Detection of the foreign object positions in agricultural soils using Mask-RCNN

Yuanhong Li[1,2,3], Chaofeng Wang[1], Congyue Wang[1], Xiaoling Deng[1], Zuoxi Zhao[1], Shengde Chen[1], Yubin Lan[1,2,3,4*]

(1. *South China Agricultural University, College of Electronic Engineering, Guangzhou 510642, China*;
2. *National Center for International Collaboration Research on Precision Agricultural Aviation*, *Pesticides Spraying Technology (NPAAC), South China Agricultural University, Guangzhou 510642, China*;
3. *Department of Biological and Agricultural Engineering, Texas A&M University, College Station, TX, 77843, USA*;
4. *Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou 510642, China*)

Abstract: Objects in agricultural soils will seriously affect the farming operations of agricultural machinery. At present, it still relies on human experience to judge abnormal Gounrd-penetrting Radar (GPR) signals. It is difficult for traditional image processing technology to form a general positioning method for the randomness and diversity characteristics of GPR signals in soil. Although many scholars had researched a variety of image-processing techniques, most methods lack robustness. In this study, the deep learning algorithm Mask Region-based Convolutional Neural Network (Mask-RCNN) and a geometric model were combined to improve the GPR positioning accuracy. First, a soil stratification experiment was set to classify the physical parameters of the soil and study the attenuation law of electromagnetic waves. Secondly, a SOIL-GPR geometric model was proposed, which can be combined with Mask-RCNN's MASK geometric size to predict object sizes. The results proved the effectiveness and accuracy of the model for position detection and evaluation of objects in soils; then, the improved Mask RCNN method was used to compare the feature extraction accuracy of U-Net and Fully Convolutional Networks (FCN); Finally, the operating speed of agricultural machinery was simulated and designed the A-B survey line experiment. The detection accuracy was evaluated by several indicators, such as the survey line direction, soil depth false alarm rate, Mean Average Precision (mAP), and Intersection over Union (IoU). The results showed that pixel-level segmentation and positioning based on Mask RCNN can improve the accuracy of the position detection of objects in agricultural soil effectively, and the average error of depth prediction is 2.87 cm. The results showed that the detection technology proposed in this study integrates the advantage of soil environmental parameters, geometric models, and artificial intelligence algorithms to provide a high-precision and technical solution for the GPR non-destructive detection of soils.

Keywords: foreign object, soil object, position, agricultural soil, Mask R-CNN, GPR image
DOI: 10.25165/j.ijabe.20231601.7173

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

1 Introduction

There are many kinds of agricultural soils with different physical parameters, which makes it difficult to control the recognition accuracy of the object detection algorithm object in soil[1]. Azizi et al.[2] applied semantic segmentation for extracting soil clods from Gounrd-penetrting Radar (GPR) data, but it only used deep pre-trained network and computer vision-based method for the surface of soils, and there was no research on the rocks inside the soil layer. Xu et al.[3] proposed a novel deep learning method for soil crack images, they improved the U-Net to accurately identify soil cracks; their research results showed the value of deep learning models in soil detection. Ma et al.[4] used the YOLO target detection algorithm to detect abnormal objects on urban roads, although the proposed neural network model can perform cavity anomaly testing quickly, the accuracy rate was still low. The insufficient training samples lead to insufficient confidence problems[4]. Dou et al.[5] proposed a novel threshold segmentation method to separate the target in the GPR image from the background, and apply machine learning algorithms to identify the hyperbolic features; the proposed method can successfully identify and fit the hyperbola, and the location、size of the detection target and propagation speed of electromagnetic waves in the medium can be evaluated; however, this method lacks a comprehensive feature study, resulting in the low detection rate of targets with inconspicuous hyperbolic characteristics[5]. In related research, Pham and Lefèvre[6] adapted the Faster Region-based Convolutional Neural Network (Faster RCNN) framework to detect GPR images and pre-trained a CNN method to deal with some specific GPR data, but this method was limited to the scale of sample data. Gong and Zhang[7] proposed a deep-learning-based Faster R-CNN algorithm for GPR images features and they analyzed Faster R-CNN's recognition ability with an average

Received date: 2021-11-06 **Accepted date:** 2022-07-16
Biographies: Yuanhong Li, PhD, research interest: agricultural engineering and machine vision, Email: liyuanhong@stu.scau.edu.cn; **Chaofeng Wang**, MS candidate, research interest: agricultural engineering and information extraction, Email: scauwcf@stu.scau.edu.cn; **Congyue Wang**, MS candidate, research interest: computer science and multi-sensor fusion, Email: wangcongyue@stu.scau.edu.cn; **Xiaoling Deng**, PhD, research interest: agricultural aviation and remote sensing, Email: dengxl@scau.edu.cn; **Zuoxi Zhao**, PhD, research interest: agricultural mechanization engineering, machine learning and data analysis, Email: zhao_zuoxi@scau.edu.cn; **Shengde Chen**, PhD, research interest: agricultural aviation and remote sensing, Email: shengde-chen@scau.edu.cn.
***Corresponding author: Yubin Lan**, PhD, Professor, research interest: agricultural aviation and remote sensing, South China Agricultural University, Guangzhou 510642, China. Tel: +86-13922707507, Email: ylan@scau.edu.cn.

accuracy rate of 93.0%, but this algorithm framework is based on ROI Pooling, the positioning accuracy is easily disturbed. This method adopts selective search with SVM as the core to extract the area to be detected, and for the first time uses bounding box regression to make Bbox correction[8,9]. Actually, Mask R-CNN has made an improvement on Faster R-CNN (2016) that Mask R-CNN adds an additional mask prediction branch and ROI Pooling into its framework[10-12]. Rice et al.[13] proposed a method based on generative adversarial networks to classify objects from B-Scan of GPR images. It improved loss function by merging frequency domain features, but this paper hasn't conducted data research and analysis on the computing power consumed by its model.

Currently, the most of methods are based on target detection for foreign objects in farmland, lacking accuracy when predicting the size, type, and depth of objects. The improved Mask RCNN and a geometric model proposed in this study can improve the GPR positioning accuracy. This model is systematically studied from the mathematical principle and provides an optimized solution for the intelligent detection of foreign objects on agricultural machinery equipped with GPS.

2 Materials and methods

2.1 Experimental materials

The types of objects studied in this research include five kinds: glass, wood, iron, dry hard stone, and PVC plastic (Table 1). These objects composed of these five kinds of materials in the farming soil can damage agricultural machinery. In addition, the volume of objects in soil studied in this study is not less than 500 cm^3, and the density is not less than 2.8-3.1 g/cm^3.

Table 1 Soils objects parameters related to this study

No.	Materials	ε_x	μ_x/S·m^{-1}	Trans speed/m·μs^{-1}
1	Glass	1.1-2.2	0.001–0.010	150–200
2	Wood	2.8-3.0	10^{-4}–10^{-3}	112–122
3	PVC	3.0-8.0	0.010–0.001	170
4	Stone	6.0-12.3	10^{-6}–10^{-8}	134
5	Iron	>500	--	Near light speed

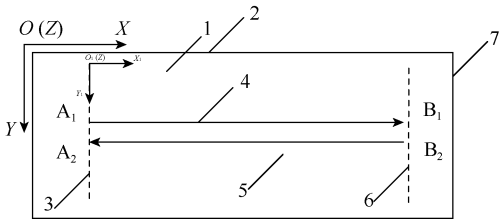
The detection equipment uses the GX750-HDR GPR (MALA Geoscience AB, Swedish). The position detection of objects in soil was tested under three schemes; among them, the first scheme controls the soil moisture content to be different, and ensures that other conditions are the same; The second option is to control the soil conductivity to be different, which corresponded to different types of soil mixing and ensure that others are the same; the third option is to control the soil dielectric constant to be different. The moisture content can be measured with a soil moisture measuring instrument, and the soil conductivity can be measured with a conductivity meter; the soil dielectric constant can be measured with a soil dielectric constant measuring instrument. The soil parameters used in the experiment are three types A, B, and C in Table 2.

Table 2 Three kinds of soil parameters involved in this article

Soil	Moisture content/%			Conductivity/Sm			Dielectric constant		
A	19-20	21-22	23-24	1.4×10^{-4}	1.4×10^{-4}	1.4×10^{-4}	5.4	5.4	5.4
B	19-20	19-20	19-20	1.4×10^{-4}	1.2×10^{-3}	5.3×10^{-3}	5.4	5.4	5.4
C	19-20	19-20	19-20	1.4×10^{-4}	1.4×10^{-4}	1.4×10^{-4}	5.4	7.6	9.2

2.2 Objects position measurement and evaluation method

In this study, the coordinate value was converted to distance to evaluate the position accuracy of the object in soil. As shown in Figure 1, firstly, a global coordinate system *XOY* was established, and then set a local coordinate system $X_1O_1Y_1$ in the soil; The testing Route A-B can be multiple parallel straight lines, and the detecting route of the radar is measured with many times according to its line in the figure. There are three different values for the measurement speed. It should be based on a low speed (20.0 cm/s±0.5 cm/s), medium speed (24.0 cm/s±0.5 cm/s), and high speed (28.0 cm/s±0.5 cm/s), respectively. The selected speed of the test should consider the working environment of the field operation. The test speed can be referenced to the speed of the wheeled tractor in the field for other types of self-propelled agricultural machinery.



1. Soil surface 2. Fixed auxiliary survey line direction reference plane 3. Survey start line 4. Survey line L1 5. Survey line L2 6. Survey termination line 7. Fixed auxiliary reference surface

Figure 1 Design of line A-B experiment

During the experiment, the radar was dragged along the soil surface according to the AB route plan, then marked the starting point of the radar according to the reference point, and artificially bury any number of foreign objects between the AB lines. Generally, the interval between each object is not less than 0.5 m; among them, the position recognized by the machine vision system is generally based on the morphological center of the objects as the recognition reference point. The radar was dragged on the soil surface according to the AB route plan, and then mark the start point of the radar was according to the reference point, generally, the start point will be selected at the position of the first detected target, and any amount of objects is buried between the AB lines for the object in soil, its interval between each object in the measurement route is not less than 0.5 m; the morphological center of the object in soil identified by the intelligent algorithm system will be the identification reference point, and then it was compared with the soil depth scale value of the identification reference point. It will record the performed data of statistics mean and variance. The GPR image reference point selected in the experiment should accurately reflect the error produced by the actual detection process. Unless otherwise specified, the reference point for the radar to start measurement is the centroid of the entire radar instrument generally, and the position where the bottom of the radar is in contact with the soil surface is the position reference point where the depth of the detection target should be zero (Figure 2).

2.3 Geometrical model of object detection

It is supposed that the size of the original GPR image was $m\times n$ and its real size will be $a\times b$. It is denoted the final convolution layer of n-th stage as conv_n after the convolutional layer of Feature Pyramid Network (FPN). Then the i-th row and j-th column (i, j) in convolutional transformed into coordinates on original images by the relationship of side length ratio $\left(\frac{m}{a},\frac{n}{b}\right)$ as

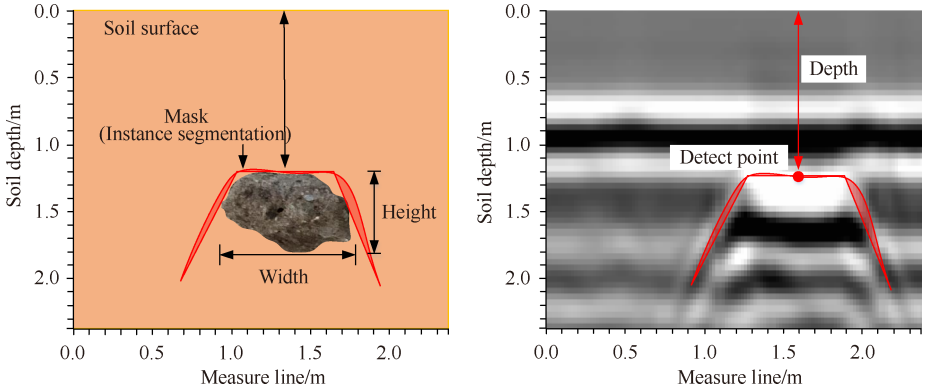
$\left(\frac{im}{a}, \frac{jn}{b}\right)$ [14-16]. It calculated the function for two dimension gray values of (i, j) in a grayscale image, then can be defined as:

$$f(i, j) = w_1p_1 + w_2p_2 + w_3p_3 + w_4p_4 \quad (1)$$

where, p_i (i=1, 2, 3, 4) are four pixels adjacent to (i, j), and w_i (i=1, 2, 3, 4) are global weight of the corresponding pixel. For clarity, we introduce the SOIL-GPR geometric model. The objects in soil were used as the test subject. It is assumed that the centroid radius of the object is R (m) and the diameter is recorded as D (m).

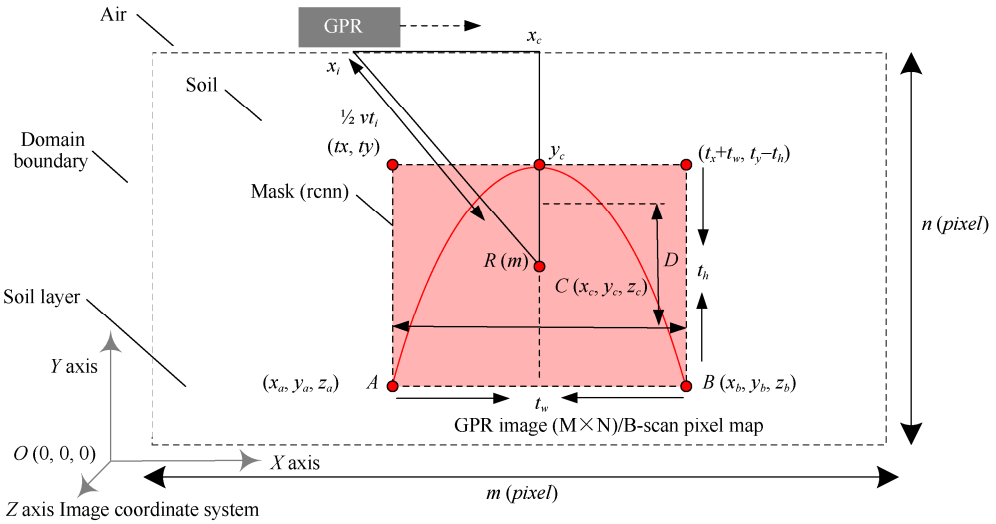
Here C (x_c, y_c, z_c) is the three-dimensional coordinates of hyperbola center in discrete space. In Figure 3, the (t_x, t_y, t_w, t_h) are pixel coordinate values of the convolution layer image. In B-scan, the z_c value can be taken as the unit value of GPR measurement step size or 1. The propagation speed of radar in soil is recorded as v, and radar wave travel time is recorded as t; the soil dielectric constant can be marked as ε, and the conductivity can be marked as σ[9,17]:

$$(x_i - x_c)^2 + y_c^2 = \left(\frac{1}{2}vt_i + R\right)^2 \quad (2)$$



a. Definite depth of soil object b. Definition of GPR image depth

Figure 2 Measurement definition of the object position



Note: R is the centroid radius of the object, m; D is the diameter, m; the (t_x, t_y, t_w, t_h) are pixel coordinate values; C (x_c, y_c, z_c) is the three-dimensional coordinates of the hyperbola center in discrete space and its unit is pixel; v is the propagation speed of radar in soil, m/s; t is the radar wave travel time, ns. GPR: Gounrd-penetrting Radar.

Figure 3 Geometric model of SOIL-GPR

From the previous hyperbolic geometry SOIL-GPR model in Figure 3, it is concluded that the value S_Δ can represent the level of hyperbola opening. When premise intercept H stays constant; the larger S_Δ, the more asymptote inclined and larger curve opening; on the contrary, the smaller S_Δ, the smaller its curve opening. In the case where C coordinates (object depth) are unchanged, Figure 4a, the opening of L_1 is significantly smaller than that L_2. When there are multiple samples, the asymptotes are represented by L_1, L_2, …, L_i. The mathematical relationship that can be deduced is $S_{1\Delta}$=CPB$_1$, $S_{2\Delta}$=CPB$_2$, …, $S_{i\Delta}$=CPB_i. With different values of C and $S_{i\Delta}$, it can extract the GPR signals from different objects in soils.

The single curve in GPR image can be mirrored into a hyperbola and then we can draw out two asymptotic for each one[18]; firstly, vertex coordinate C was calculated by image processing, and then converted the origin coordinate of the axis to the world coordinate system as showed in Figure 3 were marked XOY. Point C is the cylinder center. The 2D space in B-Scan is only discussed of hyperbolic equations $f_c(x)$.

$$\frac{\left(t_i - \left(-\frac{2R}{v}\right)\right)^2}{\left(\frac{2y_c}{v}\right)^2} - \frac{(x_i - x_c)^2}{y_c^2} = 1 \quad (3)$$

From Equation (3), it can be known that the center point coordinates of the hyperbola is $p\left(x_c, -\frac{(2R)}{v}\right)$; its focuses are F_1 and F_2. Figure 4b shows the geometric model of Equation (3),

suppose the real axis of the hyperbola is marked 2a and the imaginary axis is marked 2b[19,20]. Obviously, there exists $||AF_1|-|AF_2||=2a$ and $a=\frac{2y_c}{v}$, $b=y_c$. Distance from the focal point to the hyperbola center is $c=PF_1=PF_2$. Obviously, there is a mathematical relationship here $c^2=a^2+b^2$. Figure 4c is a mathematical asymptote model in the coordinate system. Two asymptotic lines L_{a1} and L_{a2} are symmetrical. Point B is the intersection point that passes through hyperbola vertex C which parallels to *X*-axis and intersects L_{a1}. BC is perpendicular to the longitudinal axis; PC is a straight line passing through the center of the hyperbola. Obviously, there are mathematical relations $S_{ABC}=S_\Delta=\frac{1}{2}ab=\frac{y_c^2}{v}$. α is the angle between BP and PC, (°); there is another mathematical relation $\tan\alpha=\frac{BC}{PC}$. β represents the angle between BC and BP. Q point represents any point on the curve; QN is a straight line perpendicular to L_1; QM is a straight line perpendicular to the *X*-axis and M point is on straight line L_1. Suppose H is the intercept of asymptote on *t*-axis and $H=\mp\frac{2}{v}(R+x_c)$, then the L_1 function $f_{L1}(x)$ is

$$f_{L_1}(x)=\frac{2}{v}x-\frac{2}{v}(R+x_c) \tag{4}$$

As the $|QN|$ value is approaching 0, the hyperbola will be closer to the asymptote. If $|QN|=|QM|\cos\beta$, then there is the following formula:

$$|QN|=\lim_{x\to\infty}\frac{|f_c(x)-f_{L1}(x)|}{\sqrt{1+\frac{2}{v}^2}}=\frac{\left|f_c(x)-\frac{2}{v}x+\frac{2}{v}(R+x_c)\right|}{\sqrt{1+\frac{2}{v}^2}}=0 \tag{5}$$

Combine Equations (4) and (5) to get Equations (6).

$$H=\frac{2}{v}(R+x_c)=\lim_{x\to\infty}\left[f_c(x)-\frac{2}{v}x\right],\ \text{and}\ \frac{2}{v}=\lim_{x\to\infty}\frac{f_c(x)}{x} \tag{6}$$

Equation (6) can be used to calculate the spread velocity v of electromagnetic waves among soils. The value H can be calculated by v. Now this problem can be simplified to use bilinear interpolation to find the world coordinates of point $C(x_c, y_c)$ in the coordinate system of the mask pixel. It is supposed that the pixel coordinates of a GPR image are x_{ij}. The image coordinates of any point after bilinear interpolation mapping represents as $Y\left(\frac{im}{a},\frac{jn}{b}\right)$, then there is Equation (7).

$$X_{ij}=\sum_{i=1,2,3,4}w_i(t_x,t_y) \tag{7}$$

$$X_{i,j+1}=\sum_{i=1,2,3,4}w_i(t_x+t_w,t_y-t_h) \tag{8}$$

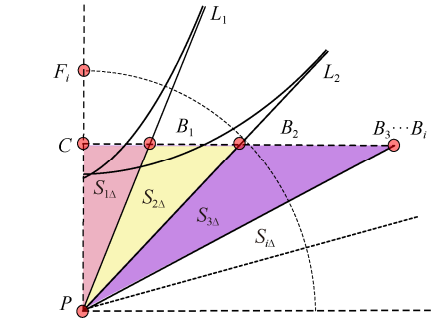
$$X_{i+1,j}=\sum_{i=1,2,3,4}w_i(t_x,t_y-t_h) \tag{9}$$

$$X_{i+1,j+1}=\sum_{i=1,2,3,4}w_i(t_x+t_w,t_y) \tag{10}$$

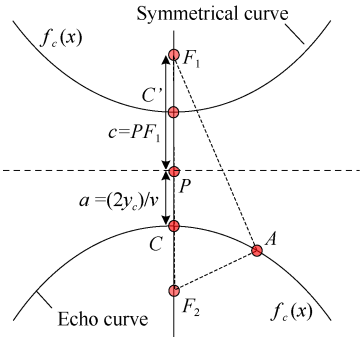
where, w_i (i=1, 2, 3, 4) is the interpolation weight that is nearest its 4 pixels of any point $Y\left(\frac{im}{a},\frac{jn}{b}\right)$, the specific definitions are as the following formula:

$$Y\left(\frac{im}{a},\frac{jn}{b}\right)=(1-a)(1-b)X_{ij}+a(1-b)X_{i+1,j}+(1-a)bX_{i,j+1}+abX_{i+1,j+1} \tag{11}$$

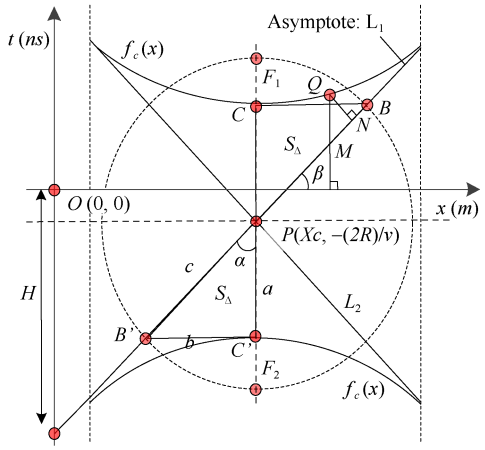
The magnitude of value α obviously represents the level of hyperbolic tilt formed by the object in the GPR echo signal. Due to the geometric symmetry of the hyperbola and its asymptote, only one of S_Δ area characteristics was focused on.



a. Asymptote enclosing area calculation



b. Geometric model



c. Geometric model in coordinate system

Figure 4 Asymptote and area feature S_Δ of SOIL-GPR model

2.4 Mask RCNN

Mask RCNN has an additional branch for predicting segmentation masks on each Region of Interest (RoI) in a pixel-to-pixel manner. We combined this with our previous geometric model, MacK RCNN can use anchor boxes to detect multiple objects in the soil with different scales or overlapping targets in GPR images. This improves the speed and efficiency of object detection. Firstly, Mask-RCNN makes a residual network convolutional search through Region Proposal Network (RPN) (Figure 5). RPN relies on windows sliding on all shared features to generate 9 anchors with fixed aspect ratios for each pixel position[21]. Among them, the size of three main anchors are

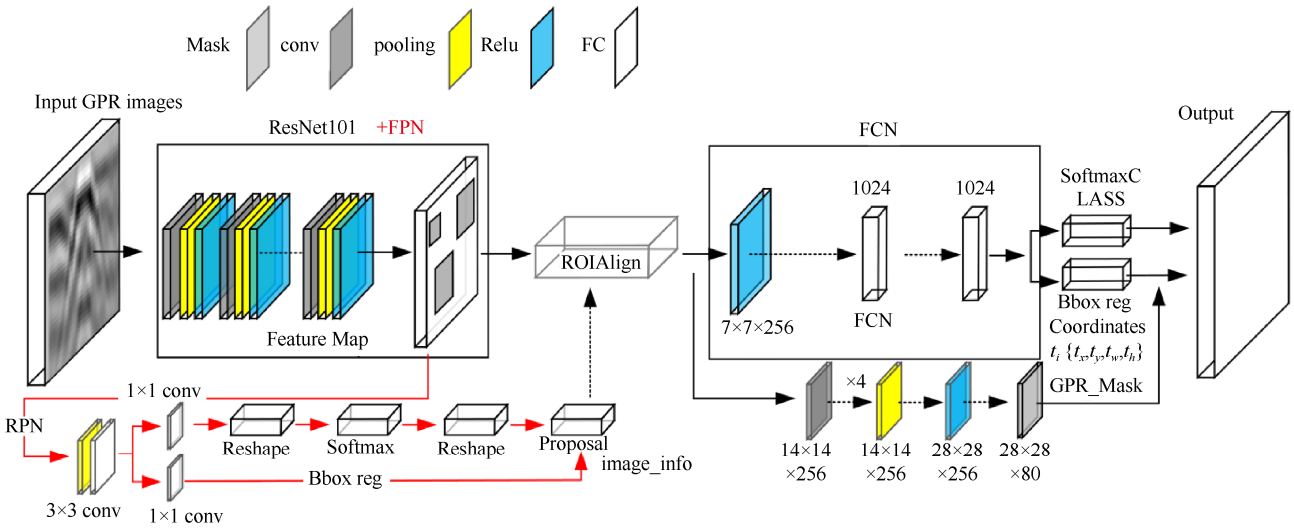
128×128 pixels, 256×256 pixels, and 512×512 pixels, respectively, then multiply by three aspect ratios of 1:1, 1:2, and 2:1 and classify each ROI found by RPN. RPN uses its sliding window to predict K anchors at the same time. Each anchor has four coordinates, so there will be 4K outputs[22]. In ROI pooling of Faster RCNN, RPN selects corresponding features for each ROI to meet the input requirements of the Fully Convolutional Networks (FCN) layer[23]. Obviously, ROI Pooling's quantization operation would not affect ROI classification, but it will produce huge errors for Mask based on pixel prediction[24]. ROI Align is not to complement coordinate points on the candidate region boundary for pooling, but to eliminate the rounding operation of ROI pooling for bilinear interpolation to accurately match up its original pixels[25,26]. ROI Align in Figure 5, followed by three branches, which included output bounding box coordinates, target's category, and mask (t_x, t_y, t_w, t_h), respectively. The "Head" behind ROI Align is to expand the output dimension. During the training procedure of the Mask Branch, N prediction maps will replace the softmax in FCN which outputs Masks for each category. This is different from semantic segmentation with FCN, which usually adopts a per-pixel sigmoid and multinomial cross-entropy loss[27,28]. For GPR images, obviously, a GPR Mask is needed to output each hyperbola $f_c^i(x)$ of electromagnetic waves. Faster RCNN's classes and bounding box regression loss function can be expressed as[29],

$$L(\{y_i\},\{t_i\})=\frac{1}{N_{class}}\sum_i L_{class}(y_i,\overline{y}_i)+\gamma\frac{1}{N_{boxes}}\sum_i \overline{y}_i L_{boxes}(t_i,\overline{t}_i) \quad (12)$$

where, y_i is the probability of the anchor which are predicted as targets, $\overline{y}_i \in [0,1]$ represents the labels of ground truth; $t_i\{t_x, t_y, t_w, t_h\}$ represents the offset vector predicted by the anchor in RPN stage; $\overline{t}_i$ and t_i have the same dimension and it is an actual offset of the bounding box relative to ground truth. It added the loss function L_{GPR_mask} in Mask RCNN and adopted average binary cross-entropy loss for training[30]. The multi-task loss function in Mask-RCNN is,

$$L=L_{class}+L_{boxes}+L_{GPR\_mask} \quad (13)$$

where, L_{class} is classification loss and L_{boxes} is bounding regression loss. This multi-task loss function can select those anchors that contain the target to maximum and adjust the bounding box position and size[31]. Finally, bounding boxes and masks are generated. We will use ROI Align to get a 7×7×1024 feature map, and then use ResNets to obtain a 7×7×2048 feature map. There have 2048 channels divided into two branches which belong to classification and regression. Another branch is responsible for generating a 14×14×80 GPR mask (Figure 5). Due to multiple convolutions and pooling, the corresponding resolution is reduced. Mask RCNN in this study used deconvolution to improve resolution while reducing the number of channels. If FPN network architecture was selected, after inputting a single-scale GPR picture, a corresponding feature pyramid will be generated[32]. In order to adopt ROI Align to connect two branches, one branch output a 7×7×256 feature map and is then vectored into 1024 channels. Finally, it output the results of classification and bounding box regression; Others output a feature map with size 14×14×256 and a 28×28×80 detected mask at last. Obviously, FPN can let fewer filters make efficient use of every convolutional layer feature (Section 2.4). FPN networks can compile GPR image information in different dimensions on features of different scales.

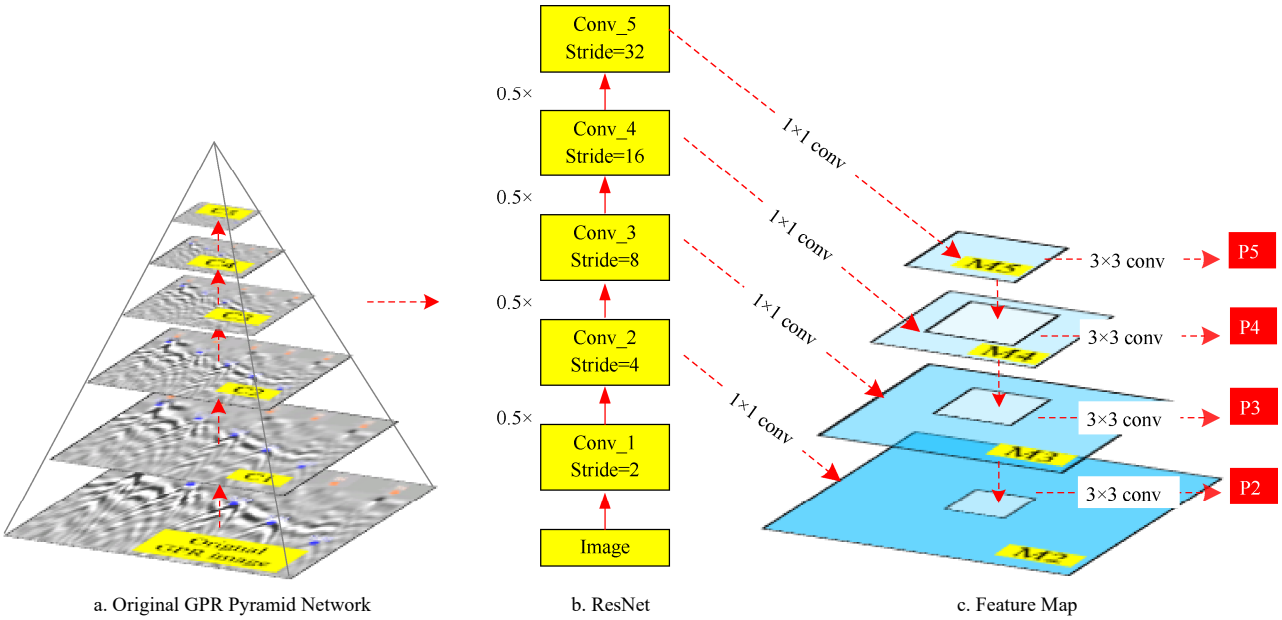


Note: Conv: Convolution; ReLU: Rectified Linear Unit; FC: Fully connected layer; RPN: Region Proposal Network; FPN: Feature Pyramid Network; ROI: Region of Interest; FCN: Fully Convolutional Network; Bbox reg: Bounding box regression; $t_i\{t_x, t_y, t_w, t_h\}$ represents the offset vector predicted by anchor in RPN stage.

Figure 5 Mask RCNN framework for GPR images

There are many mainstream frameworks for deep learning, such as GoogleNet, AlexNet, Visual Geometry Group (VGG), and ResNets[33,34]. Obviously, updating the lower-level network's parameter during the training process of the high-level network easily leads to the disappearance of the gradient. This is not caused by overfitting but by the model degradation caused by redundant network layer learning parameters that are not identity mapping[35-37]. In the actual situation, the different sizes of objects mirror GPR images in form of electromagnetic wave signals are different. It is necessary to quickly classify objects in GPR images and then make an instance segmentation[38]. SPP net, Fast RCNN, and Faster RCNN all adopt their last layer of network features to make predictions. The lowest layer features of Single Shot Detector (SSD) only use the conv4_3 layer's characteristics of VGG network[39]. The core idea of FPN is to construct a multi-scale pyramid to extract multi-layer features from the original image, then make a bounding box prediction according to different features proposed by each layer[40,41]. This is an improved CNN

idea. Because of the process of convolving and pooling from the original image, multi-scale feature maps can be generated equal to constructing a pyramid in feature space of image. Extracting pixel details from low-level networks, high-level networks can perform semantic segmentation more quickly and make classification[42]. The framework of FPN is shown in Figure 6. Firstly, it performs a deep convolution on the original GPR image, then reduces the conv_1 dimension feature and performs a down-sample operation on C_1. Secondly, FPN adds the corresponding elements to processed conv_2 and conv_3 and inputs them into M2, M3. Composed with the horizontal connection and up-sample and down-sample of feature map, it will generate the same size as feature map[43]. This study used 3×3 convolutional kernel to eliminate the aliasing effect. Its stride is 2, 4, 8, 16, and 32 in ResNets framework, respectively. The conv_2, conv_3, conv_4, conv_5 and M2, M3, M4, and M5 have one-to-one correspondence relationship, and then through 3×3 convolutional kernel to obtain P2, P3, P4, and P5. In this way, corresponding feature maps and strong semantic segmentation performance can be output from each layer. It is obvious that semantic feature information in low-level layer network is less, but the detections are accurate. Feature information of upper-level layer network is richer, but the accuracy of detected pixel unit is not high[44].



a. Original GPR Pyramid Network b. ResNet c. Feature Map

Figure 6 FPN framework with GPR image

3 Results and discussion

3.1 Soil layer detection

The object detection experiments in this study were designed into two working conditions, which included outdoor and indoor soils. The soil cube values of length, width, and depth of the indoor experiment were 6.0 m, 2.0 m, and 1.5 m, respectively. The soil dielectric constant is single and its default value was that in addition to the artificially buried objects before the experiment, there is no other unknown hard foreign objects in the soil; when indoor soil layer experiments were studied, the location where this done was at the key laboratory of southern agricultural machinery to operation. In this experiment, the indoor soil sand pond was used as the location, and an upper corner of the sand pond boundary is used as the global coordinate system {0}'s origin point. The starting point of the transmitting antenna of the system under test is the starting point of the actual coordinate system {1} of the measurement picture. After collecting the GPR data, we use ReflexW software for preprocessing. The radar was detected according to the scanning grid; the experiment was followed by dragging the radar mileage wheel on the surface of the soil layer along the *X* and *Y* directions. ReflexW's tracking function was used for tracking soil layers. The function compares the energy values of the collected electromagnetic wave grayscale images. Because the soil layers will cause a sudden change in the energy amplitude of an electromagnetic wave during the propagation process, for example, in the same plant soils, even if the moisture content was different, there was a big difference in its value. Due to the different soil parameters, we can use the software to easily distinguish the soil stratification, as shown in Figure 7, four colors represent four different physical parameters of layered soils. Obviously, the Reflexe track function can not only effectively detect and distinguish but also divide the medium into a layer with different dielectric constants.

With the relationship between moisture content and dielectric constant, this study introduced ridge regression analysis on the magnetic loss according to the attenuation coefficient of the soil layer. It is an improved least squares estimation method. Abandoning the unbiasedness of the least squares method, the cost of losing some information and reducing accuracy to obtaining the regression coefficient. This method is more in line with the actual and more reliable regression. The amplitude of the refracted wave attenuates along the depth of soil, and the phase change was related to the detection direction of GPR. When the constant amplitude surface of the electromagnetic wave was inconsistent, the angle between the propagation direction of the refracted wave and its *Z*-axis meets the corrected relationship of the law of refraction. By fitting the correlation between the magnetic loss attenuation function, it was found that when applying partial least squares regression (PLS) to predict the correlation between the physical and magnetic loss factor rate, the effective modeling confidence interval for soil moisture is wider than its originals. The probability of accepting the hypothesis becoming greater, although multiple correlations do not affect the

unbiasedness and minimum variance of the PLS estimator, the least squares estimator in all linear unbiased estimators middle was the smallest variance, this variance was not small necessarily; Although this estimator has a slight deviation, its accuracy can be much higher than the unbiased estimator. As shown in Figures 8 and 9, the reflection coefficient K and refraction coefficient T were easily interfered with by the magnetic loss of electromagnetic wave direct wave under the influence of soil moisture content; for example, the generalized reflection coefficient in Figure 10 peaked when the soil moisture was 15.4%, the corresponding refraction coefficient is 3.945, and then a trough appeared when the soil moisture percentage was 17%. To obtain logistic regression equations for the reflection coefficient and refraction coefficient, finally, it analyzes the relationship between magnetic loss and soil layering. The analysis results are shown in Figure 10.

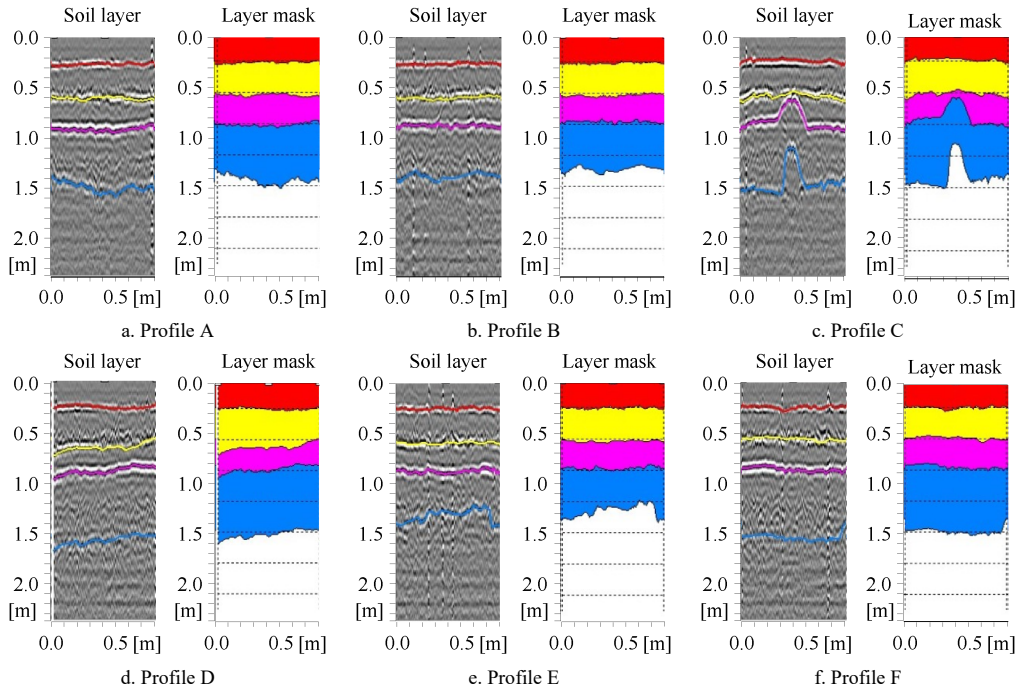


Figure 7 Stratification analysis of layered soil

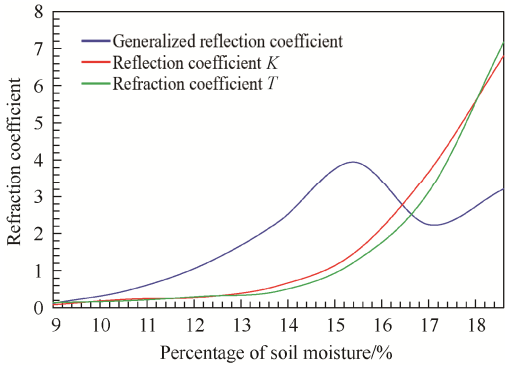


Figure 8 Refraction coefficient (T) and reflection coefficient (K) before ridge regression

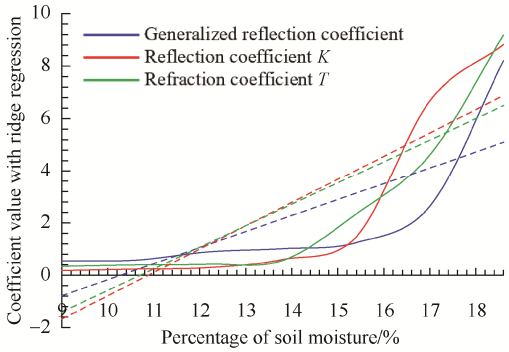


Figure 9 Refraction coefficient (T) and reflection coefficient (K) with ridge regression

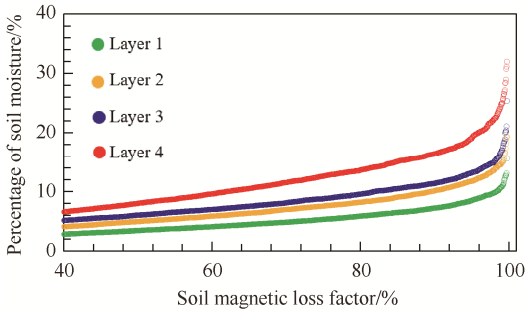
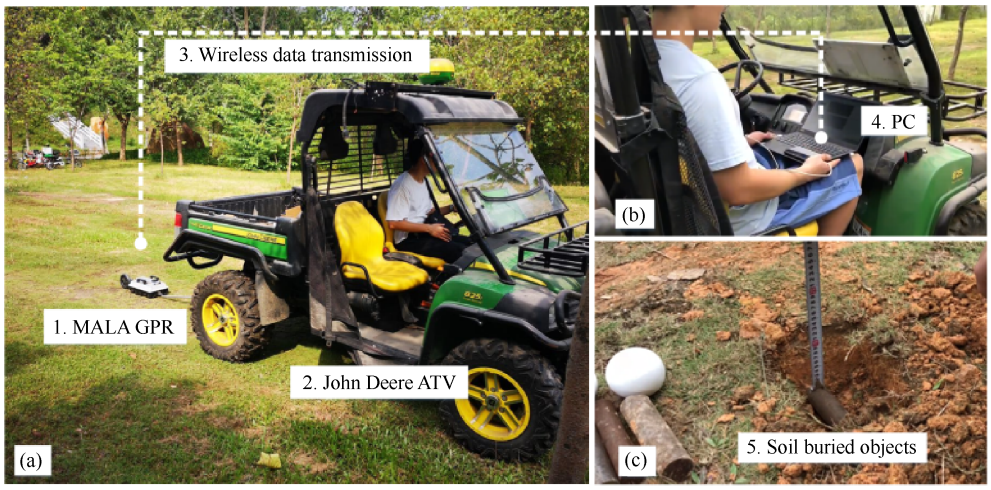


Figure 10 Soil magnetic loss versus soil moisture (layer)

3.2 Object position results

The field experiment in this section was done at South China Agricultural University. Our test route A-B was marked by the operator according to the flatness of the soil. The test length of the A-B line should be greater than or equal to 5 m, and the width should be greater than or equal to 2 m. In addition, the measuring instrument used in this study was the Swedish MALA 750HD ground penetrating radar, and the measuring indicators are the position of objects in soil and the false alarm rate of the Mask-RCNN vision system. The operating performance of the machine vision inspection system can be tested for accuracy by using the coordinate system of the measured distance as a reference value; first, a global coordinate system XOY was established, and then a local coordinate system $X_1O_1Y_1$ should be set; the measurement line AB can be multiple parallel straight lines, The detection route of the ground penetrating radar carries out multiple

inspections according to the route in the figure. GPR collects radar data through the radar trigger wheel, and the collected data is used for instance detection through Mask-RCNN and U-Net with a PC. The instance segmentation effect was tested in 3 kinds of actual situations: the single parabola, the intersecting parabola, and overlapping waveforms in the vertical direction (Figure 11).



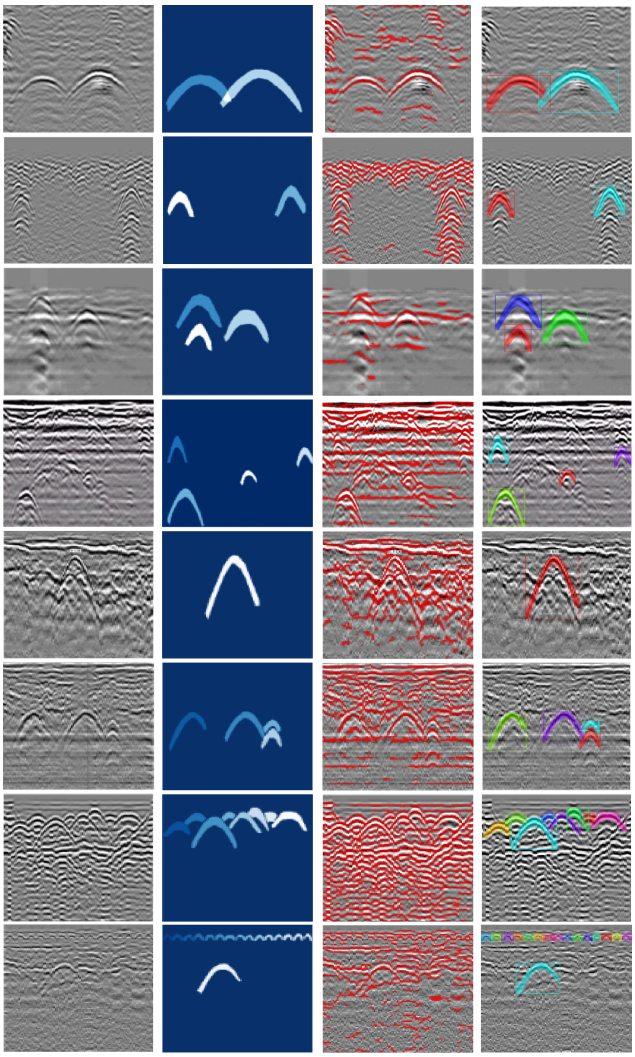
a. GPR and vehicle b. PC communication c. Soil object measurement

Figure 11 Field experiment of all-terrain vehicle equipped with radar

It can be seen from Figure 12 that the Mask RCNN with FCN can segment most objects from the GPR image, whether it was on the overlap or non-overlap situation. However, the U-Net segmentation result was not ideal, most of the single waveforms can be separated, but at the same time, with the overlapped curve, it was also easy to segment from layered soil which refers to the same detection location, with different soil physical parameters and different depths. Because U-Net can perform pixel-level texture segmentation and the network structure was simple, it is easily interfered with by the layered soil, thereby reducing the segmentation effect. It is obvious that Mask RCNN can combine the similarity merge algorithm to filter lots of unnecessary interfered boxes and process many strong oscillations of electromagnetic signal into an integrated bounding box prediction result. For small targets, Mask RCNN uses deep networks for feature mapping. In addition, the mAPs of two network frameworks were compared. The Mask RCNN with the merged similarity principle proposed in this paper has achieved the ideal mAP of 97.48% for small objects and 96.84% for big objects. In order to judge the pixel segmentation performance of Mask RCNN and U-Net model more intuitively, this study compared their output mask generated by Mask RCNN and U-Net. Figure 12 shows the Mask RCNN recognition and segmentation effect of the hyperbolic echo signal in the GPR image, where the first column in the figure represents the original GPR image; the second column represents the ground truth mask; the third column shows the U-Net segmentation effect and the fourth column showing the segmentation performance of Mask RCNN. In contrast, U-Net can accurately display the parabolic waveform of the target with a mask, but it is more sensitive to the signal of soil layering or oscillation. This is because it is known that in the backbone framework of Mask-RCNN, FPN can perform corresponding bounding box prediction based on different features proposed by each layer. Furthermore, the process of FPN convolving and pooling the original image is similar to constructing a pyramid in image feature space, and extracting detailed information on the low-level networks[45-47]. The conclusion was that the most of soil object hyperbola can be selected and segmented by Mask-RCNN. The algorithm can obtain nice detection results in the following scenarios, such as some samples which overlap each other, the distance between the waveforms being very close, or its transverse span being relatively large in the layered soil. Mask RCNN was used to identify multiple targets of smaller size with a volume was less than 125 cm^2; in small target detection, the mAP of Mask RCNN can reach 96.30%. In summary, Mask RCNN can segment multiple objects from GPR images accurately.

As mentioned earlier, we studied the pixel discrete correspondence of bilinear interpolation. Mask RCNN had achieved good mAP value in instance segmentation. Obviously, after Mask RCNN joins the ROI Align, it can obtain accurate discrete correspondence on each pixel with the original GPR image. The Mask RCNN can segment each parabolic target body from the original image, whether it was a small or unobvious object. Obviously, the parabola selected by the frame can be fitted to the hyperbola through the geometric model. As in Figure 13, through the t_h, t_w, and pixel coordinates of point A or B calculation, it can be obtained the model parameters of the asymptote $Y=kx+b$. The y_c in the figure corresponds to the vertex of the parabola, and the area feature S_Δ of the SOIL-GPR model in Figure 4 can be obtained through this asymptote.

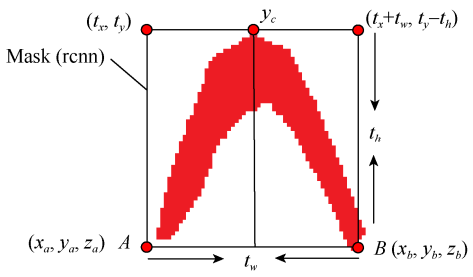
As mentioned earlier, we studied the pixel discrete correspondence of SOIL-GPR model. After Mask RCNN joins the ROI Align, it can obtain accurate discrete correspondence on each pixel. Mask RCNN can segment each parabolic object in the soil. Obviously, the parabola selected by the frame can be fitted to the hyperbola through the geometric mode. With the t_h, t_w, and pixel coordinates of point A or B calculation, it can be obtained the model parameters of the asymptote $Y=kx+b$. The parabola in Figure 14 shows the image recognition result located or predicted by Mask RCNN and the SOIL-GPR model proposed in this study. First, ridge regression was applied to the reflection coefficient R and refraction coefficient T, then the GPR image samples with the known parabolic vertex Y_c were compared. Evaluated the Mask RCNN positioning and recognition results in different layered soil, it was found that before the regression analysis of the electromagnetic wave attenuation parameters i, the value of Y_c in the first layer of the measured value and Mask RCNN output values was small in overall error (Figure 14a). Obviously, this kind of result in the second, third, or fourth layer in Figure 14 was not ideal; for example, the maximum prediction error in Figure 14b was 0.29 m.



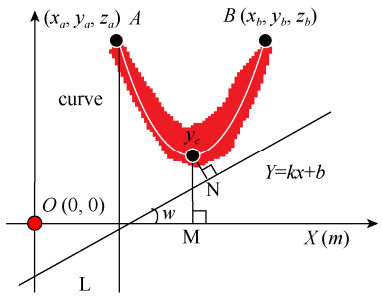
a. $H\times W$=512×512 b. Soil target c. U-Net d. Mask-RCNN

Figure 12 Detection effect with different instance segmentation algorithms

Combining the soil layer with the ridge regression analysis of Mask RCNN, it can be seen from Figure 15 that when in the first soil layer, the maximum difference between the Y_c multi-group prediction and measured values of Mask RCNN was only 0.01 m. In the second layered soil, the maximum difference between the measured value of Y_c and the predicted value was 0.06 m. As shown in Figure 15b, the accuracy had been improved by Mask-RCNN. Similarly, in the third soil layer, the maximum error of the Y_c's image coordinates identification and positioning was below 0.06 m; while in the fourth soil layer, the maximum error of the Y_c coordinates was 0.24 m, but the minimum error can reach 0.03 m. Those results show that the Mask RCNN detection model based on the analysis of the electromagnetic wave attenuation coefficient can apply to the detection of layered soil images.

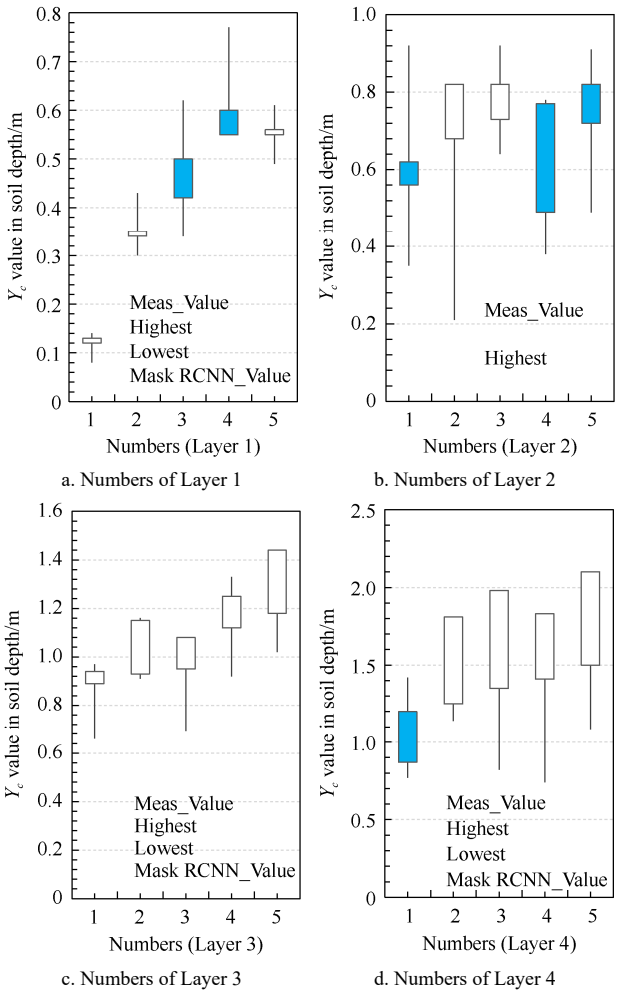


a. Segmentation detection frame positioning



b. Asymptote calculation

Figure 13 Calculation of parabolic geometric model



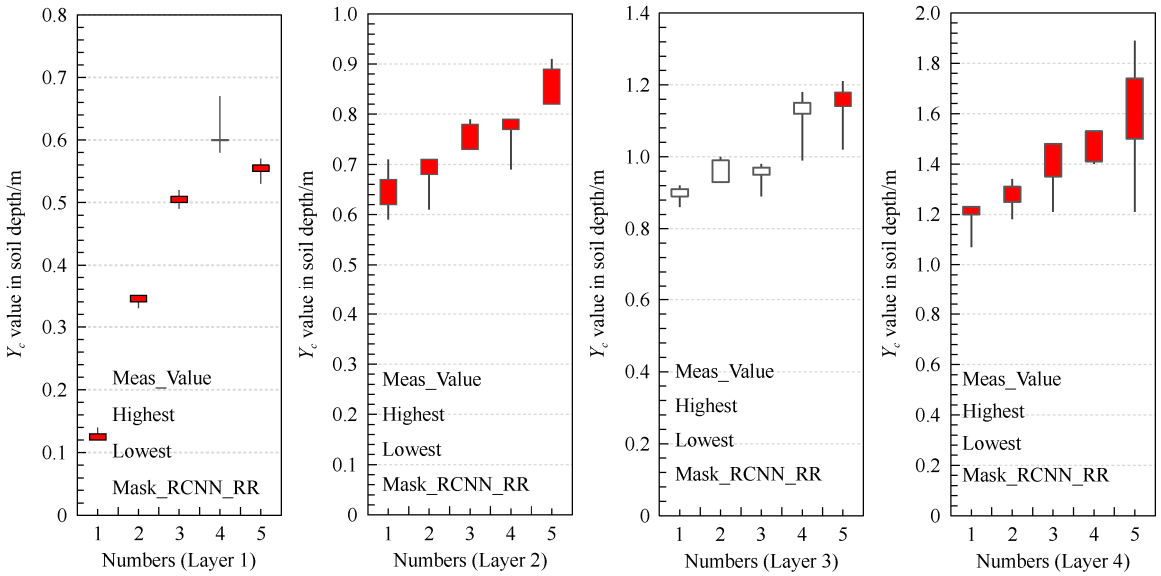
a. Numbers of Layer 1 b. Numbers of Layer 2

c. Numbers of Layer 3 d. Numbers of Layer 4

Figure 14 Y_c value analysis in different soil layers

3.3 Evaluation of location detection

This article conducts experimental statistics based on the evaluation indicators in Section 2.2. Table 3 is the experimental results on the length of the measuring line. The distance was marked and recorded between the actual buried position of objects in soil and the initial measurement position. With this distance, this value was compared with the mask output by the algorithm and the prediction results of the asymptotic features. The average error of 20 times detection was 0.04365 m, and the mean-variance was 0.000931 m. Table 4 is a comparison of the results of Mask-RCNN detection in the depth direction of soil detection. The average error of 20 times is 2.8700 cm, and the mean-variance was 2.31105 cm.



a. Numbers of Layer 1 b. Numbers of Layer 2 c. Numbers of Layer 3 d. Numbers of Layer 4

Figure 15 Ridge regression (RR) prediction analysis

Table 3 Test results in the measure line direction (20 times)

Data	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Actual distance/m	1.420	1.420	1.420	1.420	1.420	3.500	3.500	3.500	3.500	3.500	4.500	4.500	4.500	4.500	4.500	6.000	6.000	6.000	6.000	6.000
Detected distance/m	1.502	1.488	1.511	1.457	1.523	3.514	3.567	3.466	3.512	3.519	4.509	4.517	4.556	4.490	4.562	6.087	6.033	6.015	6.019	5.982
Error/m	0.082	0.068	0.091	0.037	0.103	0.014	0.067	0.054	0.012	0.019	0.009	0.017	0.056	0.010	0.062	0.087	0.033	0.015	0.019	0.018

Note: Mean error is 0.043650; Mean variance is 0.000931.

Table 4 Test results of soil depth direction (20 times)

Data	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Actual distance /cm	30.00	30.00	30.00	30.00	30.00	30.00	30.00	30.00	30.00	30.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00	40.00
Detected distance /cm	28.19	27.11	28.45	35.03	26.61	27.20	31.04	28.88	27.09	28.87	34.44	38.90	37.09	44.08	42.69	38.26	36.09	37.48	37.66	33.32
Error/cm	1.81	2.89	1.55	5.03	3.39	2.80	1.04	1.12	2.91	1.13	5.56	1.10	2.91	4.08	2.69	1.74	3.91	2.52	2.54	6.68

Note: Mean error is 2.870000; Mean variance is 2.31105.

The missed detection rate refers to the percentage of the number of soil objects that are not detected by the detection system in the total number of the inspection batch. The false alarm rate refers to the percentage of the total number of inspection batches that are not soil objects but were detected by the machine vision inspection system. The false alarm rate is based on recording the number of objects buried in the soil on the AB route, and then using the machine vision system to detect the object in the soil in the radar image scanned by the radar, and to count the number of soil objects detected by the machine vision system. It calculates the number of misrecognition due to the detection of non-target objects based on the number of objects detected by the machine vision system. Finally, we recorded the percentage of non-target object recognition by the algorithm in the total number of detections. The underreporting rate is based on recording the number of objects buried in soil on the AB route, and then the machine detects the object in the radar image scanned by the ground penetrating radar through the vision system and counts the number of objects in soil detected by the machine vision system, According to the number of buried objects minus the number of the detected soil objects by the system, the number of correct detection is obtained. Finally, the percentage of the total number of targets detected by the algorithm for the number of missed targets was recorded. The average variance of the number of objects s detected by Mask-RCNN is 67.466667; the false positive rate is 0.055%, and the false negative rate is 0.075% (Table 5).

Table 5 Statistics of false alarm rate and false alarm rate

Test times	false alarm rate	missed detection rate
10	8%	5%
15	5%	8%
20	5%	5%
30	4%	12%
Sample mean	0.0550	0.075

4 Conclusions

In this study, a method was proposed for detecting the position of objects in soil using Mask Region-based Convolutional Neural Network (Mask-RCNN). This method combines electromagnetic wave detection technology, geometric model theory, and deep learning algorithms. By integrating that method, a smart solution, and a non-destructive method were proposed for object detection in agricultural soil. The conclusions in this study mainly include the following four points:

1) The layering of agricultural soil and the law of electromagnetic wave signals can be mapped through the generalized reflection coefficient; when deep learning algorithms were used for instance segmentation of Gounrd-penetrting Radar (GPR) images, the analysis of layered soil can be performed in advance; it will clarify the relationship between magnetic loss and layered soil to improve the object information implied by the parabola.

2) This study proposed a mathematical model of objects in the soil based on parabolic detection. This model can be combined with Mask-RCNN's mask geometric size to predict the size of cultivated objects, and it can also accurately calculate the size of the detected target through mathematical methods and the position of the object. The result shows that this model can improve the prediction accuracy of Mask-RCNN instance segmentation.

3) This study compared the multiple instance segmentation algorithms in GPR images. When the objects in layered soil are relatively single, the U-Net, Fully Convolutional Networks (FCN), and Mask-RCNN can achieve ideal segmentation results; but when the conditions are complicated, the segmentation accuracy of Mask-RCNN is more ideal.

4) This study simulated the operated speed of agricultural machinery to design an A-B survey line experiment, using the survey line direction, soil depth, false alarm rate, false alarm rate, Mean Average Precision (mAP), and Intersection over Union (IoU) to evaluate the detection accuracy with Mask-RCNN.

This experiment provides a high-precision technical test and solution for the GPR non-destructive testing technology of cultivated soil.

Acknowledgments

This work was financially supported by the Laboratory of Lingnan Modern Agriculture Project (Grant No. NT2021009); Guangdong University Key Field (Artificial Intelligence) Special Project (No. 2019KZDZX1012) and the 111 Project (D18019); Guangdong Basic and Applied Basic Research Foundation (Grant No. 2021A1515110554); China Postdoctoral Science Foundation (Grant No. 2022M721201); the National Natural Science Foundation of China (Grant No. 31901411); The Open Competition Program of the Top Ten Critical Priorities of Agricultural Science and Technology Innovation for the 14th Five-Year Plan of Guangdong Province (No. 2022SDZG03).

[References]

[1] Li Y H, Zhao X X, Xu W C, Liu Z, Wang X. An effective FDTD model for GPR to detect the material of hard objects buried in tillage soil layer. Soil and Tillage Research, 2019; 195: 104353. doi: 10.1016/j.still. 2019.104353.

[2] Azizi A, Abbaspour-Gilandeh Y, Vannier E, Dusseaux R, Mseri-Gundoshmian T, Moghaddam H A. Semantic segmentation: A modern approach for identifying soil clods in precision farming. Biosystems Engineering, 2020; 196: 172–182.

[3] Xu J J, Zhang H, Tang C S, Cheng Q, Liu B, Shi B. Automatic soil desiccation crack recognition using deep learning. Geotechnique, 2021; pp.1–13.

[4] Ma W P, Wang X D, Liu D. Research on YOLO target detection algorithm in GPR image anomaly recognition. Bulletin of Surveying and Mapping, 2019; S1: 72–76. (in Chinese)

[5] Dou Q, Wei L, Magee D R, Cohn A G. Real-time hyperbola recognition and fitting in GPR data. IEEE Transactions on Geoscience and Remote Sensing, 2017; 55(1): 51–62.

[6] Pham M T, Lefèvre S. Buried object detection from B-scan ground penetrating radar data using Faster-RCNN. In: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia: IEEE, 2018; pp.6804–6807. doi: 10.1109/IGARSS.2018.8517683.

[7] Gong Z M, Zhang H Q. Research on GPR image recognition based on deep learning. In: 2019 International Conference on Computer Science Communication and Network Security (CSCN2019), 2019; 309: 03027. doi: 10.1051/matecconf/202030903027.

[8] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014; pp.580–587.

[9] Yang F, Qiao X, Zhang Y Y, Xu X L. Prediction method of underground pipeline based on hyperbolic asymptote of GPR image. In: Proceedings of the 15th International Conference on Ground Penetrating Radar, IEEE, 2014; pp.674–678. doi: 10.1109/ICGPR.2014.6970511

[10] Gkioxari G, Girshick R, Malik J. Contextual action recognition with R* CNN. In 2015 IEEE International Conference on Computer Vision (ICCV), Santiago: IEEE, 2015; pp.1080–1088. doi: 0.1109/ICCV.2015.129.

[11] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: Proceedings of the IEEE international conference on computer vision 2017; pp.2961–2969.

[12] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017; 39(6): 1137–1149.

[13] Rice W, Omwenga M, Wu D L, Liang Y. Enhanced underground object detection with conditional adversarial networks. In: ISSAT International Conference on Data Science & Intelligent Systems, 2019.

[14] Gribbon K T, Bailey D G. A novel approach to real-time bilinear interpolation. In: Proceedings. DELTA 2004. Second IEEE International Workshop on Electronic Design, Test and Applications, Perth: IEEE, 2004; pp.126–131. doi: 10.1109/DELTA.2004.10055.

[15] Janning R, Busche A, Horváth T, Schmidt-Thieme L. Buried pipe localization using an iterative geometric clustering on GPR data. Artificial Intelligence Review, 2014; 42(3): 403–425.

[16] Savelyev T G, van Kempen L, Sahli H, Sachs J, Sato M. Investigation of time–frequency features for GPR landmine discrimination. IEEE Transactions on Geoscience and Remote Sensing, 2006; 45(1): 118–129.

[17] Valueva M V, Nagornov N N, Lyakhov P A, Valuev G V, Chervyakov N I. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. Mathematics and Computers in Simulation, 2020; 177: 232–243.

[18] Nuzzo L. Coherent noise attenuation in GPR data by linear and parabolic Radon Transform techniques. Annals of Geophysics, 2003; 46(3): 3426. doi: 10.4401/ag–3426.

[19] Benedetto F, Tosti F. GPR spectral analysis for clay content evaluation by the frequency shift method. Journal of Applied Geophysics, 2013; 97: 89–96.

[20] Lachowicz J, Rucka M. Application of GPR method in diagnostics of reinforced concrete structures. Diagnostyka, 2015; 16(2): 31 – 36.

[21] Tang P, Wang X, Wang A, Yan Y, Liu W, Huang J, et al. Weakly supervised region proposal network and object detection. In: Proceedings of the European conference on computer vision (ECCV), 2018; pp.352–368.

[22] Li J Q, Wu Y, Zhao J Q, Guan L T, Ye C, Yang T. Pedestrian detection with dilated convolution, region proposal network and boosted decision trees. In: 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage: IEEE, 2017; pp.4052–4057.

[23] Lei T, Zhang Q, Xue D H, Chen T, Meng H Y, Nandi A K. End-to-end change detection using a symmetric fully convolutional network for landslide mapping. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton: IEEE, 2019; pp.3027–3031. doi: 10.1109/ICASSP.2019. 8682802.

[24] Yan C, Chen W W, Chen P C, Kendrick Amezquita S, Wu X M. A new two-stage object detection network without RoI-Pooling. In: Proceedings of the 30th Chinese Control And Decision Conference (2018 CCDC), 2018; pp.1680–1685.

[25] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 1997; 55(1): 119–139.

[26] Li X C, Wang L, Sung E. A study of AdaBoost with SVM based weak learners. In: 2005 IEEE International Joint Conference on Neural Networks, Montreal: IEEE, 2005; 1: 196–201. doi: 10.1109/IJCNN.2005.1555829.

[27] Lin T Y, Maire M, Belongie S, Hay J, Perona P, Ramanan D, et al. Microsoft COCO: Common objects in context. In: European Conference on Computer Vision, 2014; pp.740–755.

[28] Li Y H, Wang X, Zhao Z X, Han S, Liu Z. Lagoon water quality monitoring based on digital image analysis and machine learning estimators. Water Research, 2020; 172: 115471. doi: 10.1016/j.watres. 2020.115471.

[29] Sun L Q, Zou Y B, Li Y, Cai Z D, Li Y, Luo B, et al. Multi target pigs tracking loss correction algorithm based on Faster R-CNN. Int J Agric & Biol Eng, 2018; 11(5): 192–197.

[30] Johnson J W. Adapting Mask-RCNN for automatic nucleus segmentation. arXiv preprint, 2018; arXiv:1805.00500.

[31] Shaodan L, Chen F, Zhide C. A ship target location and mask generation algorithms base on Mask RCNN. International Journal of Computational Intelligence Systems, 2019; 12(2): 1134–1143.

[32] Huang Z D, Zhong Z Y, Sun L, Huo Q. Mask R-CNN with pyramid attention network for scene text detection. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa: IEEE, 2019; pp.764–772.

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

[34] Schmidhuber J. 2015. Deep learning in neural networks: An overview. Neural Networks, 2015; 61: 85–117.

[35] Chen Y S, Lin Z H, Zhao X, Wang G, Gu Y F. Deep learning-based classification of hyperspectral data. IEEE Journal of Selected Topics in Applied Earth Obserbvations and Remote Sensing, 2014; 7(6): 2094–2107.

[36] He K M, Zhang X Y, Ren S Q, Sun J. Identity mappings in deep residual networks. In European Conference on Computer Vision 2016; pp.630–645.

[37] Zagoruyko S, Komodakis N. Wide residual networks. arXiv, 2016; arXiv preprint arXiv:1605.07146.

[38] Li S, Yuan C X, Cai H B. Integrated processing of image and GPR data for automated pothole detection. Journal of Computing in Civil Engineering, 2016; 30(6): 04016015. doi:10.1061/(ASCE)CP.1943-5487.0000582.

[39] Fu C Y, Liu W, Ranga A, Tyagi A, Berg A C. DSSD: Deconvolutional Single Shot Detector. arXiv preprint, 2017; arXiv:1701.06659.

[40] Lempitsky V, Kohli P, Rother C, Sharp T. Image segmentation with a bounding box prior. In: 2009 IEEE 12th International Conference on Computer Vision, IEEE, 2009; pp.277–284.

[41] Wang G B, Lan Y B, Qi H X, Chen P C, Hewitt A, Han Y X. Field evaluation of an unmanned aerial vehicle (UAV) sprayer: effect of spray volume on deposition and the control of pests and disease in wheat. Pest Management Science, 2019; 75(6): 1546–1555.

[42] Mu G, Chen Y, Wu D, Zhan Y, Zhou X S, Gao Y. Relu Cascade of Feature Pyramid Networks for CT pulmonary nodule detection. In: International Workshop on Machine Learning in Medical Imaging, 2019; pp.444–452.

[43] Ghiasi G, Lin T Y, Le Q V. NAS-FPN: Learning scalable feature pyramid architecture for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2019; pp.7036–7045.

[44] Chiao J Y, Chen K Y, Liao K Y K, Hsieh P H, Zhang G, Huang T C. Detection and classification the breast tumors using mask R-CNN on sonograms. Medicine, 2019; 98(19): e15200. doi: 10.1097/MD. 0000000000015200.

[45] Chen L C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. arXiv preprint, 2017; arXiv:1706.05587.

[46] Chen P C, Douzals J P, Lan Y B, Cotteux E, Delpuech X, Pouxviel G, et al. Characteristics of unmanned aerial spraying systems and related spray drift: A review. Frontiers in Plant Science, 2022; 13: 870956. doi: 10.3389/fpls.2022.870956.

[47] Zhan Y L, Chen P C, Xu W C, Chen S D, Han Y F, Lan Y B, et al. Influence of the downwash airflow distribution characteristics of a plant protection UAV on spray deposit distribution. Biosystems Engineering, 2022; 216: 32–45.