



Real-time detection of crop rows in maize fields based on autonomous extraction of ROI

Yang Yang^{a,b}, Yang Zhou^a, Xuan Yue^a, Gang Zhang^a, Xing Wen^a, Biao Ma^c, Liangyuan Xu^a, Liqing Chen^{a,b,*}

^a School of Engineering, Anhui Agricultural University, Hefei 230036, China

^b Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China

^c Nanjing Institute of Agricultural Mechanization, Ministry of Agriculture and Rural Affairs, Nanjing 210014, China

ARTICLE INFO

Keywords:

Crop rows detection
ROI
Deep learning
Navigation

ABSTRACT

The current crop rows detection based on machine vision generally has the problems of low detection accuracy and poor real-time performance. Moreover, crop rows detection remains a challenging problem in complex field conditions, such as high weeds pressure, poor illumination conditions, and vegetation foliage shading. We propose a crop rows detection algorithm based on autonomous extraction of ROI (Region of interest). The prior method computes the feature points of the entire image and groups them into the crop rows which they belong to. Instead, we consider the core of crop rows detection problem to be the extraction of the travelling area of agricultural machinery in maize fields. A YOLO (You Only Look Once) neural network is employed to predict the travelling area of the agricultural machinery end-to-end. The prediction boxes are unified into ROI and the crop and soil background are segmented in the ROI by Excess Green operator and Otsu's method. Then, the feature points of crops are extracted using FAST (Features from Accelerated Segment Test) corner point detection, and finally the detection lines of crop rows are fitted with least squares method. Because image recognition is limited to a valid region after the ROI is extracted, the processing speed of our algorithm is remarkably fast. It takes only about 25 ms to process a single image (640*360 pixels) and the frame rate of video stream exceeds 40FPS. Meanwhile, it can achieve high accuracy and robust extraction of ROI in various maize fields. The average error angle of the detection lines is 1.88°, which can meet the real-time and accuracy requirements of field navigation. The proposed algorithm can provide a new solution to the current machine vision-based navigation technology for agricultural machinery. Code is available at: <https://github.com/WoodratTradeCo/crop-rows-detection>.

1. Introduction

With the introduction of precision agriculture and intelligent agriculture, scholars have obtained numerous important results in recent years (Gonzalez-de-Santos et al., 2020; Lakhiar et al., 2018; Li, Chen, Zheng, Dou, & Yang, 2019). In the area of navigation of agricultural machinery, the conventional method is still GPS (Global Position System) based path planning (Xu et al., 2021; Zhang & Qiu, 2004; Zhu et al., 2007). Since crops in the field are typically sown in rows, the path of the agricultural machinery in the field is an S-shaped curve along the crop rows (the path between crop rows is straight and the tractor turns at the

headland), which is not a tough task to achieve. However, GPS-based navigation of agricultural machinery in localized crop rows has the potential to cause seedling injury problems from the wheels due to deviations between the ideal path and the actual crop rows distribution. In the area of precision agriculture, UAVs have received the attention of many researchers for their potential to perform precision operations and remote sensing in the field due to their advantages of low cost, high mobility and portability. Mukherjee, Misra, and Raghuvanshi (2019) conducted a study on segmentation of farming land based on remote sensing system and instance segmentation network U-net. Zhang, Hasanzadeh, Kikkert, Pethybridge, and van Aardt (2021) compared the

Peer review under responsibility of Submissions with the production note 'Please add the Reproducibility Badge for this item' the Badge and the following footnote to be added: The code (and data) in this article has been certified as Reproducible by the CodeOcean: <https://codeocean.com>. More information on the Reproducibility Badge Initiative is available at <https://www.elsevier.com/physicalsciencesandengineering/computerscience/journals..>

* Corresponding author at: School of Engineering, Anhui Agricultural University, Hefei 230036, China.

E-mail addresses: yy2016@ahu.edu.cn (Y. Yang), xlyjwh@ahu.edu.cn (L. Xu), lqchen@ahu.edu.cn (L. Chen).

performance of SfM (Structure-from-Motion) using multiple cameras for perception and LiDAR in UAS (Unmanned aerial system). More competent than large agricultural machinery, UAVs are more capable of global analysis of the field environment in medium-term plant protection operations, such as monitoring and modeling of crops (Mesas-Carrascosa, 2020). However, the sensing of farming land from the air provides no real-time effectiveness for localized agricultural activities on the ground. Based on a machine vision system, Czymbek, Schramm, and Hussmann (2020) propose to use digital image processing techniques to implement crop rows detection tasks for guiding agricultural robots in the field. However, this approach still needs to be improved in real-time performance when it is actually implemented in agricultural production activities. And the simultaneous work of multiple agricultural equipment leads to a greater workload while the operability decreases.

Therefore, the application of machine vision-based crop rows detection on unmanned agricultural machinery has received more and more attention in recent years (Liao et al., 2020; Liu et al., 2016). The core idea of machine vision-based field navigation for unmanned agricultural machinery is to install a camera in front of the wheel to plan precise paths between crop rows in real time, keeping the wheels from causing mechanical damage to the crops. And numerous crop row recognition algorithms have been proposed successively. We have summarized the development of crop rows detection algorithms in recent years, and the popular algorithms can be mainly divided into the following categories.

1) Methods based on Hough transform: Hough transform (Hough, 1962) is one of the most commonly used methods for line detection in the crop rows detection area. Astrand and Baerveldt (2005) gave an outlook on intelligent agricultural machinery, and they performed Hough transform to extract crop rows detection lines for navigation operations of agricultural machinery. To surmount the problem of non-parallel crop rows in the image, Bakker, van Asselt, Bontsema, Muller, and van Straten (2010) first corrected the image and subsequently segmented the crop and soil to obtain a grayscale map. The grayscale image was divided into three parts to calculate the Hough transform, and they are stitched together into a single image afterwards. The essence of this method is to determine the region of interest of the image in segments. However, the degree of correction varies for field images taken at different camera installation angles, so its adaptability still remains to be developed. Winterhalter, Fleckenstein, Dornhege, and Burgard (2018) proposed a new Patten Hough Transform-based crop rows detection algorithm that is more adaptable and capable of performing crop rows detection and navigation operations of agricultural machinery in fields with prior-free conditions. Ji and Qi (2011) proposed a crop rows detection method based on the randomized Hough transform (Oja, 1993). This method optimizes the classical Hough transform in terms of real-time performance, but the performance of the Hough transform is still not superior enough in operational environments with high real-time requirements and complex field conditions. Therefore, in recent years, more and more scholars have explored the algorithms with higher real-time performance and accuracy.

2) Methods based on linear regression: The advantages of the linear regression method are low computational cost and fast running speed, but its most important link is the extraction of feature points. Zhang et al. (2018) first segmented the image with binarization, followed by dividing the image into several horizontal strips. And the pixel values in each strip were projected vertically, and the feature points were extracted using the positional clustering algorithm and the shortest path method. And they ended up fitting the feature points with the least squares method. Jiang, Wang, and Liu (2015) also determined the real feature points by dividing the horizontal strips and vertical projection operation on the segmented image, and fitted detection lines by least squares method. Si, Jiang, Liu, Gao, and Liu (2010) segmented the crop and soil background by the (G-R) feature operator, located the rising and falling points in each image strip using vertical projection, and finally

performed least squares method to fit the categorized feature points. Ospina and Noguchi (2019) extracted feature points by contour detection algorithm and used least squares method to fit the navigation lines in the field. It can be found that the main work of the linear regression-based method is generally to segment the image into horizontal strips and extract the feature points of the crop rows in each strip. This method is usually fast and precise, but in environments with high weeds pressure or sparse crop distribution, the extraction of crop rows features in the strip becomes extremely difficult because we need to eliminate weed interference and also predict and identify the missing regions of crop rows.

3) Methods based on vanishing points: Pla, Sanchiz, Marchant, Bri-vot, and Computing (1997) presented an outlook for implementing crop rows detection-based navigation of agricultural machinery. In this study a vanishing point detection algorithm was proposed. After the image is segmented into crop and soil, the skeletal features of each crop row are extracted by vanishing points, which are used as a baseline for linear fitting. Then the crop rows are recovered so that the detection is completed. Jiang, Wang, Wang, and Liu (2016) searched for feature points by moving windows and multiple interlaced scanning strategies. The Hough transform was used to detect all possible crop rows. Finally, to exclude the wrong candidate crop rows, vanishing points are calculated by k-means clustering. The approach based on vanishing point is essentially a manual feature engineering, which requires the extraction of the features of crop rows as generalized as possible. Therefore, this method will be more difficult to generalize when dealing with complex field images, and the real-time performance of the algorithm based on vanishing point is always a problem that needs to be solved.

4) Other methods: Garcia-Santillan, Guerrero, Montalvo, and Pajares (2018) proposed a curved crop rows detection algorithm for early growth stages. After color segmentation of the image, starting points of crop rows are detected and micro-ROI is divided to track the crop rows growth trend. Ma, Tao, Du, Yu, and Wu (2021) proposed a crop rows detection method based on linear clustering and supervised learning. The method optimizes the vertical projection method based on horizontal strips division. And the detection lines are extracted by supervised learning. These methods have excellent performance in recent years, but they all calculate all the pixel in the image simultaneously, which has two problems for the feature extraction of image processing: 1) The computational cost of processing the whole image is high, and the real-time performance of the algorithm is affected. 2) In the environment of high weeds density and crop leaves shading each other, calculating the feature points of the whole image and fitting the detection lines is extremely challenging, and the accuracy is hard to guarantee.

The above researches make the navigation technology of agricultural machinery relying on crop rows detection algorithms more and more mature, but currently these conventional algorithms still have problems such as poor real-time performance, low accuracy and poor adaptability, mainly because the most popular feature extraction methods are implemented by many processes working in conjunction with each other. When unpredicted situations are encountered, such as the appearance of objects in the image that have not been previously predicted, sudden changes in illumination and camera shake due to, these incidents can affect some of the processes, and thus the overall algorithm performance is impacted. Deep learning, on the contrary, has an end-to-end characteristic, and the recognition results are insensitive to external shocks and more suitable for target detection in richer and more complex environments.

More notably, ROI extraction is a very effective idea in image processing, which can considerably improve the accuracy and real-time performance of the algorithm. However, the distribution of crop rows in the image is not parallel to each other because of the perspective principle of the camera, which also creates challenges for ROI extraction. Montalvo et al. (2012) successfully detected crop rows with high weeds pressure by determining the ROI of the image and the least squares method. However, the number and distribution of crop rows are

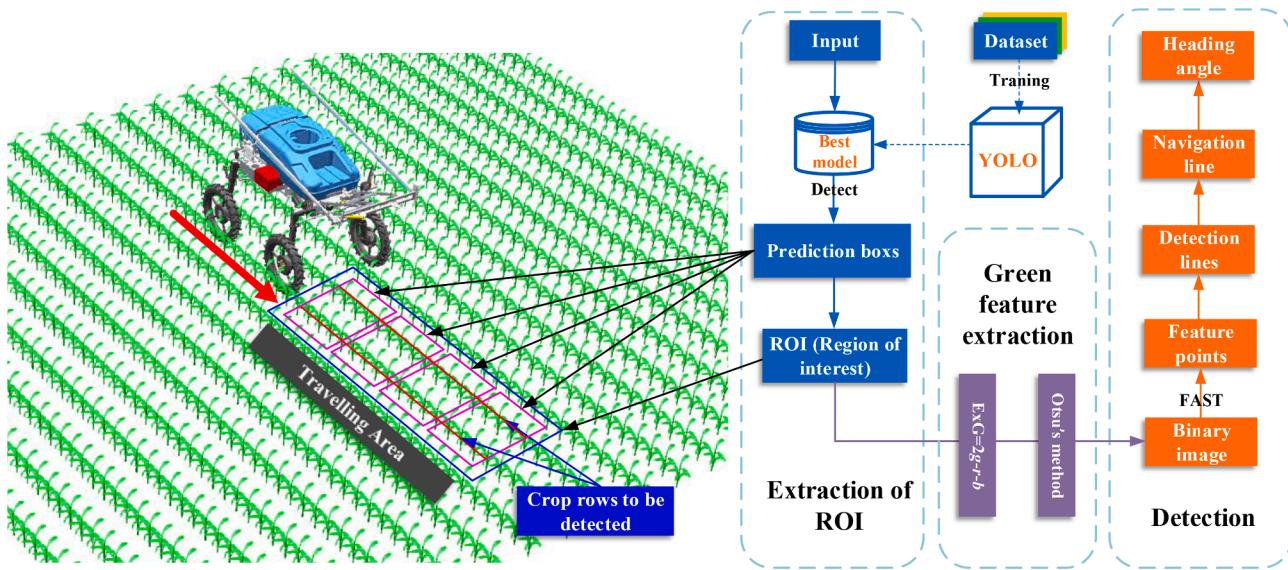


Fig. 1. Flowchart of crop rows detection based autonomous extraction of ROI.

known from previous work. Establishing adaptive ROI is essential for accurate and real-time navigation. Based on the above studies and conclusions, we found that there are not many algorithms for systematic extraction of ROI in recent studies. So, we aim to propose a robust crop rows detection algorithm based on autonomous extraction of ROI. The algorithm should be capable of meeting the following requirements: 1) It can accurately extract the travelling area as ROI in various maize fields end-to-end, and the final recognition accuracy can meet the requirements of navigation. 2) The speed of the whole algorithm should exceed 50 ms per frame (640*360 pixels) to meet the real-time requirements of field navigation. 3) The algorithm is able to show sufficient generalization ability in field images with different growth stages and complex conditions.

2. Materials and methods

2.1. Overview of method

Our method is simple and clear, the overall flow is shown in Fig. 1, which is categorized into the following steps:

1) The travelling area is labeled in segments to make a dataset for training. And the optimal model is found by YOLO (Redmon, Divvala, Girshick, & Farhadi, 2016) deep learning network. The weights of the optimal model are used to predict the prediction boxes of crop rows in

maize fields, locate the travelling areas and determine the ROI.

2) A mask is created depending on the ROI and the green feature of crop rows in the ROI are extracted using Excess Green operator and Otsu's method.

3) After the binary image is obtained after preprocessing, feature points are then extracted using the FAST corner point detection algorithm (Rosten, Porter, & Drummond, 2010). The feature points are classified into different regions in ROI and the detection lines are extracted by fitting the feature points of the crop rows in different regions using least squares method. Finally, the navigation line and heading angle can be derived from detection lines.

2.2. Maize fields information acquisition

The image acquisition location is Anhui Agricultural University's Wanbei Comprehensive Experiment Station in the town of Huigu, Yongqiao County, Suzhou City, Anhui Province (116°97'E, 33°63'N). We used a CMOS camera to collect images. The camera was mounted on the right wheel of agricultural machinery as shown in Fig. 2, 1.5 m above the ground, with an angle of 30°. The image size was 1920*1080 pixels. The frame rate was 12 FPS. The video was saved in AVI format. The distance between crop rows is 60 cm and the speed of the vehicle is 0.5 m/s. To study the crop rows detection problem under complex conditions, we collected three periods of corn crop rows with heights of

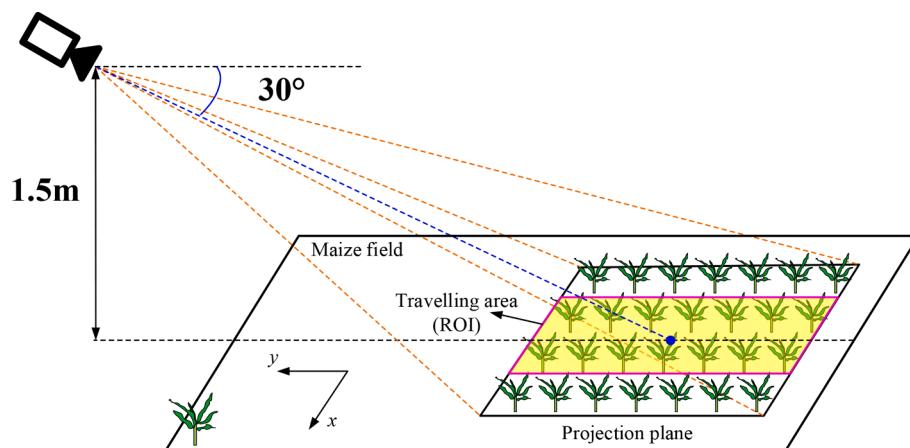


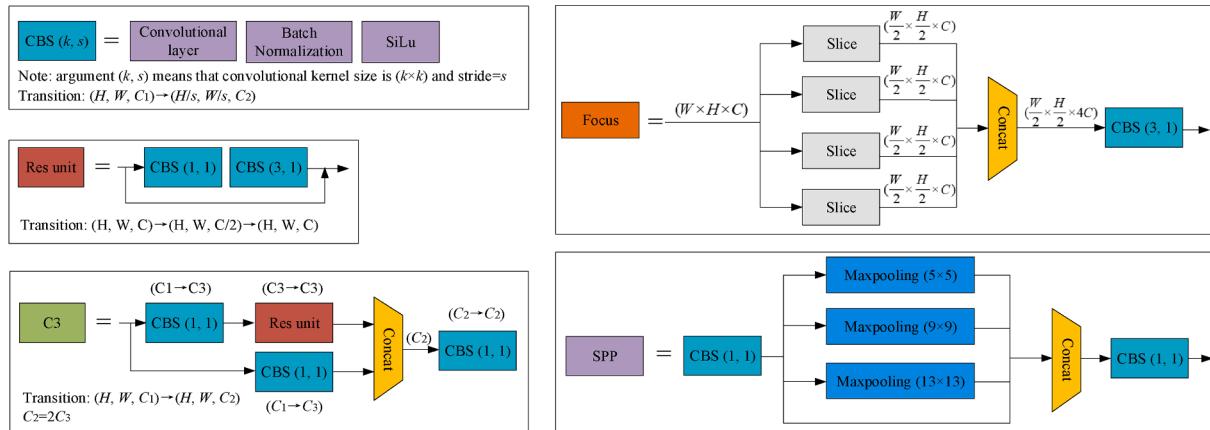
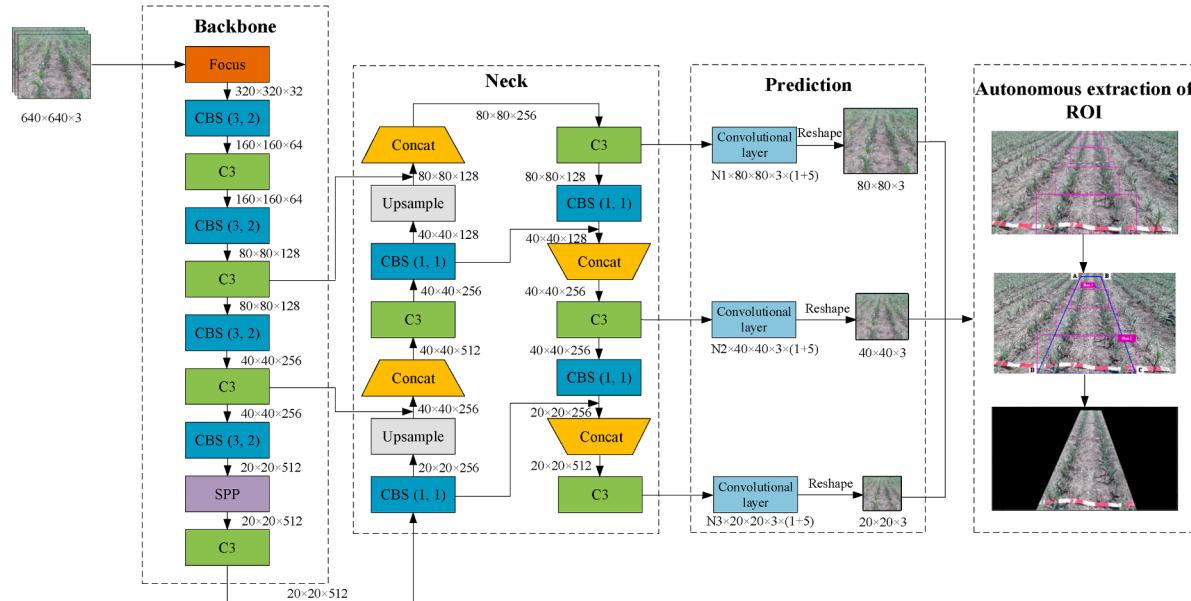
Fig. 2. Installation of the camera.



(a) Period I

(b) Period II

(c) Period III

Fig. 3. Images of maize in different growing periods.**(a) Basic block of YOLOv5 network****Fig. 4.** Autonomous extraction of ROI using YOLOv5 network model.

30 cm, 70 cm and 90 cm respectively under natural illumination conditions, as shown in Fig. 3. The algorithm proposed in this paper is implemented in Python (3.6.6version). The LAPTOP-K8UQ8410 uses an Intel(R) Core (TM) i5-8300H core processor with 2.30 GHz and 8.00 GB of RAM. The camera communicates with the laptop through a network

port.

2.3. Autonomous extraction of ROI using YOLOv5 network

The popular object detection algorithms in recent years include the

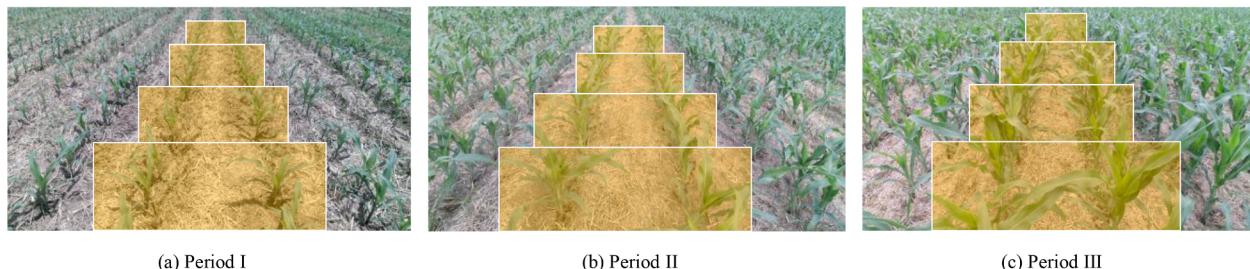


Fig. 5. Labeling of maize in different periods.

R-CNN (Region-Convolutional Neural Network) (Girshick, Donahue, Darrell, & Malik, 2013) series, SSD (Single Shot multibox Detector) (Liu, Anguelov, Erhan, Szegedy, Reed, Fu, & Berg, 2016) and YOLO series. Among them, the two-stage Faster R-CNN (Ren, He, Girshick, & Sun, 2017) object detection algorithm determines the candidate region by RPN (Region Proposal Networks) and then classifies the targets after ROI pooling, which significantly optimizes the performance of R-CNN in term of real-time performance, and it has been able to handle most object detection task. SSD is a one-stage object detection network. Compared with Faster R-CNN, SSD achieves a substantial improvement in real-time performance without a large reduction in accuracy. The one-stage YOLO algorithm considers object detection as a regression problem and performs object detection on the input image end-to-end with 24 convolutional layers and 2 fully connected layers. The detection process of this algorithm is extremely fast and achieves a major breakthrough in real-time performance while ensuring detection accuracy. The network structure and performance of YOLO is getting stronger as the research area of object detection develops. After ResNet (Residual Network) (He, Zhang, Ren, & Sun, 2016) is proposed, YOLO also incorporates the residual module to achieve multi-feature fusion. The detection accuracy and real-time performance are both upgraded. Considering that the work of this paper is navigation of agricultural vehicles based on machine vision, and the real-time performance is crucial for navigation, we choose the YOLOv5 deep learning network with faster computation speed. The basic blocks of YOLOv5 are shown in Fig. 4(a) and YOLOv5 network architecture of extracting ROI is shown in Fig. 4(b).

The backbone of YOLOv5 incorporates the Focus module, which slices the input $640 \times 640 \times 3$ original image to get the feature map, and then undergoes a convolution operation to get the double down-sampling feature map, which improves the perceptual field and reduces the loss of information in the convolutional neural network. The SPP module takes maximum pooling of 5/9/13 and concatenates pooling feature maps at multiple scales to improve the perceptual field. And the Neck part adopts the structure of FPN-PAN, in which FPN fuses the high-level semantic features by up-sampling and low-level features to get the feature map for prediction, and PAN complements FPN to transfer the localization information from shallow to deep layers to enhance the localization ability on multiple scales.

2.4. Training dataset labeling

Due to the perspective principle of the camera, the crop rows in the image are not parallel to each other and cannot be directly bounded by a simple rectangle. The crop rows that can be used for agricultural machinery navigation are mainly distributed in the center region of the image (the travelling area of the agricultural machinery), and the crop rows at the boundary of the image can be considered less helpful for navigation. Therefore, we create the training dataset by segmenting the crop rows in the travelling area. Crop rows of different growth stages are labeled as shown in Fig. 5. The labeling method proposed has the following advantages: 1) The machine learns the distribution pattern between the crop rows and the background, which is not easily affected

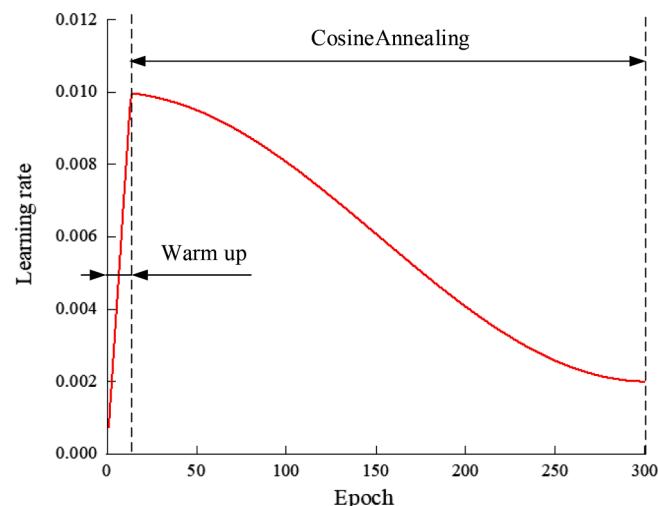


Fig. 6. The learning rate using the cosine annealing decay method.

by different camera angles and crop growth conditions. 2) Only detecting the crop rows in the travelling area can largely eliminate noise interference for the subsequent feature extraction, improve the accuracy of the algorithm, and effectively reduce the running time of the algorithm. 3) Compared to labeling the crop itself, this labeling approach is more efficient. Moreover, under complex field conditions, such as high weeds pressure and leaf-shading environments, the network is trained with the goal of global prediction of field images and is less prone to over- and under-detection.

The labeled image category C is 1 (crop), which is saved in txt format and corresponds to the original image (W^*H), forming the training dataset. Each txt file contains the number of categories of crop rows and the coordinate information of the center point of the bounding box ($x/W, y/H$) as well as the width and height information $w/W, h/H$. A total of 1800 images are labeled in this paper, of which 600 images of each of the three growth periods are included in the dataset. These 1800 images include field conditions with no wind, windy, and high weeds pressure. A random selection of 300 images from the custom dataset is used as a validation set for evaluating the performance of the model only.

2.5. Training

2.5.1. Setting of hyperparameters

The training batch size, momentum decay and weight decay are set to 20, 0.95 and 0.0005, respectively. The learning rate usually needs to be adjusted when training a neural network which is gradually decreased during the traditional training process. We use warm up and cosine annealing as training strategies to adjust the learning rate. 1) Warm up: The model is unstable at the beginning of training, and in order to prevent early overfitting, we use a lower learning rate at the beginning of training to ensure good convergence of the model, and the learning rate will change to the initial learning rate l_0 only after N_i

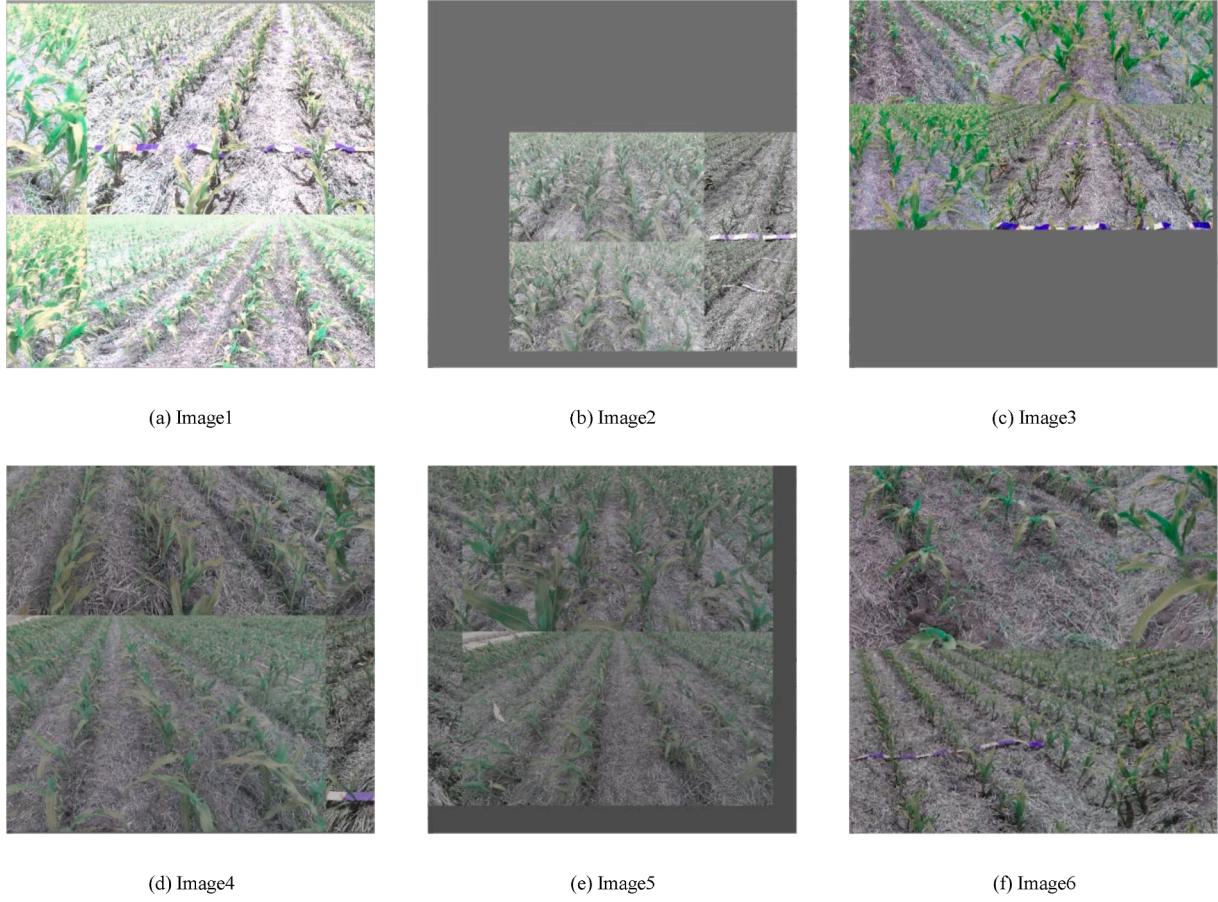


Fig. 7. Training data using data augment.

iterations of the model. N_i is 975. The momentum decay at the warm up stage is 0.8, and l_0 is 0.01. 2) Cosine annealing: When the model converges to the global optimum, we should reduce the learning rate to make the model approach the optimum more easily. The learning rate decreases slowly at the beginning of training as the number of iterations increases, then it accelerates to bring the model closer to the optimum, and then decreases slowly again to prevent the model from oscillating around the optimum. The principle of cosine annealing of the learning rate is shown in Eq. (1). In this paper, the model is trained for 300 rounds and the learning rate variation curve is shown in Fig. 6.

$$l_t = l_{min} + \frac{1}{2}(l_0 - l_{min})(1 + \cos\left(\frac{T_{cur}}{T_{max}}\pi\right)) \quad (1)$$

Where l_t is learning rate, l_0 is the maximum learning rate, l_{min} is the minimum learning rate, T_{cur} is the current iteration number, T_{max} is the maximum iteration number. In this paper, l_{min} is taken as 0.002.

2.5.2. Data augment

For deep neural networks, using 1500 images as the training set is far from enough, and too small amount of data tends to cause overfitting of the model. We first use Mosaic data augment and randomly change the hue (range of variation: 0.975–1.015 times), saturation (range of variation: 0.3–1.7 times) and brightness (range of variation: 0.6–1.4 times) of the images within a certain range to expand the training set. In order to ensure that the model does not see the exact same image twice in each training epoch and improve the generalization ability of model. The Mosaic data augment method crops four images randomly and then stitches them into one image. This method enriches the background of the image and also increases the training batch size in disguise, which reduces the training cost and improves the training efficiency. In this

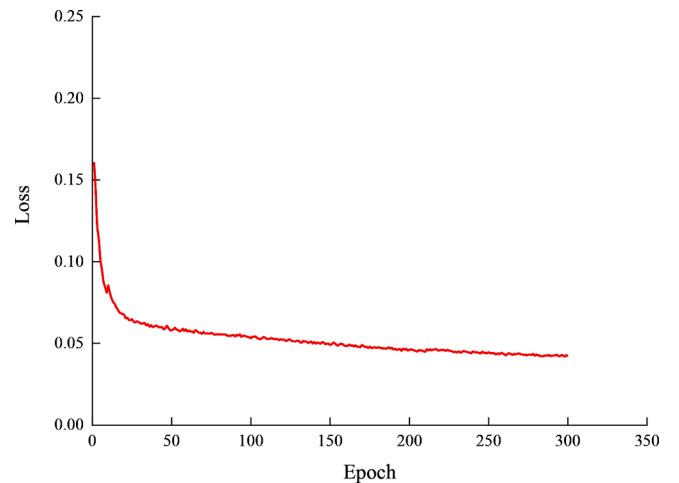


Fig. 8. Loss variation curve.

paper, the model is trained for 300 epochs, with 22,500 iterations, and a total of 450,000 images are used in the training process. Some of the transformed images after the data augment operation are shown in Fig. 7. The gray area in the figure is the filled area after random image stitching, and the network does not learn the information in it.

The variation of the loss value of training is shown in Fig. 8.

2.5.3. Selection of model

A deep neural network model does not perform better with more iterations, and continuing to train the neural network for a large number

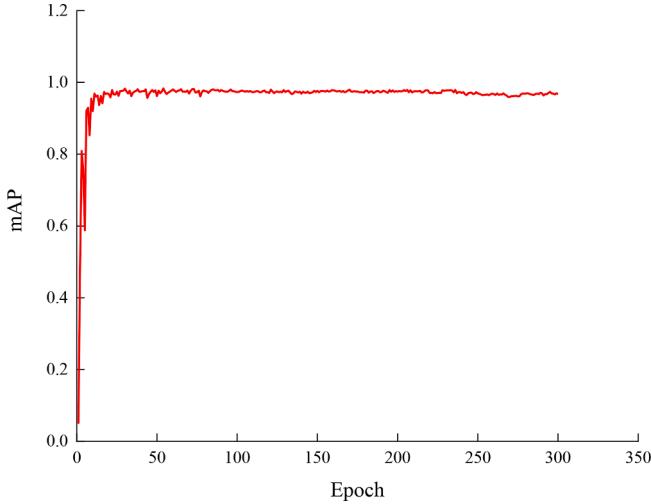


Fig. 9. mAP curve.

of iterations after it has reached stability can cause overfitting of the model. After the training, the optimal model needs to be selected for prediction on the test set. The performance measures are precision, recall, F1 score, IOU and mAP. In this paper, the performance of the model is evaluated by mAP (Mean average precision). mAP contains the information of precision and recall, which is calculated as follows:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$$AP = \int_0^1 P(R)dR \quad (4)$$

$$mAP = \frac{1}{|C|} \sum_{i=C} AP_i \quad (5)$$

where P is the precision rate, R is the recall rate, TP is the number of samples correctly classified as positive, FP is the number of samples incorrectly classified as positive, and FN is the number of samples incorrectly classified as negative.

The mAP values of the model for each epoch of the training process are shown in Fig. 9. From the mAP curve, we can see that the mAP value is gradually stabilized after 100 epochs, so we need to choose the optimal model between 100 and 300 epochs. mAP reaches the maximum value of 97.98 % at the 170th epoch, so the model at this time is chosen as the optimal model.

2.6. Object detection and determination of ROI

Subsequently, we will use the optimal model to inference the field image and output the prediction boxes of travelling area end-to-end. The results are shown in Fig. 10.

From the results, it can be seen that the prediction boxes for travelling area are extracted successfully from the field with different illumination conditions, weeds pressure and growth periods. However, the mutually independent predictor boxes are not favorable for the overall extraction of crop rows, so we have to unify the prediction boxes as the

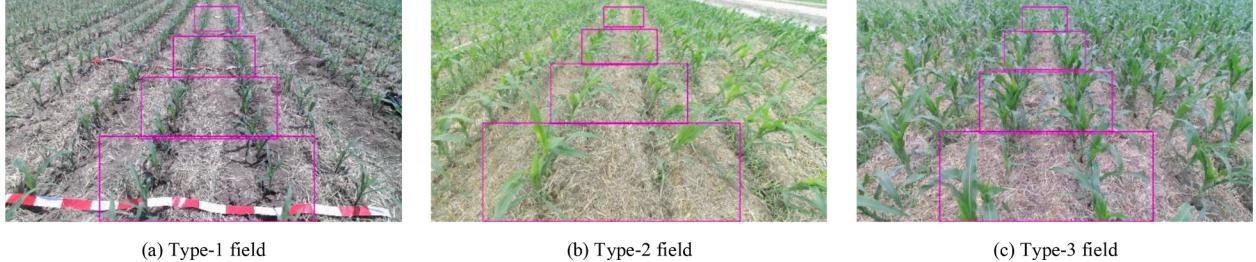


Fig. 10. Travelling area prediction results based on YOLOv5 network.

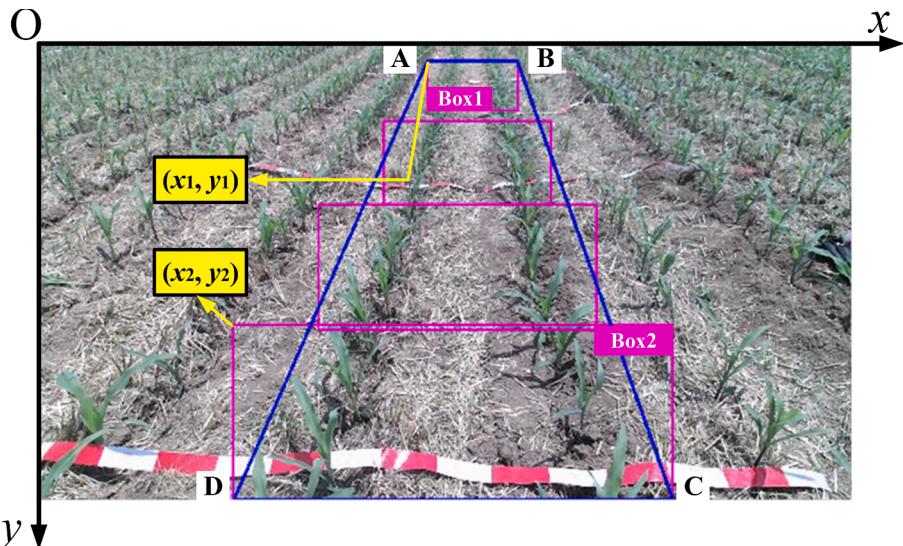


Fig. 11. Determination of ROI (The pink rectangular boxes are prediction boxes; the blue trapezoid box is ROI).



Fig. 12. Extraction of travelling area using ROI mask.

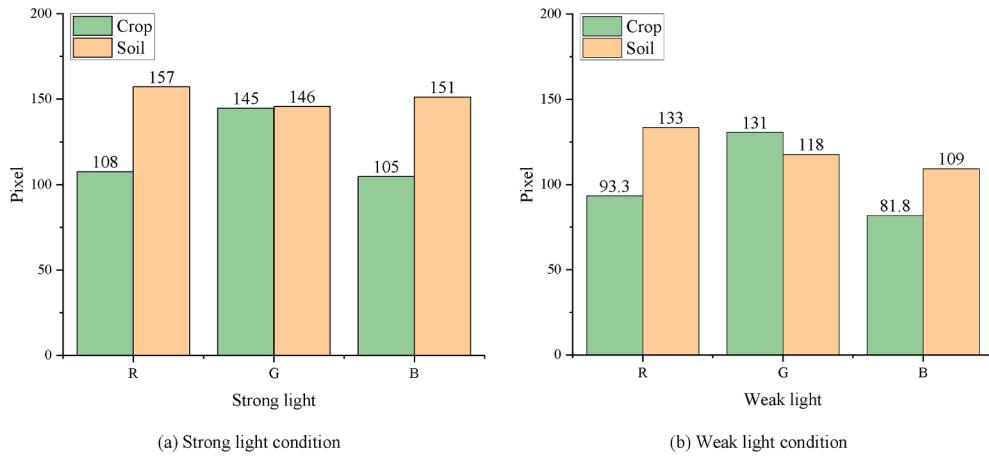


Fig. 13. RGB components statistics of soil and crop under different light conditions.

ROI to provide a reliable preparation for the subsequent feature extraction work. The solution is to record the uppermost prediction box of the image as $box1$ and the lowermost prediction box as $box2$, then the coordinates of the upper left corner, the width and height information of $box1$ is (x_1, y_1, w_1, h_1) , and similarly the position and size information of $box2$ is (x_2, y_2, w_2, h_2) . Then the trapezoidal ROI is divided as shown in Fig. 11. The coordinates of the ROI position are expressed in the following equation:

$$\left\{ \begin{array}{l} (x_A, y_A) = (x_1, y_1) \\ (x_B, y_B) = (x_1 + w_1, y_1) \\ (x_C, y_C) = (x_2 + w_2, y_2 + h_2) \\ (x_D, y_D) = (x_2, y_2 + h_2) \end{array} \right. \quad (6)$$

This ROI extraction method has good robustness, more complete information extraction and lower computational cost. We inevitably introduce the land and soil information on the outside of the crop rows when labeling the bounding boxes, and this additional information is removed once and for all after dividing the ROI. It is also notable that if the field environment is complex and the prediction boxes are missing, the ROI extraction performance is still not greatly affected because the missing prediction boxes in the middle region do not affect the ROI determined by the top and bottom prediction boxes, while if the top or bottom boxes are missing, the prediction boxes that should be in the middle region will replace them. The result is that we get a smaller ROI, but can still predict the overall crop rows growth trend. After determining the location of the ROI, a mask with the same parameters as the ROI is created and a logical bit operation is performed on the original

image to extract the crop rows image in the travelling area. The results of ROI extraction are shown in Fig. 12.

2.7. Image preprocessing

Accurate segmentation of crop and soil is essential for subsequent image processing, and the selection of a suitable color space becomes a major issue. The R, G, B color space is commonly used to extract green plant (Burgos-Artiz, Ribeiro, Guijarro, & Pajares, 2011), and in field images we can clearly identify the color difference between crop and soil, where the green component of crop is much larger than that of soil. To study the precision segmentation between soil and crop, the R, G, B components are counted. We selected 40 frames from each of the maize field videos under strong and weak light conditions, and 10 sampling points for each crop and soil are selected in a single frame. The R, G, B components of the sampling points under strong and weak light conditions are statistically calculated, and the results are shown in Fig. 13.

The statistical chart shows that the color component pattern between soil and crop has a significant difference between the R and B channels, while the difference between the G channels is not obvious. The main reason for this is that the overall value of the R, G, B component of the soil is higher than that of the crop, resulting in a G component that is supposed to be higher in the crop pixels being similar to the soil. So, in order to unify the distribution pattern of color components, The R, G, B channels first need to be normalized as follows:

$$b = \frac{B}{B + G + R}, g = \frac{G}{B + G + R}, r = \frac{R}{B + G + R} \quad (7)$$

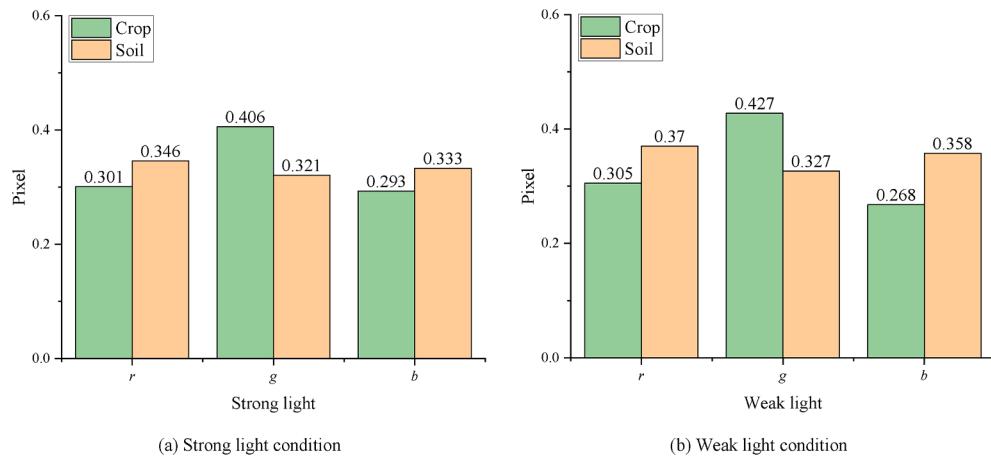


Fig. 14. Statistics of normalized color components of soil and crop under different light conditions.

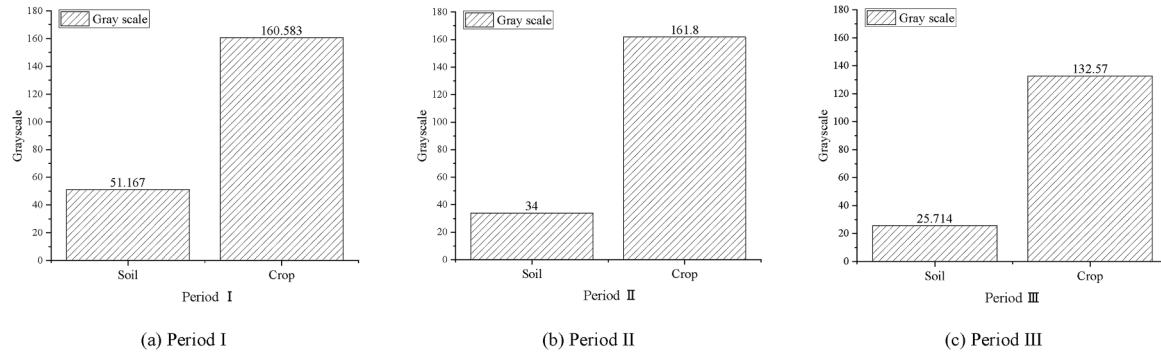


Fig. 15. Statistics of pixel values of grayscale field images of the different growth period.

where B , G , and R are image color components; b , g , and r are normalized values of B , G , R .

The normalized color component statistical results for soil and crop are shown in Fig. 14.

From the results, it is evident that there is a significant difference between the green component of crop and soil, and the G component of the crop is significantly higher than both the R and B . It is also worth mentioning that multispectral cameras are often used in the area of remote sensing to obtain image information of farmland. With the addition of data in the near infrared red band, the field images can be segmented using different vegetation indices (Cavaliere & Senatore, 2022). However, because of its higher equipment cost and longer image processing time, its real-time performance is affected when deployed in mobile devices such as unmanned agricultural machinery, so only RGB images are processed in this paper.

In the area of image processing, the Excess Green algorithm (ExG)

(Woebbecke, Meyer, Bargen, Mortensen, & o. A, 1995) is often used for real-time segmentation of vegetation. Burgos-Artizú et al. (2011) conducted a systematic study of the RGB coefficients in this method to segment crops in maize fields with high weeds pressure and proved its effectiveness. Because of the different external conditions such as illumination and soil background, the classical EXG algorithm is chosen to be performed in this paper. Excess Green operator is calculated as following equation:

$$E(x, y) = 2g - r - b \quad (8)$$

where E is the grayscale value of pixel point (x, y) .

The grayscale image can be obtained by reverting the normalized pixel values. The pixel points of the field images of three growth stages calculated by the Excess Green operator are sampled, where each group of growth stage features contained 20 soil pixels and 20 crop pixels. The

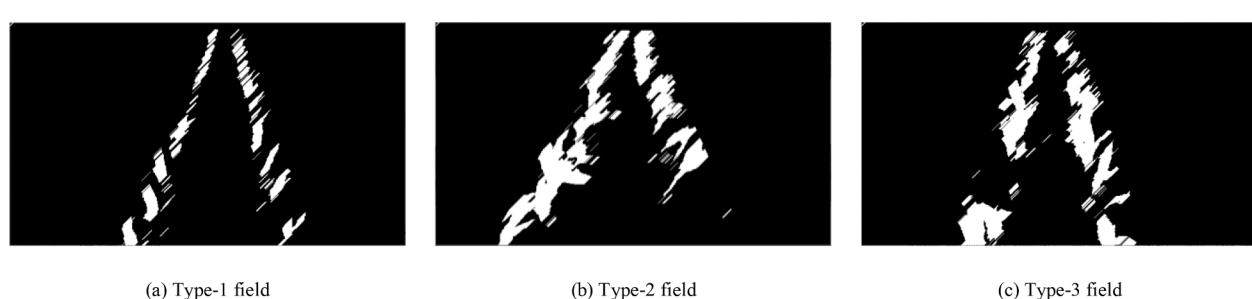


Fig. 16. Binary images obtained by Excess Green algorithm and Otsu's method.

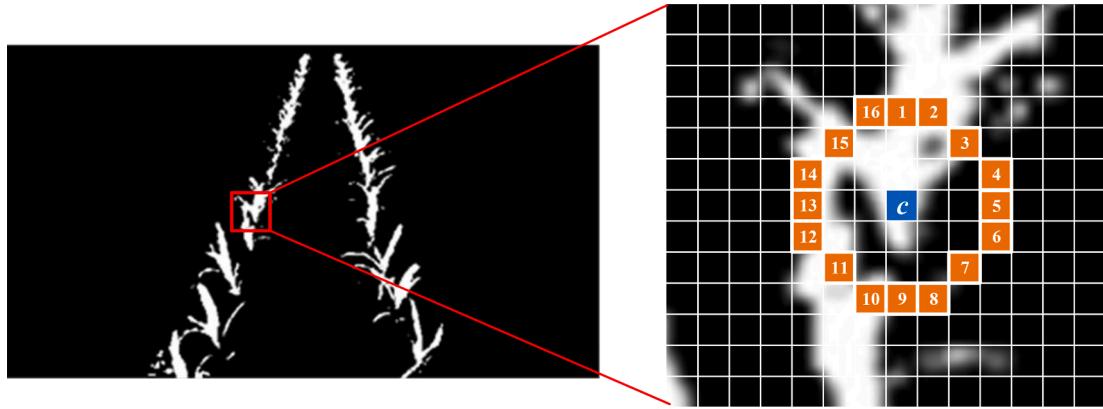


Fig. 17. Feature points extraction based on FAST algorithm.

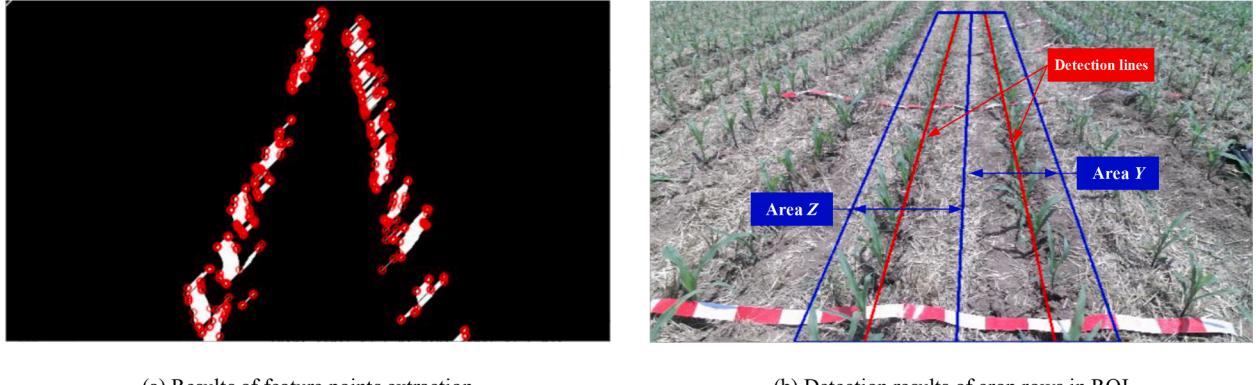


Fig. 18. Crop rows detection using least squares method.

average gray values are counted, and the results are shown in Fig. 15.

From the results, it is clear that the gray values of the crop calculated by Excess Green operator are much higher than those of the soil, which is reliable enough for subsequent segmentation.

Then we performed a binarization operation on the images using the Otsu's method (Otsu, 2007), and to further reduce the impulse noise in the image from weeds, a morphological processing erosion operation was performed twice on the image. To supplement the missing valid information, the expansion operation was performed four more times, and the morphological convolution kernel is shown in Eq. (9). The binary images under different field conditions are shown in Fig. 16.

$$\text{Kernel} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad (9)$$

2.8. Extraction of detection lines

After obtaining the binary feature map, the FAST algorithm is used to extract the feature points of the crop rows. It is a corner point detection method with an extremely fast computational speed. And the algorithm can achieve higher accuracy after extracting ROI and background segmentation based on the previous work. In the binary image, the steps of FAST are as follows: 1) We select a pixel c with a pixel value $P(c)$ of 0 or 255. 2) A discretized Bresenham circle with a radius equal to 3 pixels is made at the center of point c . There are 16 pixels on the boundary of this circle, as shown in Fig. 17. 3) These 16 pixels are compared with the center pixel, and if the condition of Eq. (10) is satisfied, then point c is a feature point.

$$\begin{cases} n = \sum_{x \in \text{circle}(c)} (|P(x) - P(c)| \geq \varepsilon) \\ n > t (0 \leq t \leq 16) \end{cases} \quad (10)$$

Where n is the number of pixel points that satisfy the condition, x is any pixel point on the circumference of a circle, $P(x)$ is the pixel value of any point on the circumference of the circle, $P(c)$ is the pixel value of the center of the circle, ε is the pixel threshold, ε is 255 in this paper, t is the quantity threshold, t is 12 in this paper.

The crop feature points are obtained by FAST, and the results are shown in Fig. 18(a). These points are stored in the point set K , and the ROI is subsequently divided into two regions. The method is as follows: First we connect the line segment AB's midpoint M and the line segment CD's midpoint N. Then we call the closed region consisting of points AMND as region Z and the closed region consisting of points BMNC as region Y. The points belonging to region Z in point set K are stored in point set K_Z and the points belonging to region Y in point set K are stored in point set K_Y . Considering the speed and accuracy of calculation, the detection lines are fitted using least squares method in K_Z and K_Y respectively, and the results are shown in Fig. 18(b).

3. Results and discussion

3.1. Calculation of heading path

The angle between the central axis of the chassis and navigation line is the heading angle of the wheel. After extracting the detection lines, we calculate the navigation path of the wheel by the detection lines. The method is to record the upper and lower two endpoints of the detection line on the left (Z area) as (x_t^1, y_t^1) and (x_b^1, y_b^1) . The upper and lower two

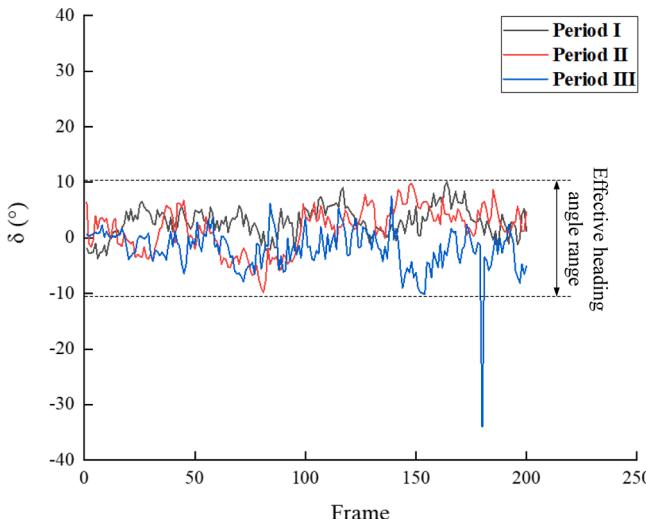


Fig. 19. Variation curves of heading angle in maize fields at different periods.

endpoints of the detection line on the right (Y area) as (x_t^2, y_t^2) and (x_b^2, y_b^2) . Then the two endpoints (x_t, y_t) and (x_b, y_b) coordinates of the navigation line are:

$$\begin{cases} (x_t, y_t) = \left(\frac{x_t^1 + x_t^2}{2}, y_t^1\right) \\ (x_b, y_b) = \left(\frac{x_b^1 + x_b^2}{2}, y_b^1\right) \end{cases} \quad (11)$$

The final output of the algorithm for the wheel heading angle is then calculated as:

$$\delta = \left(\frac{\pi}{2} + \arctan k\right) \times \frac{180^\circ}{\pi} \quad (12)$$

where δ is the heading angle calculated by algorithm, k is the slope of navigation line.

In order to verify the stability and feasibility of the heading angle output by algorithm proposed in this paper, 200 frames of each of the three different growth periods of maize field videos shown in Fig. 2 are selected. The variation curve of the heading angle is shown in Fig. 19. From the figure, it can be seen that the height of maize in the fields of Period I and Period II is lower, and there is little weed interference. The output heading angle varies between $\pm 10^\circ$, which is in accordance with the local growth rule of the crop rows. The height of maize in Period III is

higher and the leaves are obscured by each other, so it is difficult to extract navigation lines. However, the algorithm proposed in this paper can exclude most of the noise at the edge of the image by selecting ROI, and still can output the heading angles effectively. Although the camera shake causes a large offset of the heading angle at the 180th frame, it is quickly corrected in the next frame. The stability of the output heading angle is generally favorable, which can provide a reliable navigation path for the unmanned agricultural machinery based on machine vision.

3.2. Performance analysis of the algorithm

To visually test the accuracy of detection algorithm, we manually delineated the crop rows in images. Detection lines and navigation line were drawn as shown in Fig. 20. Considering that the objective of the work in this paper is extraction of detection line and navigation lines, and the accuracy of the heading angle is the main factor affecting the performance of field navigation based on machine vision, we define the accuracy as in Eq. (13). Error angle of heading and error angles of the detection lines are respectively assigned a weighting coefficient.

$$AEA = 0.5\Delta\delta + \frac{1}{2N} \sum_{i=1}^N \Delta\theta_i \quad (13)$$

Where $\Delta\delta$ is the deviation between the heading angle calculated by algorithm and the heading angle of drawn line, N is the number of detection lines ($N = 2$), $\Delta\theta$ is the error angle between the detection line calculated by algorithm and detection line.

3.2.1. Experiment of crop rows detection using different network

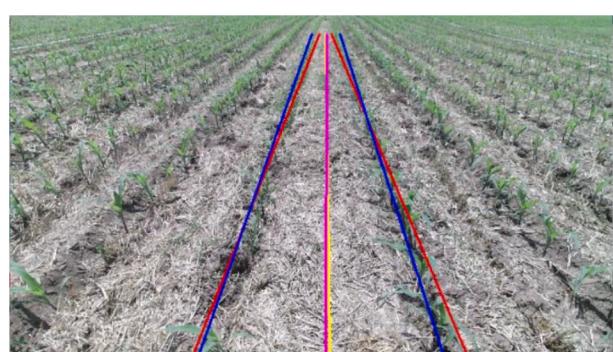
In this paper, YOLOv5 is employed as the object detection network, the detection lines in the validation set are manually delineated in order to investigate its superiority and reliability in the crop rows ROI extraction scenario. The two-stage Faster R-CNN and one-stage SSD network are used for experimental comparison, respectively. Faster R-

Table 1
Performance of different networks in detecting travelling area.

Network	Inference time (ms)/FPS	Mean average precision			Total processing time(ms)/FPS	AEA (°)
		0.5	0.75	0.5:0.95		
Faster R-CNN	232.02/ 4.31	0.983	0.758	0.652	244.82/4.10	2.72
SSD	47.48/ 21.06	0.963	0.630	0.594	61.21/16.34	2.94
YOLOv5	11.33/ 88.24	0.978	0.653	0.616	24.63/40.60	2.85



(a) Drawn detection lines (blue) and drawn navigation line (pink)



(b) Drawn detection lines, detection lines (blue), drawn navigation line and navigation line (yellow)

Fig. 20. Comparison of detection lines, navigation line and drawn lines.

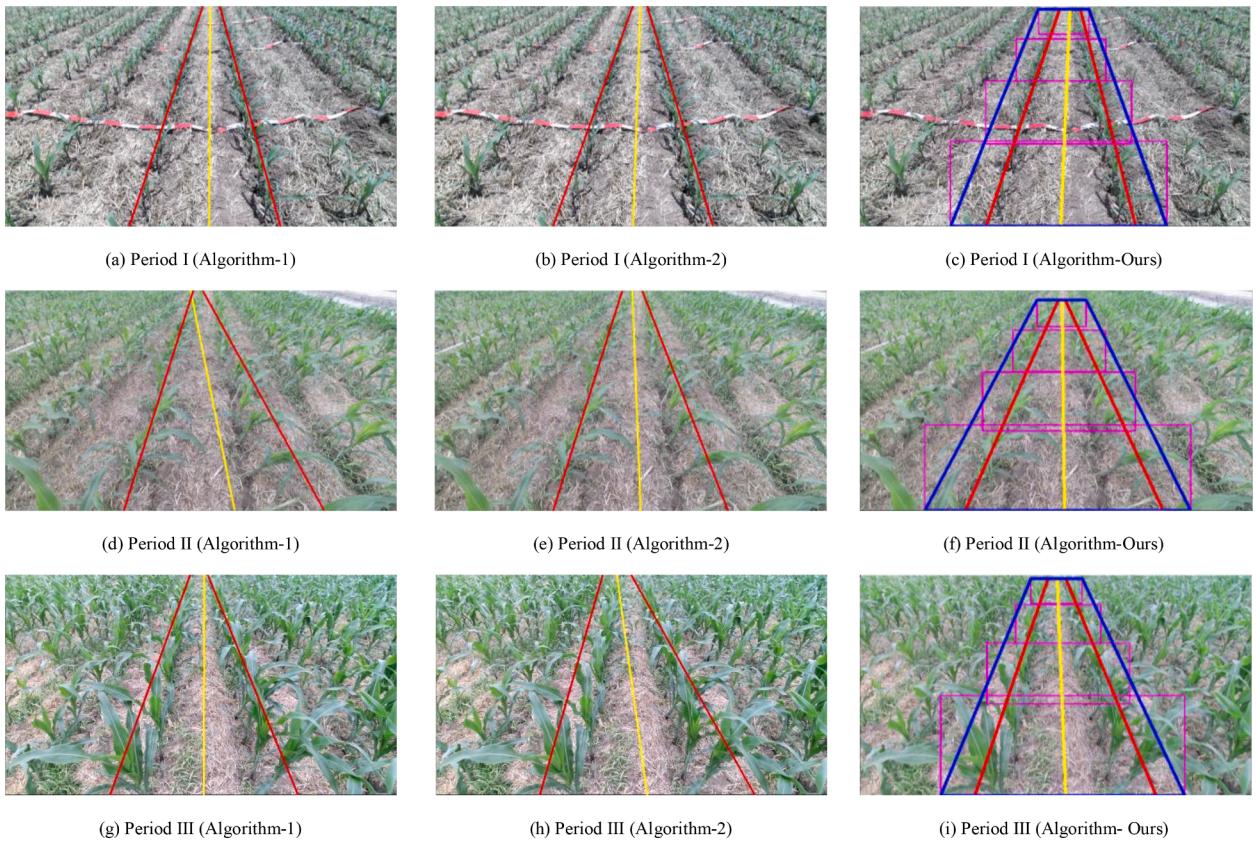


Fig. 21. Crop rows detection results at different growth stage.

CNN employs ResNet-50 as the backbone and SSD employs VGG-16 as the backbone. The training strategies and model selection methods are the same for different networks. And mAP, model inference time, AEA and total computation time of different models in the validation images are listed in [Table 1](#). From the comparison results, it can be seen that Faster R-CNN has more advantages in detection accuracy, which is mainly because the two-stage objection detection network performs the classification of foreground and background in RPN network first, which can effectively avoid the problem of unbalanced positive and negative samples compared with one-stage network. However, the two-stage algorithm takes longer computing time because of its large number of parameters. Faster R-CNN network takes 230.02 ms when inferring a 640×360 field image, which is not ideal for real-time navigation, especially when considering its deployment on mobile devices, the inference time may be longer. Instead, one-stage object detection network needs to perform the localization and classification task in one shot. And SSD as one-stage detector, its real-time performance has been significantly improved. Compared to Faster R-CNN, SSD achieves a 5 times improvement in real-time performance at the expense of 2 % accuracy ($\text{IOU} = 0.5$). YOLOv5 network, because of its lightweight

backbone and the feature fusion module in Neck part, achieves about 20 times better real-time performance with a loss of 0.5 % ($\text{IOU} = 0.5$) accuracy compared to Faster R-CNN. After a comprehensive comparison of the performance between different models, we further investigated the impact of different object detection results on the error of the subsequent image processing part. The AEA of the three different networks is 2.72° , 2.94° and 2.85° , respectively, while the image processing part takes about 13 ms, and a negative correlation can be found between AEA and mAP. Compared with Faster R-CNN, YOLOv5 has 0.13° higher AEA with 0.5 % loss of mAP, which is within a reasonable range of error, while the real-time performance is improved by 10 times. After considering the performance of different networks, it can be proved that the ROI extraction method based on YOLOv5 selected in this paper is reliable enough.

3.2.2. Experiment of crop rows detection at different growth stages

To further verify the performance of the algorithm in this paper, we randomly selected 50 images each of the crop in various environment, which contain three different growth stages, sparse crop distribution environment, windy environment and high weeds pressure environment. The selected images were manually calibrated as described in the [Section 3.2](#). The algorithm proposed in this paper is then compared with the algorithms proposed by [Zhou et al. \(2021\)](#) (Algorithm-1) and [Zhang et al. \(2018\)](#) (Algorithm-2) using evaluation metrics including accuracy and real-time performance.

Experiments were conducted for three growth stages of crop rows, and the detection results of different algorithms are shown in [Fig. 21](#). Because the field environment in Period I and Period II was simple, with little background noise and suitable feature point distribution, all three algorithms could complete the detection work. However, in Period II, because Algorithm-1 is based on the horizontal strip and vertical projection method to extract feature points, the larger leaves will affect the algorithm's prediction for rootstocks of maize. In Period III, because of

Table 2
Performance of different methods in terms of error angle and processing time.

Condition	AEA (°)			Processing time(ms)/FPS		
	Proposed	Zhang et al.	Zhou et al.	Proposed	Zhang et al.	Zhou et al.
Period I	1.03	1.43	2.15	21.38/ 46.77	355.05/ 2.82	175.56/ 5.70
Period II	1.86	4.67	3.20	24.03/ 41.61	382.97/ 2.61	202.39/ 4.94
Period III	0.96	5.65	3.64	25.36/ 39.43	403.21/ 2.48	198.47/ 5.04

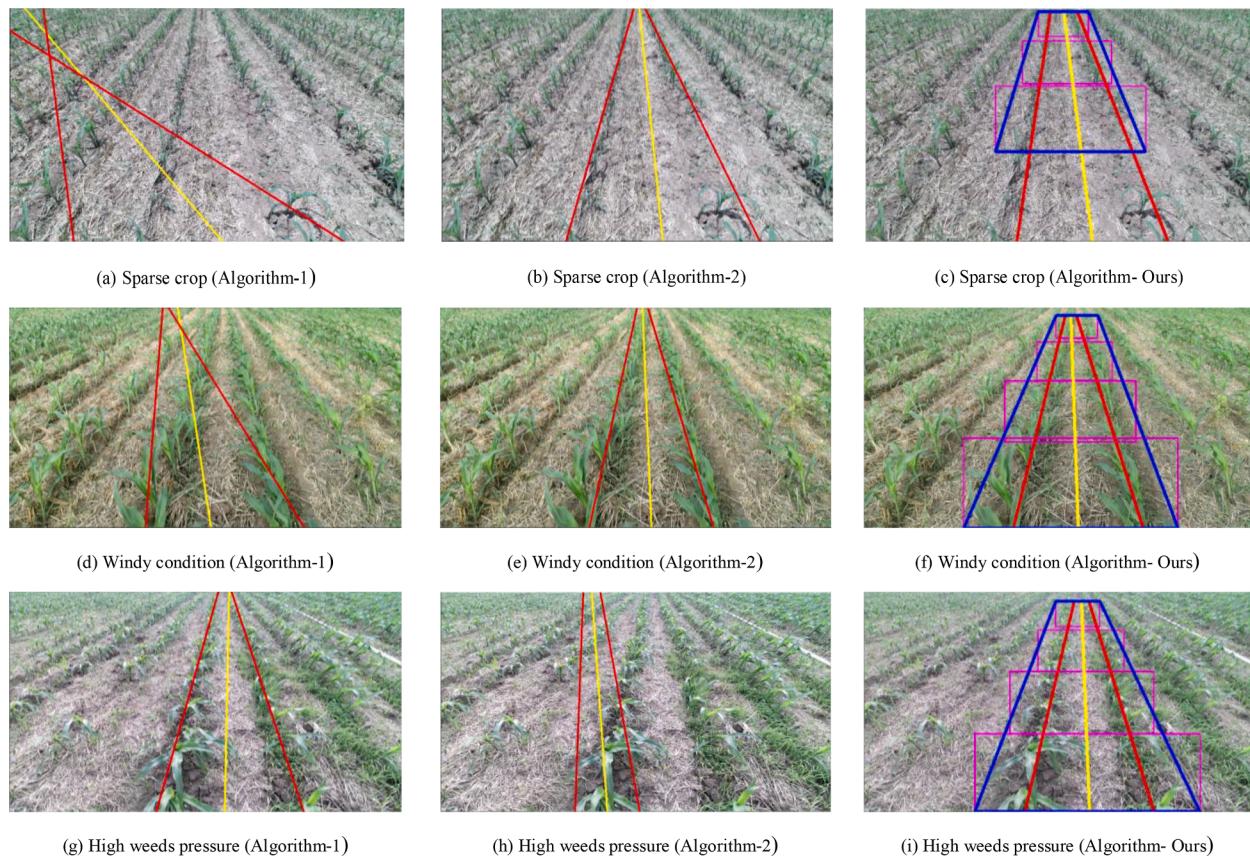


Fig. 22. Crop rows detection results under complex conditions.

the dense crop distribution, high plant height and mutual shading of leaves, Algorithm-2 showed a significant bias in extracting feature points. Instead, the algorithm in this paper was able to accurately retain and predict the rootstock part after determining the ROI, and the leaf information outside the travelling area was removed, so the accuracy could still be guaranteed.

The experimental statistical results of the three algorithms are shown in Table 2. The average error angle of the algorithm proposed in the three periods did not exceed 2° , which was superior in terms of accuracy compared with Algorithm-1 and Algorithm-2. Furthermore, our proposed algorithm performed even better in the dense crop distribution environment of Period III, because it was not affected by the leaves in predicting ROI, and the richer feature pixels provided FAST an advantage in extracting corner points, which theoretically means that the detection lines are less likely to be offset. And the comparison algorithms were susceptible to camera shake when dealing with dynamic video streams. The accuracy was significantly affected in the Period III fields. This also led to the final AEA generally higher than the algorithm proposed in this paper. In term of real-time performance, the algorithm in this paper fully combined the speed advantages of YOLO and FAST corner point detection algorithms, the average time to calculate a 640×360 image frame was about 25 ms, and the processing frame rate of the video stream reached 40 FPS. As for the comparison algorithm, the

average processing time of the algorithms reached 200 ms and 400 ms, respectively, when processing the same image. The reason is that the two methods have to traverse the entire pixel points of the image and calculate the vertical projection of the pixel values. Although the processing time can meet the basic requirements of field navigation, too long processing time is detrimental to navigation when the vehicle speed is high, the field environment is complex and unexpected situations with unprecedented objects arise.

3.2.3. Crop rows detection in complex field environments

For complex scenes, the detection results of the three algorithms are shown in Fig. 22, and the statistical results are shown in Table 3. In more complex environments, the comparison algorithms were not as effective in recognition. In the environment with sparse crop rows, Algorithm-1 was susceptible to the influence of local crop row distribution when extracting ROI based on the horizontal strip method. As shown in Fig. 22 (a), the distribution of crop rows in the figure was sparse and some crops were missing. When extracting the ROI, the Algorithm-1 incorrectly calculated the weeds as real crop rows because the missing crops were not detected in some image strips, resulting in an offset in multi-ROI proposed in the paper. As we mentioned before, the conventional feature extraction algorithm is completed by several working links, and if one of them us malfunctioning, it will cause a strong impact on the

Table 3

Performance of different methods in terms of error angle and processing time in complex conditions.

Condition	AEA ($^\circ$)			Processing time(ms)/FPS		
	Proposed	Zhang et al.	Zhou et al.	Proposed	Zhang et al.	Zhou et al.
Sparse vegetation	2.74	4.81	6.32	23.19/43.12	347.38/2.88	184.47/5.42
Wind	1.67	5.83	8.61	26.59/37.61	365.65/2.73	176.53/5.66
High weeds pressure	3.03	9.08	4.34	25.90/38.61	412.74/2.41	189.48/5.28

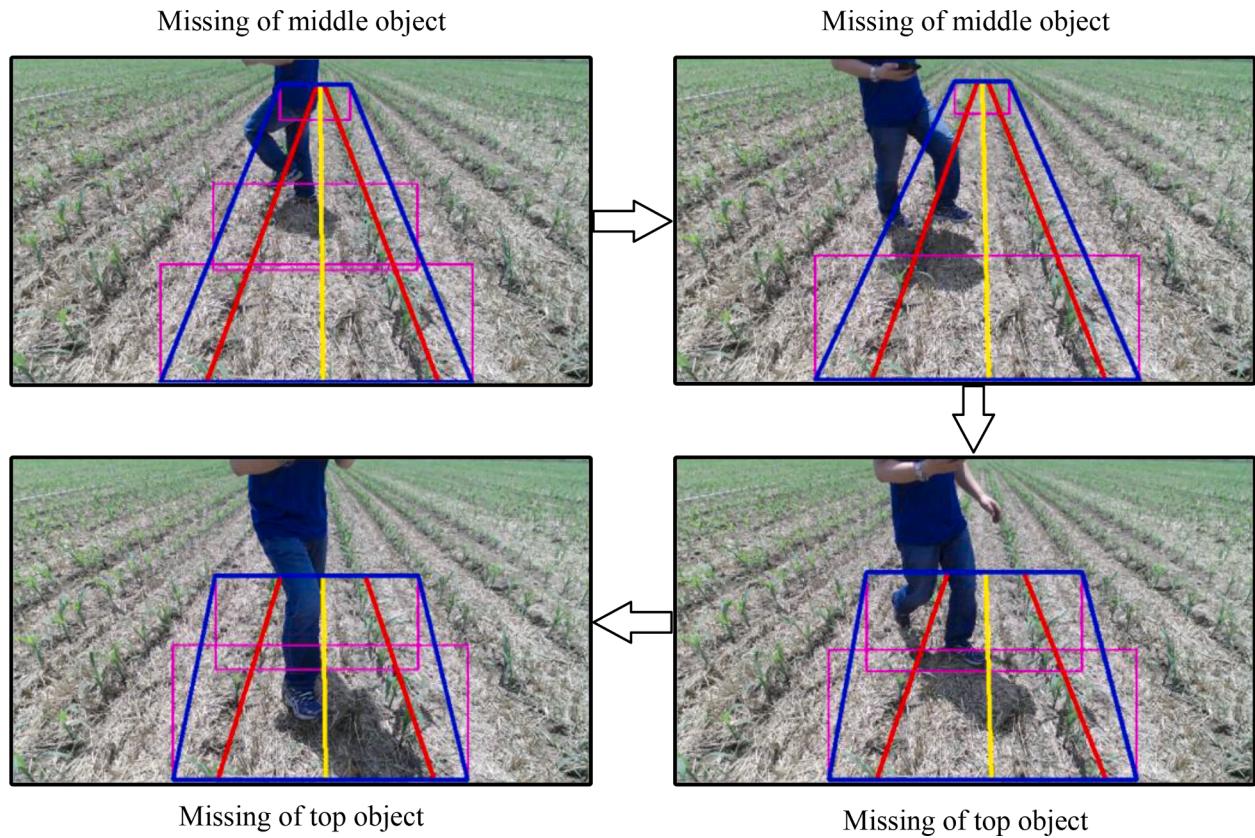


Fig. 23. ROI and detection lines extraction results when the image is occluded.

overall performance of the algorithm, which also leads to very unpromising extraction of the final detection lines. The windy environment in Fig. 22(d)(e) cause the crop to show a tendency to shift to the left. When the rootstock of maize is no longer the center of its highest pixel density and the shifted leaves made the crop features more concentrated at leaves, Algorithm-1 and 2 incorrectly considered the leaves as rootstocks and ended up with an offset in detection lines. In high weed pressure and leaf mutual shading conditions, Algorithm-2 incorrectly recognized weeds as real crop rows when performing vertical projection of pixel points, and the whole algorithm relied heavily on the selection of starting points and tracking of feature points, which ultimately led to unsatisfactory recognition results. In contrast, algorithm proposed in this paper output the travelling area end-to-end by YOLO deep neural network, which was not affected by the distribution of crop rows within the local area, and the FAST corner point detection algorithm was able to extract enough feature points at once, which was less susceptible to interference from missing crop rows and weeds when using least squares compared to Algorithm-1 and Algorithm-2. As we mentioned before the robust ROI extraction method, it is evident in Fig. 22(c) that even if the crop rows in the image were missing, it did not affect the final detection result. Because of the missing objects, we obtained a smaller ROI, but this did not impact the prediction of the crop rows trend. And we will further discuss the robustness of ROI in a later section. In term of real-time performance, algorithm proposed in this paper can achieve real-time output of navigation path and heading angle, which can meet the accuracy and real-time requirements of navigation of agricultural machinery in complex field conditions.

3.2.4. Performance of ROI in the absence of prior knowledge

It is well known that one of the challenging problems in image processing algorithms is to face unexpected situations that have not been considered in previous work. Theoretically, we can continuously enrich and improve the algorithm so that it can cope with most environments.

But this is a tough goal to achieve for the real-time algorithm. Therefore, our proposed algorithm needs to be robust while taking into account the real-time performance, and the most critical core of this paper is the ROI extraction. We analyzed the conditions when the field images are occluded to prove that the ROI extraction method is effective and stable. We let a moving person to simulate the occlusion appearing in the image, and the ROI extraction results are shown in Fig. 23. It can be found that even if the target region in the image was missing, the ROI extraction method proposed in this paper was still able to identify the travelling area. When the target is not detected in the middle part, the prediction boxes at the top and bottom of the image can still determine the location of the ROI. And when the prediction box at the top or bottom is missing, the middle part of the prediction box will replace it. As a result, we get a smaller ROI. Even though the range of ROI is smaller, based on FAST corner point detection, we can still extract enough feature points at one time, which means it does not affect the algorithm's prediction of crop row sowing trend. Therefore, the crop rows detection algorithm based on autonomous extraction of ROI proposed in this paper can achieve stable and accurate identification under unprecedented conditions.

4. Conclusion and future work

A crop rows detection algorithm based on autonomous extraction of ROI is proposed for the problem of low accuracy and poor real-time performance in maize fields with different growth stages and complex environments. Firstly, a dataset of maize field environments with three different growth stages, windy, high weeds pressure and sparse crop distribution is established. A training dataset is made by segmentally labeling the travelling areas of agricultural machinery in maize fields. A YOLOv5 neural network is constructed to train the data and obtain the optimal model to predict the travelling area in maize fields and delineate the ROI, which greatly reduces the interference of weeds and dense

crops at the edges of the images. The Excess Green operator and Otsu's method are used to segment the crop and soil background in the ROI, and morphological operations are performed to eliminate the impulse noise in the binary image. The FAST corner point detection algorithm is used to extract the feature points of crop rows in the binary image, and least squares is applied to fit the feature points and extract the detection lines of crop rows. Finally, the heading angles can be calculated. The algorithm proposed in this paper is compared with two existing methods in recent years to conduct field video experiments under various environments. The experimental results show that the average error angle of the algorithm in this paper are in range of 1.03°-3.03°, the average running time of processing a 640*360 size image is about 25 ms, and the frame rate of video stream processing is over 40FPS, which is significantly improved over the comparison algorithms, and the error is reduced by 50 %. The algorithm proposed in this paper can meet the accuracy and real-time requirements of field navigation, and provides an advanced theoretical basis for the field navigation technology of automatic agricultural machinery based on machine vision. In the future, we will optimize the ROI extraction network in a targeted way, and the feature extraction in ROI will also be done in a neural network. And we will optimize the stability of ROI, the idea is to compare and evaluate the ROI of the previous frame with the ROI of the current frame to ensure that the ROI will not be shifted drastically when the camera shakes.

CRediT authorship contribution statement

Yang Yang: Conceptualization, Methodology, Resources. **Yang Zhou:** Conceptualization, Software, Writing – review & editing. **Xuan Yue:** Software, Visualization. **Gang Zhang:** Software, Data curation. **Xing Wen:** Software. **Biao Ma:** Resources. **Liangyuan Xu:** Supervision. **Liqing Chen:** Supervision, Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors are unable or have chosen not to specify which data has been used.

Acknowledgements

The research in this article was funded by National Natural Foundation of China Youth Fund Project: (51905004), the University Synergy Innovation Program of Anhui Province (Grant No. GXXT-2020-011).

References

- Astrand, B., & Baerveldt, A. J. (2005). A vision based row-following system for agricultural field machinery. *Mechatronics*, 15(2), 251–269. <https://doi.org/10.1016/j.mechatronics.2004.05.005>
- Bakker, T., van Asselt, K., Bontsema, J., Muller, J., & van Straten, G. (2010). Systematic design of an autonomous platform for robotic weeding. *Journal of Terramechanics*, 47 (2), 63–73. <https://doi.org/10.1016/j.jterra.2009.06.002>
- Burgos-Artizzu, X. P., Ribeiro, A., Guijarro, M., & Pajares, G. (2011). Real-time image processing for crop/weed discrimination in maize fields. *Computers and Electronics in Agriculture*, 75(2), 337–346. <https://doi.org/10.1016/j.compag.2010.12.011>
- Cavaliere, D., & Senatore, S. (2022). Incremental knowledge extraction from IoT-based system for anomaly detection in vegetation crops. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 876–888. <https://doi.org/10.1109/jstas.2021.3139155>
- Czymmek, V., Schramm, R., & Hussmann, S. (2020). Vision Based Crop Row Detection for Low Cost UAV Imagery in Organic Agriculture. *Paper presented at the 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*.
- Garcia-Santillan, I., Guerrero, J. M., Montalvo, M., & Pajares, G. (2018). Curved and straight crop row detection by accumulation of green pixels from images in maize fields. *Precision Agriculture*, 19(1), 18–41. <https://doi.org/10.1007/s11119-016-9494-1>
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. J. I. C. S. (2013). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 580–587. doi:10.1109/cvpr.2014.81.
- Gonzalez-de-Santos, P., Fernandez, R., Sepulveda, D., Navas, E., Emmi, L., & Armada, M. (2020). Field Robots for Intelligent Farms-Inheriting Features from Industry. *Agronomy-Basel*, 10(11). <https://doi.org/10.3390/agronomy10111638>
- He, K., Zhang, X., Ren, S., & Sun, J. J. I. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. doi:10.1109/CVPR.2016.90.
- Hough, P. J. U. s. p. (1962). Method and means for recognizing complex patterns. *U.S. patent* 3069654.
- Ji, R. H., & Qi, L. J. (2011). Crop-row detection algorithm based on Random Hough Transformation. *Mathematical and Computer Modelling*, 54(3–4), 1016–1020. <https://doi.org/10.1016/j.mcm.2010.11.030>
- Jiang, G. Q., Wang, X. J., Wang, Z. H., & Liu, H. M. (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, G. Q., Wang, Z. H., & Liu, H. M. (2015). Automatic detection of crop rows based on multi-ROIs. *Expert Systems with Applications*, 42(5), 2429–2441. <https://doi.org/10.1016/j.eswa.2014.10.033>
- Lakhkar, I. A., Gao, J. M., Syed, T. N., Chandio, F. A., Buttar, N. A., & Qureshi, W. A. (2018). Monitoring and control systems in agriculture using intelligent sensor techniques: A review of the Aeroponic system. *Journal of Sensors*, 2018. <https://doi.org/10.1155/2018/8672769>
- Li, Z. Q., Chen, L. Q., Zheng, Q., Dou, X. Y., & Yang, L. (2019). Control of a path following caterpillar robot based on a sliding mode variable structure algorithm. *Biosystems Engineering*, 186, 293–306. <https://doi.org/10.1016/j.biosystemseng.2019.07.004>
- Liao, J., Wang, Y., Zhu, D. Q., Zou, Y., Zhang, S., & Zhou, H. Y. (2020). Automatic segmentation of crop/background based on luminance partition correction and adaptive threshold. *IEEE Access*, 8, 202611–202622. <https://doi.org/10.1109/access.2020.3036278>
- Liú, L., Mei, T., Niú, R. X., Wang, J., Liu, Y. B., & Chu, S. (2016). RBF-based monocular vision navigation for small vehicles in narrow space below maize canopy. *Applied Sciences-Basel*, 6(6). <https://doi.org/10.3390/app6060182>
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. J. S., Cham (2016). SSD: Single Shot MultiBox Detector.
- Ma, Z. H., Tao, Z. Y., Du, X. Q., Yu, Y. X., & Wu, C. Y. (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>
- Mesas-Carrascosa, F. J. (2020). UAS-remote sensing methods for mapping, monitoring and modeling crops. *Remote Sensing*, 12(23). <https://doi.org/10.3390/rs12233873>
- Montalvo, M., Pajares, G., Guerrero, J. M., Romeo, J., Guijarro, M., Ribeiro, A., ... Cruz, J. M. (2012). Automatic detection of crop rows in maize fields with high weeds pressure. *Expert Systems with Applications*, 39(15), 11889–11897. <https://doi.org/10.1016/j.eswa.2012.02.117>
- Mukherjee, A., Misra, S., & Raghuvanshi, N. S. (2019). A survey of unmanned aerial sensing solutions in precision agriculture. *Journal of Network and Computer Applications*, 148. <https://doi.org/10.1016/j.jnca.2019.102461>
- Oja, L. X. J. C. I. U. (1993). Randomized Hough Transform (RHT): Basic Mechanisms, Algorithms, and Computational Complexities.
- Ospina, R., & Noguchi, N. (2019). Simultaneous mapping and crop row detection by fusing data from wide angle and telephoto images. *Computers and Electronics in Agriculture*, 162, 602–612. <https://doi.org/10.1016/j.compag.2019.05.010>
- Otsu, N. J. I. T. o. S. M., & Cybernetics. (2007). A Threshold Selection Method from Gray-Level Histograms. 9(1), 62–66.
- Pla, F., Sanchez, J. M., Marchant, J. A., Brivot, R. J. I., & Computing, V. (1997). Building perspective models to guide a row crop navigation vehicle. 15(6), 465–473.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. J. I. (2016). You Only Look Once: Unified, Real-Time Object Detection.
- Ren, S. Q., He, K. M., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149. <https://doi.org/10.1109/tpami.2016.2577031>
- Rosten, E., Porter, R., & Drummond, T. (2010). Faster and better: A machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1), 105–119. <https://doi.org/10.1109/tpami.2008.275>
- Si, Y., Jiang, G., Liu, G., Gao, R., & Liu, Z. J. N. J. X. T. o. t. C. S. o. A. M. (2010). Early stage crop rows detection based on least square method. 41(7), 163–167+185.
- Winterhalter, W., Fleckenstein, F. V., Dornhege, C., & Burgard, W. (2018). Crop row detection on tiny plants with the pattern hough transform. *IEEE Robotics and Automation Letters*, 3(4), 3394–3401. <https://doi.org/10.1109/lra.2018.2852841>
- Woebbecke, D. M., Meyer, G. E., Bargen, K. V., & Mortensen, D. A. J. T. o. A. (1995). Shape features for identifying young weeds using image analysis. 38(1), 271–281.
- Xu, G. F., Chen, M. Z., He, X. K., Pang, H. X., Miao, H. Q., Cui, P. D., ... Diao, P. S. (2021). Path following control of tractor with an electro-hydraulic coupling steering system: Layered multi-loop robust control architecture. *Biosystems Engineering*, 209, 282–299. <https://doi.org/10.1016/j.biosystemseng.2021.07.014>
- Zhang, F., Hassanzadeh, A., Kikkert, J., Pethybridge, S. J., & van Aardt, J. (2021). Comparison of UAS-based structure-from-motion and LiDAR for structural characterization of short Broadacre crops. *Remote Sensing*, 13(19). <https://doi.org/10.3390/rs13193975>

- Zhang, Q., & Qiu, H. (2004). A dynamic path search algorithm for tractor automatic navigation. *Transactions of the ASAE*, 47(2), 639–646.
- Zhang, X. Y., Li, X. N., Zhang, B. H., Zhou, J., Tian, G. Z., Xiong, Y. J., & Gu, B. X. (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>
- Zhou, Y., Yang, Y., Zhang, B. L., Wen, X., Yue, X., & Chen, L. Q. (2021). Autonomous detection of crop rows based on adaptive multi-ROI in maize fields. *International Journal of Agricultural and Biological Engineering*, 14(4), 217–225. <https://doi.org/10.25165/ijjabe.20211404.6315>
- Zhu, Z. X., Chen, J., Yoshida, T., Torisu, R., Song, Z. H., & Mao, E. R. (2007). Path tracking control of autonomous agricultural mobile robots. *Journal of Zhejiang University-Science A*, 8(10), 1596–1603. <https://doi.org/10.1631/jzus.2007.A1596>