundaries of the missed paternal regions were relatively blurred and short in length due to the

**Fig. 14** Confusion matrix of region detection results obtained by the proposed method



**Table 1** Detailed analysis results of the confusion matrix of the proposed method

Accuracy	Recall	Precision	f1-Score
90.48%	86.36%	98.96%	92.23%

perspective imaging property, which led to failing recognition. Additionally, the incomplete triangular crop regions that appeared in the upper corners of the image were also mis-detected. The precision represented the proportion of the actual paternal region in all positive instances, reaching a value of 98.96%, demonstrating that the detection results of the paternal region are highly credible. It can be seen from Table 1 that the proposed method obtained a low recall but a high precision, so the f1-score was introduced to measure the values of the two comprehensively. The f1-score was found to be 92.23%, which confirmed the approach as a reliable method in crop region detection.

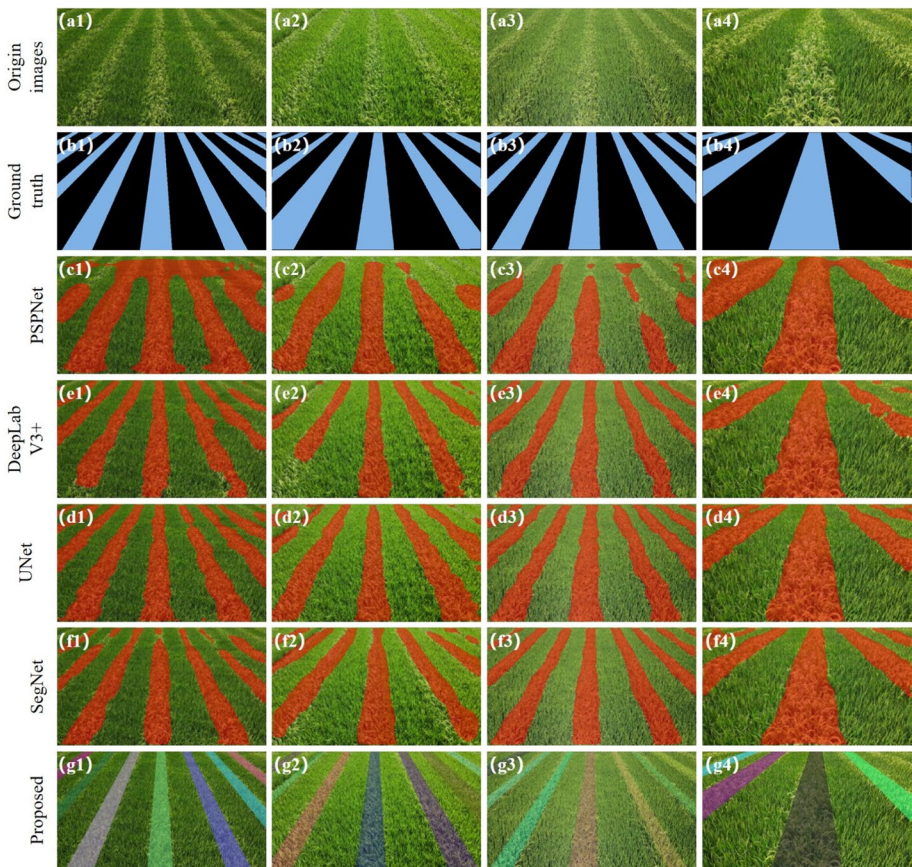
**Comparison with state-of-the-art methods**

A comparative study was conducted on the proposed method and several state-of-the-art semantic segmentation methods, including UNet (Ronneberger et al., 2015), SegNet (Badrinarayanan et al., 2017), PSPNet (Zhao et al., 2017) and DeepLab V3+(Chen et al., 2018). MobileNet (Howard et al., 2017) was adopted as the backbone of feature extraction for these semantic segmentation models. One thousand three hundred manually annotated images of hybrid rice crop rows at the pollination stage were randomly divided into training, validation and test sets in a ratio of 7:2:1 for network training. All networks were trained for 100 epochs, and the model weights were used to infer on images in the test set. The IoU achieved by each model is listed in Table 2, and their visual segmentation results are depicted in Fig. 15.

First, it can be seen that the semantic segmentation models were only able to distinguish between paternal and maternal regions, while the proposed method further completed the identification of paternal areas for a single instance. In terms of segmenting hybrid rice rows, PSPNet performed the worst. It occasionally made general predictions about different paternal region locations while missing most of the intended ones. The paternal regions detected by DeepLab V3+ were smaller than the ground truth labels, and there were some excessively protruding or depressed small areas at the region

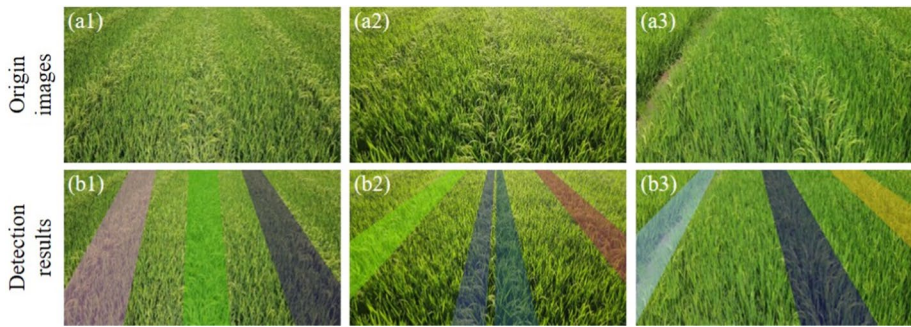
**Table 2** IoU obtained by CNN-based region segmentation models and the proposed method in hybrid rice row images

	PSPNet	DeepLab V3+	UNet	SegNet	Proposed
IoU	72.4	78.6	82.3	82.5	83.2



**Fig. 15** Comparison of regional detection performance of hybrid rice crops between CNN-based semantic segmentation models and the proposed method. **a1–a4** are test images

boundaries. The overall detection of DeepLab V3+ was less effective. The region segmentation performance achieved by UNet and SegNet was approximately the same. The paternal regions they segmented were relatively complete, but the region boundaries remained concave and uneven. Compared to the bottom of the image, the detection ability of both models degraded at the top. The prediction masks of both models were often discontinuous at the top of the image, which could be caused by the narrow crop region at the top and the consequent blurring of the region boundaries.



**Fig. 16** Typical failure detection cases of the proposed method. **a1–a3** are test images of hybrid rice rows

Compared with other crop region segmentation methods, the paternal area obtained by the proposed method was complete, with a more regular shape and precise boundaries. The proposed method was specifically designed for the geometry of hybrid rice rows in the image, taking fully into account the perspective imaging characteristics and the crop row arrangement pattern. The method presented utilized the gradient law between the inclination angles of the crop rows as a constraint to guarantee basically correct paternal region detection results. Afterwards, a fine-tuning step was used to modify the initial detection area to ensure the reliability of the final segmentation result.

### Failure detection cases and cause analysis

Several major failure cases that occurred during region detection of hybrid rice crops are illustrated in Fig. 16. The failure scenario analysis aids in the further optimization of the suggested method.

In Fig. 16a, the proposed method made an error in detecting the central crop area, and the location of the vanishing point was thus shifted significantly, leading to the failure of subsequent crop area scanning and crop boundary fine-tuning. Since the planting of crop rows did not perfectly fit the geometric model in Fig. 2, the location of the vanishing point was not strictly fixed. Usually, when the position of the detected vanishing point varied within a reasonable interval, the crop area could eventually be correctly located, thanks to the boundary fine-tuning algorithm. However, in Fig. 16a, due to the combined effect of light conditions and shooting angle, the color of the central paternal area was consistent with the surrounding maternal regions. Blurred region boundaries resulted in false vanishing point detection. Such errors are expected to be avoided by seeking a more stable way of image preprocessing.

In Fig. 16b, the proposed method incorrectly detected the central parent region as two. Since the area of hole-type noise in the paternal region was more significant than the residual pixel noise in the maternal region, when the scanning mask passed through relatively independent large holes of the paternal region, the scanning curve reflected more prominent local extreme points, which led to false detection. False detections like this occur very occasionally, and attempts can be made to improve the problem by correlating the detection results of the before and after frames.

In Fig. 16c, the ridge was identified by mistake as the paternal region. Since the proposed algorithm was designed to maximize the original image details and reduce the image

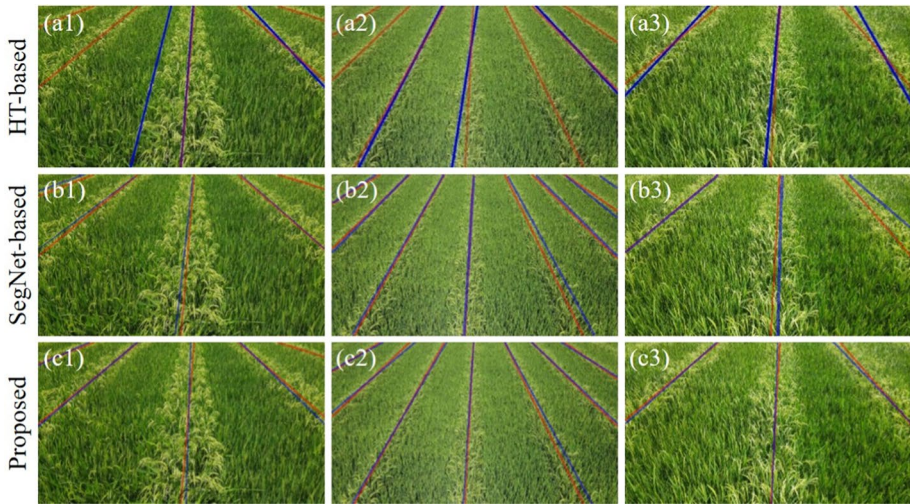


**Fig. 17** Crop row detection algorithm implementation steps: **a** HT-based method, **b** SegNet-based method

processing time, it was assumed that there are only two kinds of object in the image, the paternal regions and the maternal regions. In the future, it is expected to complete the filling of the ridge area in advance by using the double threshold method to eliminate its influence on the detection of hybrid crop areas.

### Evaluation of the performance on crop row extraction

The centerline extraction performance of the approach presented was compared with the Hough Transform-based method (Zhang et al., 2017), one representative crop row detection method based on traditional image processing techniques. Among CNN-type methods, the SegNet with the best crop region segmentation performance was selected for comparison with the proposed approach. When applying the HT-based crop row extraction method, the specific steps of image processing were modified according to the characteristics of hybrid rice at the pollination stage. The implementation process of these two crop row detection methods is illustrated in Fig. 17. Before HT was applied, the crop row image was sequentially subjected to ExR grayscale, Otsu binarization, connected domain denoising, SUSAN corner extraction and DBSCAN crop row clustering. The detected crop rows were limited to the three most salient adjacent rows in the image to ensure HT working performance. When adopting the SegNet for crop centerline extraction, the paternal region masks were predicted first using the trained model. After that, small areas were removed using connected domain denoising, too. Then the least-square method was carried out to fit each single mask area to a straight line representing the centerlines.



**Fig. 18** Comparison of centerline extraction results of hybrid rice crops between the HT-based, SegNet-based and the proposed methods. **a1–a3** images in the test set. Note: Red lines are ground truth lines annotated manually and the detected crop rows are in blue (Color figure online)

The comparison results of crop centerline extraction obtained by the three methods are shown in Fig. 18. When applying the HT-based method to the hybrid rice images, misdetection occurs in Fig. 18a1, and the detection errors near the left and right edges of the image are more significant than those in the central crop region as seen in Fig. 18a3. Pixels belonging to the paternal crop panicles are scattered in the image, and numerous gaps exist in the boundaries of the crop region. The unique morphological characteristics of hybrid rice plants seriously affected the accuracy of the HT-based method, resulting in low centerline extraction accuracy.

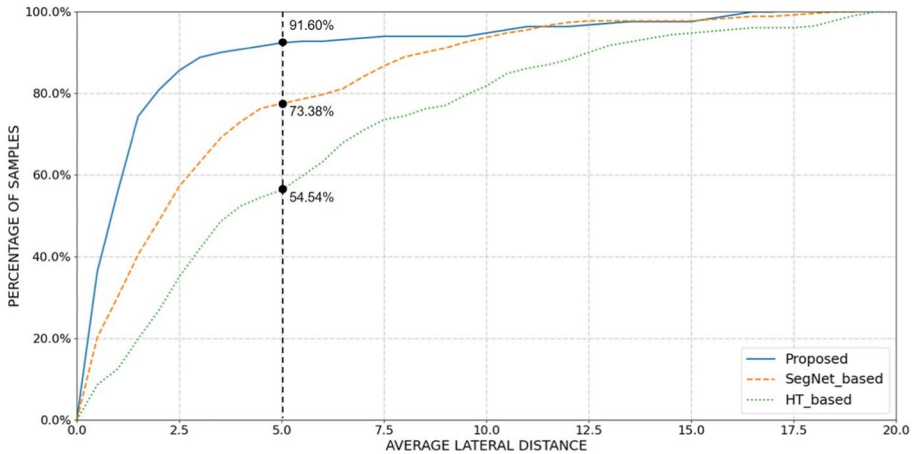
Affected by the uneven boundary and the region detection accuracy dropping at the top of the image, the results of crop row centerline extraction based on the SegNet were not satisfactory. Semantic segmentation models did not take advantage of the planting pattern of hybrid rice crops or perspective imaging property to compensate for detection errors, resulting in sub-optimal centerline detection performance. Aside from the tedious manual labeling process required for training the semantic segmentation models, the slow image inference speed remains a bottleneck that limits their practical application in the navigation tasks for hybrid rice pollination fields.

Compared with the HT-based and the SegNet-based methods, the presented method solved the centerline extraction of hybrid rice crops more accurately. Average values, median values and standard deviations of the average lateral distances of all identified centerlines are presented in Table 3. The numerical statistical results were consistent with the visualization results. The proposed method achieved the best average and median lateral deviations, indicating its superiority over the other two approaches. The standard deviation of the lateral distance of the proposed method was the smallest, which illustrated its stability in performing centerline extraction of hybrid rice crops.

A normalized cumulative histogram of the average lateral distances of all identified centerlines is drawn in Fig. 19. A normalized cumulative histogram is a data representation where the horizontal axis corresponds to values of a measured variable  $x$ , and the vertical

**Table 3** Statistical results of the lateral distance deviation in pixels obtained by the three approaches

	HT-based	SegNet-based	Proposed
Average	5.704	3.555	1.943
Median	3.983	2.097	1.003
Std deviation	4.813	3.648	3.099



**Fig. 19** Normalized cumulative histogram of average lateral distance obtained by the HT-based, SegNet-based and the proposed methods

axis represents the percentage of measurements that are  $\geq x$  (Vidović et al., 2016). It can be intuitively seen that the presented method outperformed the other two. Experiments revealed that crop rows could be considered to be accurately detected when the ALD was less than or equal to 5 pixels on an image with  $640 \times 360$  resolution. The proposed method achieved an ALD of fewer than 5 pixels on 91.60% of the detected centerline pairs. In contrast, the corresponding percentages for the SegNet-based and HT-based methods were 73.38% and 54.54%, respectively. It should be pointed out that the SegNet-based method and the proposed method obtain almost the same IoU, where the former was 0.825 and the latter was 0.832. But the proposed approach was designed for the specific task of crop row detection, which led to better applicability.

### Evaluation of the performance on execution times

Image processing work was executed on a personal computer with an Intel Core i7-8750H 2.20 GHz processor, 16.0 GB of installed random access memory (RAM) memory, running the Windows 10 64-bit operating system. The average execution times on images with a resolution of  $640 \times 360$  for the five main image processing modules and total time in milliseconds were 0.9, 15.3, 20.7, 4.5, 0.8 and 42.2, respectively. Grayscale transformation, boundary position fine-tuning and crop region segmentation were not computationally intensive, while vanishing point detection and crop region identification took a relatively long time due to the utilization of the line-scanning method. As shown in Table 4, the execution time of the proposed method was slightly longer than that of the HT-based crop row



**Table 4** Comparison of time required for single image processing

	HT-based	SegNet-based	Proposed
Time cost (ms)	35.7	104.6	42.2

detection method but much less than that of the SegNet-based approach, although the light-weight feature extraction network MobileNet had been used when SegNet was deployed. The time consumption of the line-scanning operation is expected to be further reduced on parallel architectures such as a graphics processing unit(GPU), which is a noteworthy point for future research.

Restricted by the planting pattern of hybrid rice rows in the seed production field and the scattered growth characteristics of rice panicles during the pollination period, methods based on traditional image processing techniques that could only accomplish crop row centerline extraction and were easily affected by crop growth morphology were not applicable. In terms of accuracy, ease of use and real-time performance, the proposed method was superior to CNN-based methods, making it a better choice for hybrid rice row detection in seed production fields during the pollination period.

## Conclusions

This paper proposed a novel approach for real-time hybrid rice row detection at the pollination stage. The method extended crop row detection from centerline extraction to region identification. A line-shaped mask scanning method was constructed to determine the crop regions based on perspective imaging property. Its hybrid rice region detection performance was superior to CNN-based semantic segmentation methods with an IoU of 0.832, an accuracy of 90.48%, a recall of 86.36%, a precision of 98.96% and an f1-Score of 92.23%. Its centerline extraction ability outperformed the Hough Transform-based and SegNet-based methods in lateral distance deviation between the ground truth line and the detected line. Future work may include the promotion of its crop region detection ability in the upper corners of images, enhancing the stability of image preprocessing, and expanding the image datasets.

**Funding** The authors gratefully acknowledge project funding provided by the Zhejiang key research and development project in China (Grant No. 2022C02005).

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495. <https://doi.org/10.1109/tpami.2016.2644615>
- Bonadies, S., & Gadsden, S. A. (2018). An overview of autonomous crop row navigation strategies for unmanned ground vehicles. *Engineering in Agriculture, Environment and Food*, 12, 24–31. <https://doi.org/10.1016/j.eaef.2018.09.001>

- Chen, J., Qiang, H., Wu, J., Xu, G., & Wang, Z. (2021a). Navigation path extraction for greenhouse cucumber-picking robots using the prediction-point Hough transform. *Computers and Electronics in Agriculture*, *180*, 105911. <https://doi.org/10.1016/j.compag.2020.105911>
- Chen, J., Zhang, D., Zeb, A., & Nanehkaran, Y. A. (2021b). Identification of rice plant diseases using light-weight attention networks. *Expert Systems with Applications*, *169*, 114514. <https://doi.org/10.1016/j.eswa.2020.114514>
- Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 801–818). [https://doi.org/10.1007/978-3-030-01234-2\\_49](https://doi.org/10.1007/978-3-030-01234-2_49)
- Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus—a paradigm for model-fitting with applications to image-analysis & automated cartography. *Communications of the Acm*, *24*(6), 381–395. <https://doi.org/10.1145/358669.358692>
- Gerrish, J. B., Fehr, B. W., Van Ee, G. R., & Welch, D. P. (1997). Self-steering tractor guided by computer-vision. *Applied Engineering in Agriculture*, *13*(5), 559–563. <https://doi.org/10.13031/2013.21641>
- Han, S., Zhang, Q., Ni, B., & Reid, J. F. (2004). A guidance directrix approach to vision-based vehicle guidance systems. *Computers and Electronics in Agriculture*, *43*(3), 179–195. <https://doi.org/10.1016/j.compag.2004.01.007>
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T. et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861).
- Jiang, G., Wang, X., Wang, Z., & Liu, H. (2016). Wheat rows detection at the early growth stage based on Hough transform and vanishing point. *Computers and Electronics in Agriculture*, *123*, 211–223. <https://doi.org/10.1016/j.compag.2016.02.002>
- Jiang, Q., Wang, Y., Chen, J., Wang, J., Wei, Z., & He, Z. (2021). Optimizing the working performance of a pollination machine for hybrid rice. *Computers and Electronics in Agriculture*, *187*, 106282. <https://doi.org/10.1016/j.compag.2021.106282>
- Kakani, V., Nguyen, V. H., Kumar, B. P., Kim, H., & Pasupuleti, V. R. (2020). A critical review on computer vision and artificial intelligence in food industry. *Journal of Agriculture and Food Research*, *2*, 100033. <https://doi.org/10.1016/j.jafr.2020.100033>
- Kim, W.-S., Lee, D.-H., Kim, Y.-J., Kim, T., Hwang, R.-Y., & Lee, H.-J. (2020). Path detection for autonomous traveling in orchards using patch-based CNN. *Computers and Electronics in Agriculture*, *175*, 105620. <https://doi.org/10.1016/j.compag.2020.105620>
- Li, J., Lan, Y., Wang, J., Chen, S., Huang, C., Liu, Q., & Liang, Q. (2017). Distribution law of rice pollen in the wind field of small UAV. *International Journal of Agricultural and Biological Engineering*, *10*(4), 32–40. <https://doi.org/10.25165/j.ijabe.20171004.3103>
- Luo, J. W., Ying, K., & Bai, J. (2005). Savitzky-Golay smoothing and differentiation filter for even number data. *Signal Processing*, *85*(7), 1429–1434. <https://doi.org/10.1016/j.sigpro.2005.02.002>
- Ma, Z., Tao, Z., Du, X., Yu, Y., & Wu, C. (2021). Automatic detection of crop root rows in paddy fields based on straight-line clustering algorithm and supervised learning method. *Biosystems Engineering*, *211*, 63–76. <https://doi.org/10.1016/j.biosystemseng.2021.08.030>
- Meyer, G. E., Neto, J. C., Jones, D. D., & Hindman, T. W. (2004). Intensified fuzzy clusters for classifying plant, soil, and residue regions of interest from color images. *Computers and Electronics in Agriculture*, *42*(3), 161–180. <https://doi.org/10.1016/j.compag.2003.08.002>
- Mousazadeh, H. (2013). A technical review on navigation systems of agricultural autonomous off-road vehicles. *Journal of Terramechanics*, *50*(3), 211–232. <https://doi.org/10.1016/j.jterra.2013.03.004>
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. <https://doi.org/10.1109/tsmc.1979.4310076>
- Pla, F., Sanchiz, J. M., Marchant, J. A., & Brivot, R. (1997). Building perspective models to guide a row crop navigation vehicle. *Image and Vision Computing*, *15*(6), 465–473. [https://doi.org/10.1016/s0262-8856\(96\)01147-x](https://doi.org/10.1016/s0262-8856(96)01147-x)
- Ponnambalam, V. R., Bakken, M., Moore, R. J. D., Glenn Omholt Gjevestad, J., & Johan From, P. (2020). Autonomous crop row guidance using adaptive multi-ROI in strawberry fields. *Sensors*, *20*(18), 5249. <https://doi.org/10.3390/s20185249>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-assisted Intervention* (pp. 234–241). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Vidović, I., Cupec, R., & Hocenski, Z. (2016). Crop row detection by global energy minimization. *Pattern Recognition*, *55*, 68–86. <https://doi.org/10.1016/j.patcog.2016.01.013>
- Wang, A. C., Xu, Y. F., Wei, X. H., & Cui, B. B. (2020). Semantic segmentation of crop and weed using an encoder-decoder network and image enhancement method under uncontrolled outdoor illumination. *Ieee Access*, *8*, 81724–81734. <https://doi.org/10.1109/access.2020.2991354>

- Yu, Y., Bao, Y., Wang, J., Chu, H., Zhao, N., He, Y., et al. (2021). Crop row segmentation and detection in paddy fields based on treble-classification Otsu and double-dimensional clustering method. *Remote Sensing*, 13(5), 901. <https://doi.org/10.3390/rs13050901>
- Zhang, X., Li, X., Zhang, B., Zhou, J., Tian, G., Xiong, Y., et al. (2018). Automated robust crop-row detection in maize fields based on position clustering algorithm and shortest path method. *Computers and Electronics in Agriculture*, 154, 165–175. <https://doi.org/10.1016/j.compag.2018.09.014>
- Zhang, Q., Chen, M. E. S., & Li, B. (2017). A visual navigation algorithm for paddy field weeding robot based on image understanding. *Computers and Electronics in Agriculture*, 143, 66–78. <https://doi.org/10.1016/j.compag.2017.09.008>
- Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2881–2890). <https://doi.org/10.1109/cvpr.2017.660>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.