

Article

Cotton Stubble Detection Based on Improved YOLOv3

Yukun Yang ^{1,2}, Jingbin Li ^{1,2,*}, Jing Nie ^{1,2}, Shuo Yang ^{1,2} and Jiaqiang Tang ³

¹ College of Mechanical and Electrical Engineering, Shihezi University, Shihezi 832000, China; 18819267947@163.com (Y.Y.); niejing19@163.com (J.N.); 20202309010@stu.shzu.edu.cn (S.Y.)

² Industrial Technology Research Institute, Xinjiang Production and Construction Group (XPCC), Shihezi 832000, China

³ College of Mechanical Engineering, Hebei University of Technology, Tianjin 300401, China; 820520tang@gmail.com

* Correspondence: lj8095@163.com

Abstract: The stubble after cotton harvesting was used as the detection object to achieve the visual navigation operation for residual film recovery after autumn. An improved (You Only Look Once v3) YOLOv3-based target detection algorithm was proposed to detect cotton stubble. First, field images of residual film recycling were collected. Considering the inconsistency between stubble size and shape, a segmented labeling data set of stubble is proposed. Secondly, the Darknet-53 backbone of the original YOLOv3 network is improved to accommodate tiny targets. Next, the prediction anchor box of the improved detection backbone is clustered using K-means++, and the size of the prediction anchor box suitable for improved YOLOv3 is determined. Finally, for the false detection points after detection, a mean value denoising method is used to remove the false detection points. Feature points are extracted from the denoised stubble, and the candidate points are fitted by the least square method to obtain the navigation line. The optimal model with a mean average precision (mAP) of 0.925 is selected for testing at the test stage. The test results show that the algorithm in this article can detect the stubble of residual film recovery images at different locations, different time periods, and different camera depression angles without misdetection. The detection speed of a single image is 98.6 ms. Based on an improvement over YOLOv3, the improved model has a significantly higher detection rate in different scenarios than YOLOv3. This can provide practical technical support for the visual navigation of residual film recovery.



Citation: Yang, Y.; Li, J.; Nie, J.; Yang, S.; Tang, J. Cotton Stubble Detection Based on Improved YOLOv3. *Agronomy* **2023**, *13*, 1271. <https://doi.org/10.3390/agronomy13051271>

Academic Editor: Paul Kwan

Received: 9 March 2023

Revised: 21 April 2023

Accepted: 24 April 2023

Published: 28 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: machine vision; deep learning; target detection; cotton stubble; residual film recovery; navigation line extraction

1. Introduction

China's cotton production is concentrated in Xinjiang. Xinjiang has low rainfall and a large temperature difference between day and night. Plastic mulch cultivation technology can increase temperature, preserve moisture, save water, and increase cotton production. After cotton harvesting, mulch film must be recycled immediately. Otherwise, the remaining mulch film will change the soil's physical structure and result in many problems related to soil pollution and environmental contamination.

Recycling residual film in cotton fields is divided into manual and mechanical recycling. Mechanized residual film recycling is more effective than manual recycling. At present, mechanized recycling is dominated by post-autumn recycling which is divided into joint and segmented operations. The joint operations mean the machine can complete stalk whipping and residual film recovery simultaneously after the cotton harvest. Segmented operations are different. The cotton stalk whipping operation is first carried out in segmented operations. After the whipping is completed, the film recovery operation is performed. The machine for segmented operations has a simple structure, high reliability, and low energy consumption [1]. At present, segmented operations rely on the manual

driving of tractors to pull working tools. Due to the large cotton plantation area, the driver must work continuously for an extended period. The high labor intensity makes it difficult to ensure that the cutter is aligned effectively with the edge of the film to pick up the film. This leads to the missed collection of the remaining film. In addition, the working environment is harsh, there is a high dust pollution level, and the work is monotonous. By developing a navigation operation system for residual film recovery, it is possible to significantly reduce the labor intensity of drivers and improve the efficiency of residual film recovery operations.

At present, satellite navigation is widely used in the cultivation of cotton in Xinjiang. However, unmanned operations can be affected by poor satellite signals or farm environment changes. Machine vision can detect navigation paths in real time and correct deviations [2,3]. The navigation line extraction relies on crop detection. Fast and high-accuracy crop detection algorithms can ensure the effectiveness of the navigation line in practical applications [4].

Common crop detection algorithms use color [5–7], texture [8–10], and shape [11–13] to extract the target. Denoising the extracted targets to determine the crop points. The final navigation line is obtained by fitting the feature points [14–16]. The single crop detection algorithm detects crops in images and is fast. However, it is sensitive to environmental changes. For example, different backgrounds or lighting conditions can affect the crop's color and texture [17].

With the development of machine learning, related models have been used for crop visual detection [18–21]. Machine learning models classify target features and by optimizing the model parameters or structure, the image is detected according to the input features. The quality of feature extraction and the number of features determine the model's recognition effect. The accuracy and applicability of the machine learning models are better than common crop row detection algorithms [22]. However, the collected features limit its applications to other scenarios. The detection effect of the machine learning models will be reduced if the feature of the image has not been trained by the model and the feature has changed significantly. In addition, as the detection features increase, the model's processing time will increase. This is unsuitable for scenes with high real time requirements, such as the visual navigation of residual film recovery.

Deep learning has high accuracy and broad applicability compared with other machine learning and common crop detection algorithms. Bah Mamadou Dian et al. [23] proposed a model consisting of SegNet and Convolutional Neural Network-based (CNN) Hough transform to detect crop rows with 93.58% accuracy. Adhikari et al. [24] use a deep neural network-based semantic map method to extract crop rows. This method accurately detects the number of rows of rice. Mora-Fallas et al. [25] proposed an instance segmentation method based on the Mask Region Convolutional Neural Networks (Mask R-CNN) model which can effectively detect farmland weeds and crops. Menshchikov Alexander et al. [26] reported a fast and accurate hogweed detection method based on a fully convolutional neural network. This method identifies the size of hogweed individuals and leaves. Khan Shahbaz et al. [21] developed a deep learning system for identifying weeds and crops on farmland. The system's recognition accuracy of crops and weeds reaches 94.73%. Afonso Many et al. [27] reported the results of using the Mask RCNN algorithm to detect tomato images in the greenhouse. The results show that the detection effect based on the deep learning method is better than the results reported in earlier work. Alzadjali Aziza et al. [28] developed and compared two tassel automatic detection methods based on deep learning models. The F1 score based on the convolutional neural network is 95.9%, and the F1 score of Faster Region CNN (Faster R-CNN) is 97.9%. Osorio Kavir et al. [29] compared Support Vector Machines (SVM), YOLOv3, and Faster RCNN in weed detection. Experiments show that deep learning's F1 score is better than SVM. André Silva Aguiar [30] uses deep learning methods to detect vineyard grape clusters at different growth stages. The results show that the constructed model detects grape bunches better. A review of the literature on related crop detection indicates that deep learning has significant potential for crop detection.

Rapid and accurate stubble detection has high research value. However, there are no related literature reports on the crop detection problem in residual film recovery. The color characteristics of the stubble after cotton harvest and stalk whipping are similar to those of cotton stalks, boll husks, leaves, and soil. These disturbances can obscure or extend the shape of the stubble, making it difficult for the accuracy and real time performance of image processing techniques and machine learning algorithms to meet the requirements. Among the deep learning detection models, YOLOv3 is the widely used target detection algorithm at present [17,31–35]. The model balances accuracy and real-time metrics. The model structure is stable and easy to implement for deployment. It can detect cotton stubble in real time. Based on this information, an improved YOLOv3 algorithm is proposed for cotton stubble detection. This work aims to achieve rapid and accurate detection of cotton stubble and provide reliable technical support for visual navigation of residual film recovery after autumn.

The rest of the article is organized as follows. In Section 2, data collection and labeling were carried out. In Section 3, the YOLOv3 model is improved and trained. Section 4 analyzes the improved YOLOv3 model test results. Section 5 discusses the results. Section 6 gives conclusions.

2. Materials and Methods

2.1. Data Acquisition

The images were taken at the recovery sites of the 145th Regiment, 146th Regiment, and 152nd Regiment of the Eighth Division of the Corps. The image resolution is 640 pixels $\times$ 480 pixels. The image acquisition camera is a Wild Forest wide-angle lens (130° wide-angle), and the camera is installed on the front counterweight of the tractor as shown in Figure 1 [36]. The camera depression angles are 10°, 30°, and 45°. The captured images include sunny, cloudy, forward light, backlight, and abnormal driving. A total of 1800 images were acquired. After adjusting the brightness and enhancing the noise, 2110 images were finally obtained.

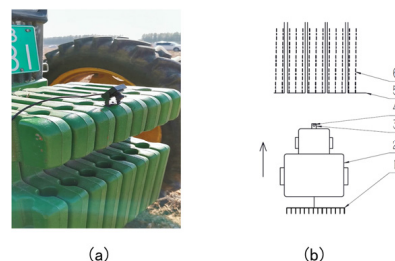


Figure 1. (a) the image of the camera installation; (b) the schematic diagram of the image acquisition process (arrow: direction of machine movement). 1. Residual film recovery equipment 2. Tractor 3. Counterweight 4. Camera 5. Residual film 6. Stubble rows.

2.2. Algorithm Flow

The navigation line extraction algorithm for residual film recovery after autumn based on improved YOLOv3 is divided into two stages: target detection and navigation line fitting. The overall process is shown in Figure 2.

- (1) Target detection: Labeling the stubble in the collected RGB image of the cotton field stubble. Construct the YOLOv3 training set. The optimal detection model is obtained by training the improved YOLOv3. According to the driving habit of residual film recovery, the stubble row area facing the tractor is selected to construct the ROI. Use the model to locate the stubble in the ROI and output its prediction frame.
- (2) Navigation line extraction: Save the stubble position information successfully detected and output by the model. Remove the stubble false detection points by denoising. Extract (x, y) and (x + w, y + h) from the stubble detection frame as feature points. The

navigation line can be obtained by fitting the characteristic points of the stubble by using the least square method.

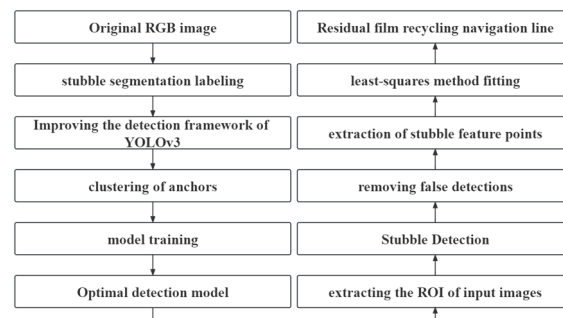


Figure 2. The overall flow of the algorithm.

2.3. Labeling Data Sets

Labeling the stubble data set is a crucial step before model training. Labeling directly affects training and test accuracy if the labeling box contains noise other than the target. There is a difference in the shape and size of the stubble. The labeling box can directly select the entire stubble outline for regular stubbles with an upright shape. When the stubble has branches and the branches are relatively thick, marking the outline of all the branches will result in too much noise, such as residual film, soil, and broken leaves and too much noise will result in inaccurate predictions. The article adopts a segmented labeling method for stubble shapes with branches. Figure 3 shows the schematic diagram. The central part is labeled with one label box, and the branch section is labeled with another. However, not all stubble with branches is labeled. The stubble with little difference in thickness from the central part is labeled as the sample. If the branch is thin, it is not included in the stubble data set since the thin branch and broken stubble characteristics are similar which can affect the prediction.

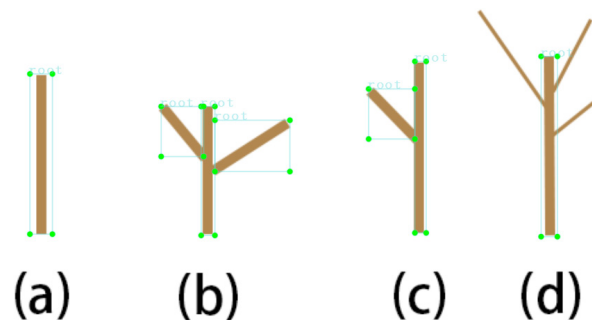


Figure 3. Label schematics (a) regular stubble with an upright shape, (b) little difference in thickness from the central part, (c) little difference in thickness from the central part and (d) branch is thinner than the central part.

Use the Labellmg tool to label each image one by one and generate the corresponding “.xml” location information file. The target frame position information in “.xml” format is normalized and converted into “.txt” text. Each “.txt” file contains the category number “c” of the stubble, the upper-left coordinate information (x, y) of the bounding box, and the width and height of the box (w, h). In this article, we only need to detect stubble targets, so the total number of categories is one. The data set is sorted based on the VOC2007 data set format. The images in the dataset are randomly divided into a training set, validation set, and test set according to a ratio of 7:2:1. An improved YOLOv3 model is trained on the training set. The validation set is used to calculate the indicators of the trained model and select the best training model. The test set is used to check the generalization ability of the best training model.

3. Stubble Detection Model Based on Improved YOLOv3

3.1. YOLOv3 Detection Model

(You Only Look Once v3) YOLOv3 converts target detection into a regression problem based on an end-to-end network [37] from the input of the original image to the output of the final stubble detection frame. The YOLOv3 model first unifies the input image to $416 \text{ pixels} \times 416 \text{ pixels}$ and then divides the image into grids of different sizes. Each grid is responsible for detecting objects that fall within the grid. The bounding box of the obtained object contains five predictions: x , y , w , h , and c . The (x, y) is the coordinates of the upper left corner of the detection frame, w and h are the width and height of the detection frame, and c is the confidence of the stubble category. The input image of the whole step passes through a Darknet-53 detection backbone to obtain five predictions of the stubble in the image.

The original YOLOv3 backbone is shown in Figure 4. After the input image passes through the Darknet-53 backbone, there are three detection grids (52×52), (26×26), and (13×13) responsible for target detection. The grid allocates three anchors of different sizes to predict the bounding box. Finally, the three output anchors were merged to obtain prediction information about the stubble.

Type	Filters	Size	Output
Convolutional	32	3×3	416×416
Convolutional	64	$3 \times 3/2$	208×208
Convolutional	32	1×1	
Convolutional	64	3×3	
Residual			208×208
Convolutional	128	$3 \times 3/2$	104×104
Convolutional	64	1×1	
Convolutional	128	3×3	
Residual			104×104
Convolutional	256	$3 \times 3/2$	52×52
Convolutional	128	1×1	
Convolutional	256	3×3	
Residual			52×52
Convolutional	512	$3 \times 3/2$	26×26
Convolutional	256	1×1	
Convolutional	512	3×3	
Residual			26×26
Convolutional	1024	$3 \times 3/2$	13×13
Convolutional	512	1×1	
Convolutional	1024	3×3	
Residual			13×13
Avgpool		Global	
Connected		1000	
Softmax			

Figure 4. Darknet-53 network before modification.

3.2. Improve the Detection Framework of the YOLOv3 Model

The shape of the stubble after the cotton harvest is diverse. The stubble is divided into two representative shapes, the first is the thicker stem, and the second is the thinner stem. Statistics were run on the w and h generated by LabelImg to determine the variation range of the two types of stubble, where w is the width of the rectangular box, and h is the height of the rectangular box. The values of w and h reflect stubble width and height. The varied range of w and h is obtained through statistics, and the distribution is shown in Table 1.

Table 1. Stubble size distribution.

Type	Size Distribution
finer stubble	$(w = 1, h = 18) \sim (w = 4, h = 30)$
Thick stubble	$(w = 4, h = 14) \sim (w = 15, h = 36)$

The three grid sizes detected for the original YOLOv3 feature extraction network are 52×52 , 26×26 and 13×13 , respectively. When the input image size is 416×416 , the pixel size corresponding to the grid is 8×8 , 16×16 and 32×32 . We can observe that as the number of grids decreases, the size of the corresponding detection target increases. The 13×13 grid detection target is a 32×32 -pixel block, but the largest target for stubble is 15×36 . From the statistical results in Table 1, most stubble sizes are unsuitable for detection on a 13×13 grid. The smaller stubble features are easily compressed when the resolution is reduced from 416×416 to 13×13 . This results in a serious loss of stubble

information and missed detection. The 13×13 grid tends to detect large targets, and the stubble is relatively small. Therefore, for the original YOLOv3 detection network, this article removes the 13×13 detection layer and replaces 13×13 with a 104×104 detection layer. The 104×104 grid can detect 4×4 -pixels blocks, compared to the 13×13 grid which can complete the prediction of finer stubble. The modified detection grids are 104×104 , 52×52 , and 26×26 . The modified detection backbone is shown in Figure 5. Compared to Figure 4, we remove the 13×13 grid output and keep the rest of the network parameters unchanged. The improved overall YOLOv3 framework is shown in Figure 6. The basic structure of DBL includes convolution (conv), batch normalization (BN), and leaky relu operation. The residual unit (res unit) includes two DBL components. Multiple residual units (resn) contain n residual units. The outputs of res1, res2, res8, and res8 correspond to the 208×208 , 104×104 , 52×52 , and 26×26 grid outputs in Figure 5.

Type	Filters	Size	Output
Convolutional	32	3×3	416×416
Convolutional	64	$3 \times 3/2$	208×208
Convolutional	32	1×1	
Convolutional	64	3×3	
Residual			208×208
Convolutional	128	$3 \times 3/2$	104×104
Convolutional	64	1×1	
Convolutional	128	3×3	
Residual			104×104
Convolutional	256	$3 \times 3/2$	52×52
Convolutional	128	1×1	
Convolutional	256	3×3	
Residual			52×52
Convolutional	512	$3 \times 3/2$	26×26
Convolutional	256	1×1	
Convolutional	512	3×3	
Residual			26×26
Avgpool		Global	
Connected		1000	
Softmax			

Figure 5. The modified Darknet-53 network.

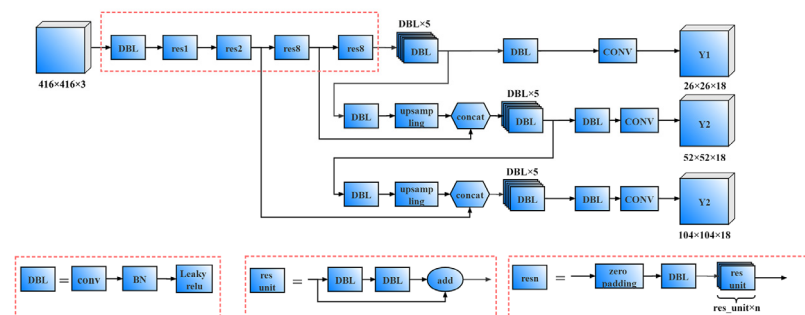


Figure 6. Improved YOLOv3 model framework.

3.3. Clustering of Anchor

The anchor parameter is introduced in the YOLOv3 algorithm [37]. The anchor is a set of a priori boxes with fixed width and height values. In stubble detection, the size of the prior frame directly affects training and detection accuracy. The original priori box size cannot match the improved detection network well. Prior verification needs to be re-clustered to obtain its size before model training.

The original YOLOv3 model uses the K-means algorithm which has large randomness in selecting the initial clustering center. Randomness will cause clustering results to differ from the optimal global solution. According to the improved detection network and stubble label, K-means++ is used instead of the K-means algorithm. The clustering algorithm aims to make the Intersection over Union (IOU) between the anchor box and the ground truth

(the labeled bounding boxes) as large as possible. The objective function uses IOU as the measurement standard. The distance formula is defined as follows:

$$D = \min \sum_{box=0}^n \sum_{cen=0}^k [1 - \text{IOU}_{cen}^{box}]. \quad (1)$$

Among them, the *box* is the target box of the sample label, *cen* is the cluster center, and *n* and *k* represent the number of samples and categories in the data set. Appropriate IOU can well weigh the complexity of the model and the detection recall rate. Figure 7 shows the clustering effect of the kmeans++ algorithm on the stubble data set. The prior frame size and IOU value under different *k* values are sequentially calculated. The comparison (Table 2) shows that the IOU reaches a higher level when *k* = 9. As the value of *k* increases, the increasing trend of the IOU is not obvious. Too much *k* value will affect detection speed. Finally, nine anchor values are determined by kmeans++, the size of which is (3, 15), (4, 10), (5, 16), (7, 20), (12, 29), (21, 23), (29, 27), (37, 31), and (42, 44). The three grid types 104 × 104, 52 × 52, and 26 × 26 are detected based on the size distribution of anchors, and each grid scale may predict three anchors.

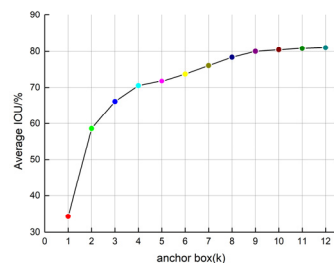


Figure 7. Average IOU corresponds to different *k* values.

Table 2. Anchor point size of different *k* values.

k = 6	k = 7	k = 8	k = 9	k = 10	k = 11	k = 12
(5, 15)	(5, 14)	(3, 11)	(3, 15)	(4, 10)	(4, 11)	(3, 13)
(8, 21)	(7, 18)	(5, 15)	(4, 10)	(4, 16)	(5, 21)	(4, 10)
(17, 27)	(9, 26)	(8, 21)	(5, 16)	(5, 21)	(6, 16)	(5, 14)
(25, 21)	(21, 21)	(21, 21)	(7, 20)	(6, 16)	(8, 23)	(5, 19)
(29, 29)	(25, 31)	(21, 31)	(12, 29)	(8, 24)	(17, 31)	(8, 22)
(37, 35)	(29, 25)	(29, 25)	(21, 23)	(17, 23)	(21, 21)	(17, 23)
	(37, 37)	(33, 25)	(29, 27)	(25, 25)	(29, 21)	(25, 29)
		(37, 37)	(37, 31)	(29, 37)	(29, 27)	(29, 21)
			(42, 44)	(33, 21)	(29, 40)	(29, 40)
				(42, 37)	(33, 33)	(33, 29)
					(42, 37)	(37, 29)
						(46, 44)

3.4. Removing False Detections

Due to the complex environmental conditions in the field image recovery of residual film and a lot of noise interference during the detection phase, it is easy for parts of the interference to be misidentified as stubble during the subsequent navigation line extraction process. This article adopts a denoising method based on the mean value to remove false detection noise at the detection stage.

First, calculate the stubble position distribution in the image. In the collected 640 × 480 resolution image, the detected ROI area has been set as the stubble row area (100 × 200 resolution ROI) driven directly by the tractor. The camera depression angle will affect the stubble position information. The camera depression angles of 10°, 30°, and 45° are considered in statistical images. Through the statistics of 300 images, the distance between the two rows of stubble rows facing the tractor is distributed between 25 and 50 pixels as shown in Figure 8. Select

the distance $d1 = 12.5$ and $d2 = 25$ as the distance threshold. This threshold is used only to determine the pixel coordinates and is not involved in the subsequent processing of pixels. Therefore, the threshold is not rounded.

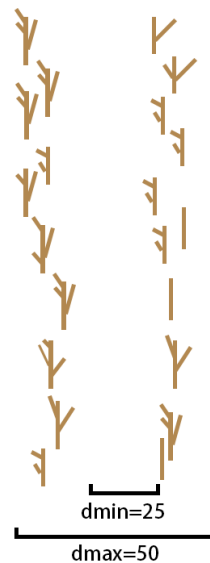


Figure 8. Schematic diagram of stubble distribution.

After completing the stubble detection, x of the detection frame provides useful output information, x is stored as a representative coordinate point of one detection frame, and all representative coordinate points and index information are saved. Calculate the average value X_{mean} of all representative coordinate points. The calculation formula is as follows:

$$X_{\text{mean}} = \frac{\sum_{i=0}^n x_i}{n} \quad (2)$$

Check whether x_i information satisfies $X_{\text{mean}} - d1 < x_i < X_{\text{mean}} + d1$, $x_i < X_{\text{mean}} - d2$, and $x_i > X_{\text{mean}} + d2$. If x satisfies one of the three conditions, it can be judged that the detection frame of x is outside the area of two rows of stubbles which is a false detection point. In the subsequent navigation line fitting, this point's x value and index information are removed from the feature points to be fitted.

3.5. Model Training

This experiment is based on the Windows 10 operating system. The GPU is NVIDIA GeForce RTX 2080 (8 GB video memory), the processor is an Intel Core i7-9700k, and the running memory is 32 G. The model building and training are implemented in Python based on the PyTorch deep learning framework, and the parallel computing framework uses CUDA version 10.0.

Before training, the image size is scaled to 416 pixels $\times$ 416 pixels, the batch_size is set to 4, the initial learning rate is set to 0.001, the weight decay is set to 0.0005, and the momentum decay is set to 0.9. The nine anchors of model training are the anchors after K-means++ re-clustering.

Figure 9 shows the average Loss value change curve during the improved YOLOv3 detection network. It can be seen in Figure 9 that when the number of network training exceeds 6000 training epochs, the Loss value stabilizes around 0.1. The network training result is ideal because of the subsequent convergence of the Loss value.

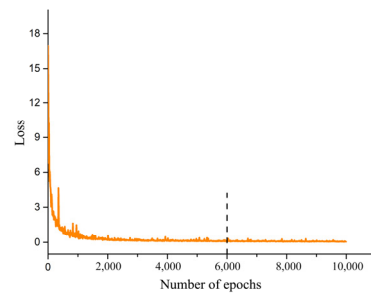


Figure 9. Change of Loss value. dashed line: Loss values stabilise.

It is not that more epochs are better for model training. Too much training can overfit the model. A training model must be evaluated to determine the most effective detection model. The model's evaluation index selects mean average precision (mAP), Precision (P), and Recall (R), and the calculation formula is as follows:

$$mAP = \frac{1}{C} \sum_{i=1}^N P(i) \Delta R_0(i) \quad (3)$$

$$P = \frac{T_P}{T_P + F_P} \quad (4)$$

$$R = \frac{T_P}{T_P + F_N} \quad (5)$$

where C : Sample Categories, T_P : correctly divided into the number of positive samples, F_P : the number of negative samples that are incorrectly divided into positive samples, F_N : the number of positive samples that are incorrectly divided into negative samples, P : Precision (that is, the proportion of correctly detected stubble samples in the detection of positive samples), R : Recall (that is, the proportion of correctly detected stubble samples to the stubble samples in the dataset), and mAP : mean average precision. Reference to the common metric mAP_{50} for target detection used in the original YOLOv3 [37]. Our definition of a correct detection sample is one in which the IOU between the stubble detection box and the labeled bounding boxes is greater than 50%.

Figure 10 shows the model's mAP on the validation set. According to Figure 10, after 9000 training epochs, the overall mAP of the model is relatively stable, and several models reach 0.925. This is the highest mAP among the models. Ultimately, the model trained 9000 epochs were selected as the optimal model. The optimal model's Precision, Recall, and mAP are 0.86, 0.971, and 0.925, respectively. The single image detection time under this model was 98.6 ms.

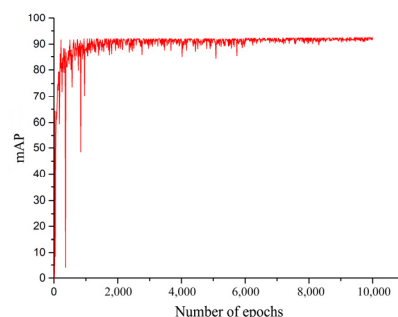


Figure 10. Mean average precision of the model.

4. Results and Analysis

Select different locations, time periods, camera depression angles, and the YOLOv3 detection model before and after the improvement test. The size of the image collected by

the camera is 640×480 resolution. The stubble row directly opposite the tractor operation is extracted with a 100×200 resolution ROI and used as the stubble row detection area. The ROI area is scaled to 416×416 resolution as the input image for the detection model. The following subsections will give specific test results.

4.1. The Detection Effect of Different Locations

Select residual film recovery field images from different locations for algorithm detection. The main differences between the images are the soil color and the distribution of broken leaves and stalks. All types of images can detect stubble with ideal results. Select the four images Figure 11a,c,e,g to represent the four types of places, and Figure 11b,d,f,h are the corresponding detection results. According to the test results, the color of the soil in Figure 11a is similar to that of the stalk and much of the stubble in the image has been incorporated into the soil. Human eyes cannot see stubble intuitively. Under such conditions, the improved YOLOv3 algorithm can still complete detection without missed or false detection. In Figure 11c, the residual film is evenly distributed. The stubble at the top of the image is thinner than that at the bottom. The detection results were better for detecting both small and large target stubbles. A large amount of residual film on the ground does not affect this algorithm. Many boll husks, broken leaves, and broken stalks are visible in Figure 11e. The color is similar to that of stubble which obscures the shape of the stubble. In this case, the algorithm can still detect the stubble, and the effect is ideal. Figure 11g shows this is the least noisy of the four place types. The stubble characteristics and shape are apparent in the image. The algorithm detects all of the stubble.

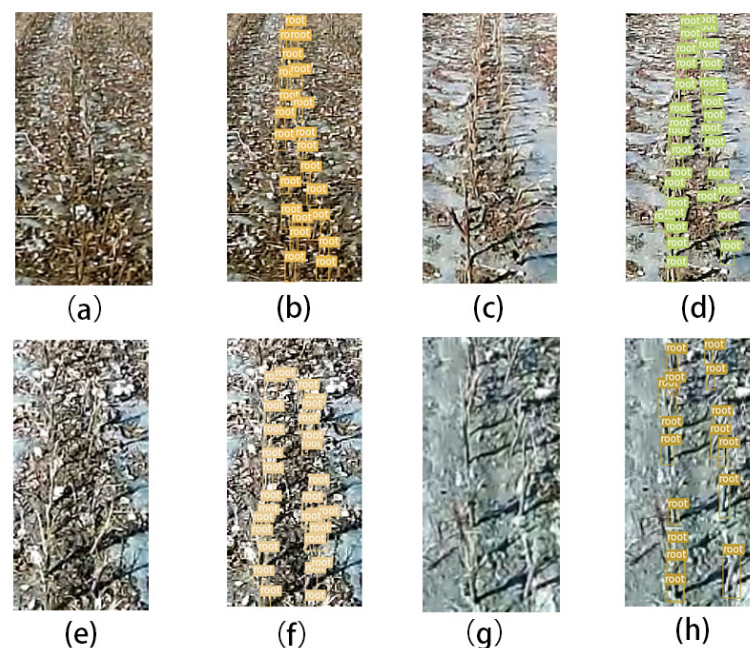


Figure 11. The detection effect of different locations: (a,c,e,g) are the original images, (b,d,f,h) are the improved YOLOv3 algorithm detection images.

The overall detection rate of stubble is relatively high in several different locations. The soil color, residual film, broken leaves, and stalks distribution do not affect the detection rate. In addition, the algorithm for detecting the finer branches of stubble may be missed since the labeled data set does not label finer branches. However, the central stalk and the thick branches can be correctly identified.

4.2. Detection Effect in Different Time Periods

Select residual film recovery field images from different time periods for algorithm detection, and the images are from the same place. Select the four images Figure 12a,c,e,g

are the original images, and Figure 12b,d,f,h are the corresponding detection results. The periods are morning, noon, and afternoon. Figure 12a shows a morning photograph. The influence of morning light is not apparent. The algorithm can generally detect the presence of stubble on the image. There is no missed and false detection, and the effect is better. Figure 12c shows the area at noon. The light makes the overall image brighter. The stubble is brighter overall due to sunlight, but intense light does not affect the overall detection effect. It is worth noting the shadow effect brought by the light. The stubble in the shadow and the central part of the stubble are blocked by the shadow, easily leading to missed detection. Figure 12e,g shows the scene in the afternoon. The light is not as intense as noon, and the overall detection effect of stubble is acceptable. There are two examples of missing stubbles in Figure 12h because the wheel has crushed the part, and the stubble is absent. The algorithm is generally less affected by time and ensures stubble detection stability and accuracy under long-term continuous operation.

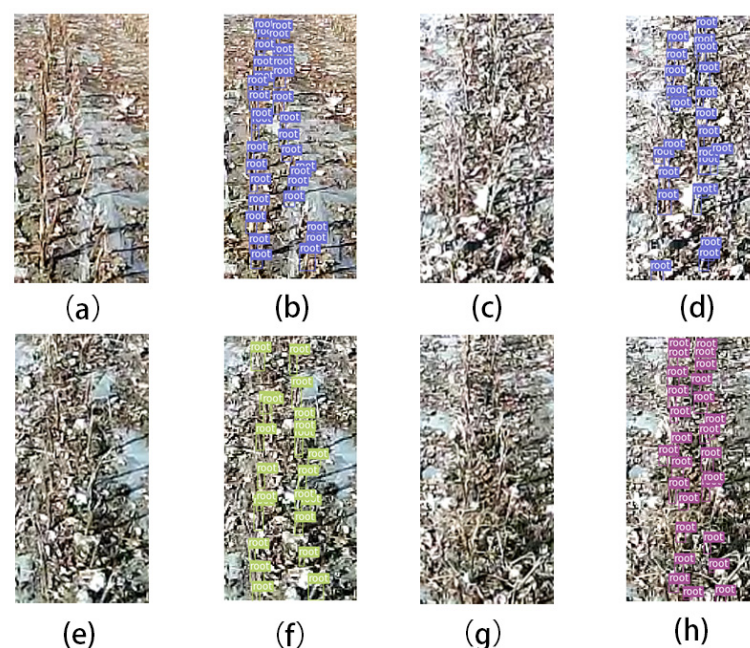


Figure 12. The detection effect of different time periods: (a,c,e,g) are the original images, (b,d,f,h) are the improved YOLOv3 algorithm detection image.

4.3. The Detection Effect of Different Camera Depression Angles

Select the residual film recovery field images under different camera depression angles for algorithm detection. The main difference between the images is the stubble size. The size of the stubble in the lower part of the ROI will increase as the depression angle of the camera increases. The camera depression angle includes 10° , 30° , and 45° . Figure 13a,b show the results of stubble detection at 10° and 30° . There is little overall difference between the stubble images collected from the two angles, and stubble can be detected without error. Figure 13c illustrates an image taken at a 45° angle for the depression. The stubble in the image is larger than the stubble at a low depression angle. According to the test results, the larger stubble detection effect is optimal, and no obvious stubbles were missed. The difference with the low depression angle is that the stubble is sparser at a 45° depression angle. As the depression angle increases, the detected stubble target becomes larger, and the detected stubble target frame distribution becomes scattered. The algorithm is generally unaffected by the camera depression angle, and stubble under different depression angles can be detected correctly.

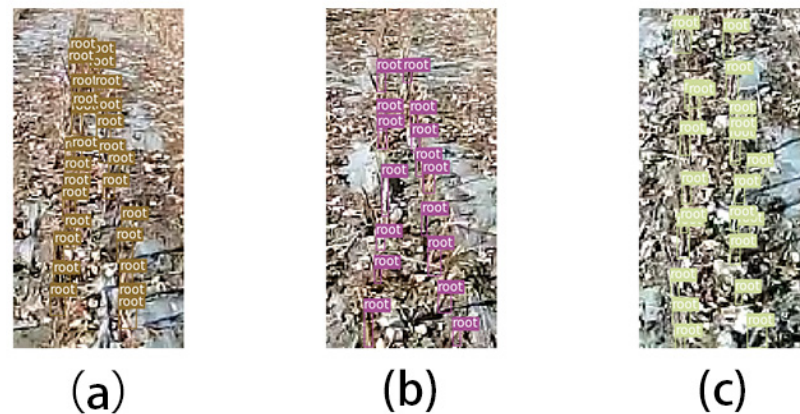


Figure 13. Detection effect of different camera angles (a) The camera depression angle of 10° (b) The camera depression angle of 30° (c) The camera depression angle of 45° .

4.4. Comparison of the Effect of the YOLOv3 Detection Algorithm before and after the Improvement

To further verify the stubble detection effect of the improved YOLOv3 algorithm. Choose the YOLOv3 algorithm and the algorithm of this article to carry out the comparison experiment of stubble detection. The YOLOv3 algorithm uses the original Darknet-53 network in the comparative test. The same sample set is used for algorithm training and testing. After 4000 training epochs, the model's Loss value stabilizes around 0.1. Finally, the model trained for 6000 epochs is selected as the optimal detection model for YOLOv3.

The comparison result is shown in Figure 14 where (a,d,g,j) are the original images. The background of Figure 14a is complex, and the soil color is similar to stubble. Part of the stubble appears to blend into the background. It is difficult for the human eye to distinguish the stubble intuitively. Figure 14b shows the result of YOLOv3. Although there is no false detection in the result, the missing stubble detection is serious. Compared with the detection result in Figure 14c of the algorithm in this article, most of the stubble was not successfully detected. It is worth noting that neither the algorithm in this article nor YOLOv3 can detect stubble in the upper part of the image. The main reason is that the part of the stubble in the image is relatively small, and the soil and broken stalks cover the outline of the stubble seriously. The stubble outline has completely blended into the background. Both types of algorithms are unable to detect stubble in this case. Figure 14d,g are located in areas with relatively more impurities, such as broken stems and boll leaves. Unlike the previous chapter's images, these two representative images were taken with a camera shake. The overall image is unclear compared to other images. Shaking can blur some stubble information in the image. Both detection algorithms cannot detect blurred stubble. The algorithm proposed in this article can detect some stubble with less blur, but YOLOv3 still cannot. Figure 14j has a more obvious stubble characteristic. Judging from the detection results, the two types of algorithms for the stubble below the image can detect well. The biggest difference is the upper area of Figure 14j. The shape of the stubble at the upper area is smaller and thinner. YOLOv3 has seriously missed detections for such small targets. The improved algorithm in this article can complete the detection of such small targets.

The improved YOLOv3 stubble detection effect is generally better than the YOLOv3 model. It can complete the stubble detection task on a complex background and provide the basis for the next step of navigation line extraction.

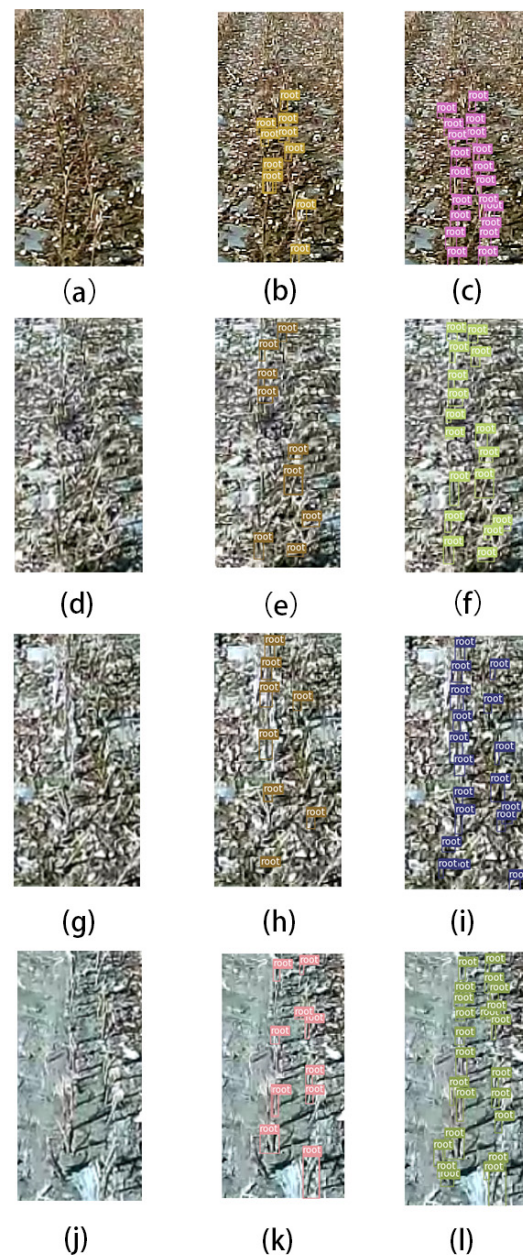


Figure 14. The detection effects of different algorithms: (a,d,g,j) are the original images, (b,e,h,k) are the detection results of the original YOLOv3 algorithm, and (c,f,i,l) are the improved YOLOv3 algorithm detection results.

4.5. Navigation Line Fitting

For the 100×200 resolution ROI area, stubble can be detected by the improved YOLOv3 model. After the misidentification points are removed, the feature points are used as candidate points for stubble navigation line fitting. The input feature point information is (x, y) , $(x + w, y + h)$, the coordinates of the upper left corner and the lower right corner of each stubble detection frame selection box to fit the navigation line. This article selects the least-squares method with an ideal fitting effect and fast calculation speed to fit feature points to a straight line. The successfully fitted straight line is the navigation line for the residual film recovery operation.

The navigation line detection result is shown in Figure 15. According to Figure 15c,f, the high accuracy of stubble detection ensures the reliability and stability of navigation line extraction. The navigation lines in the two images fit well. In Figure 15i, although the navigation line is successfully fitted, there is an error in the angle of the navigation line. It

is primarily because the navigation line fits between two stubble rows. The cotton planting mode of the image collection fields is wide and narrow row planting. The distance between the two narrow rows is closer. When the camera depression angle is small, the two narrow lines are very close in the image, and the two lines can be directly fitted to extract the navigation line. When the camera depression angle is large, the distance between the two lines will be relatively obvious in the image. If the two-line stubble fitting method is applied, it can cause a deviation from the overall fitting line due to the stubble miss detection.

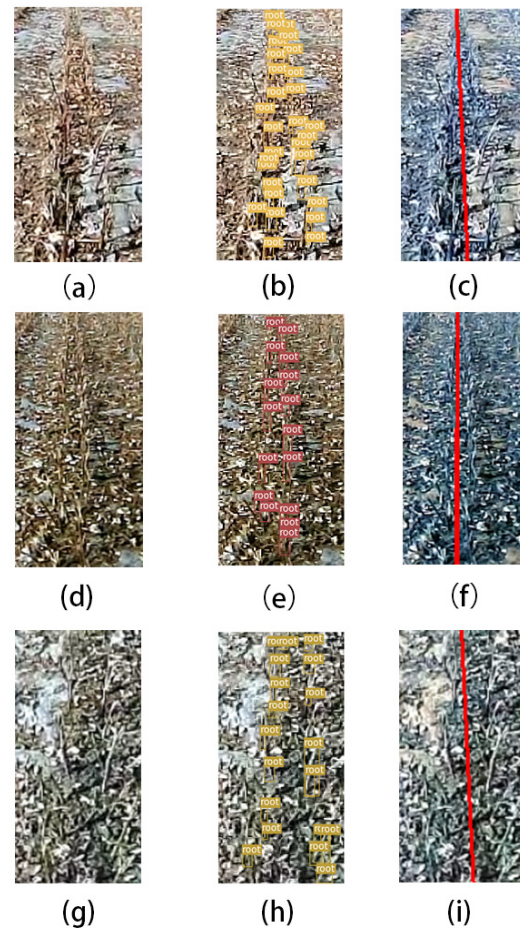


Figure 15. Navigation line detection effect: (a,d,g) are the original images, (b,e,h) are the detection results of the algorithm in this article, (c,f,i) are the detection results of the navigation line (red line).

Through experiments, we have found that when the camera depression angle is less than 30° , the two narrow lines in the collected image are relatively close. A more accurate navigation line can be obtained by direct fitting using the least square method. When the depression angle is large or the planting mode does not use wide and narrow row planting, the least square method may not provide a satisfactory fitting result. At this time, we should analyze the specific scene in detail. However, as far as the stubble detection algorithm in this article is concerned, the accuracy of stubble detection is high. This can ensure the accuracy of the navigation line fitting feature points. In follow-up research, different navigation line fitting algorithms should be used in different scenarios to complete the extraction process.

5. Discussion

Compared to previous stubble detection algorithms using fused features [36]. Although the detection accuracy was lower than the fusion features, the detection speed was significantly faster. Real-time performance is one of the most important metrics for visual navigation tasks. The algorithm in this article sacrifices some detection accuracy, but the

speed of navigation line fitting can ensure the effective deployment of the model. Mazzia et al. [32] demonstrate that YOLOv3 can be deployed in low-power embedded devices for real-time detection.

Comparing this article's algorithm with the related YOLOv3 literature, we find many shortcomings and areas for improvement. First, the processing time of the improved YOLOv3 algorithm in this article can be further optimized. Yang Li's study [34] showed that the YOLOv3 model processed an image in 13 ms by replacing the more lightweight MobileNetV2 backbone. Second, fusion manual features have high accuracy but consume large computational resources. However, a few manual features can be combined with the YOLOv3 backbone network for detection. Zhiheng Lu's study [38] showed that the YOLOv3 model combined with manual features improved detection accuracy, and single image detection took 420 ms. Processing speed and accuracy are opposite metrics. If the operation speed is slow, this time will be sufficient to meet the requirements of the operation. Third, image preprocessing can be used to optimize blurry stubble in images. Xudong Zhang's study [33] performs bilateral filtering and other operations on the input image before applying YOLOv3 detection. The image details can be improved, and the detection accuracy can be improved by about 2% to 3%.

Fourth, the camera depression angle affects target detection results. The most effective camera depression angle obtained by Wenkai Xu [35] was 30–60°, similar to the results in Section 4.3 of this article. The stubble detection at 45° was the most accurate of all images. Although the results are consistent with those reported by Wenkai Xu [35], the range of camera depression angles tested in this article is relatively small. Fifth, Wenkai Xu's study [35] also mentions that model hyperparameters impact model accuracy. For example, batch_size and epoch are mentioned in the article, and we also find similar results in this article. Even though the training Loss value of the model is stable, the accuracy will show some small fluctuations as the epoch increases. However, there are many model hyperparameters, and the impact deserves further study.

6. Conclusions

This article proposes a stubble detection method based on improved YOLOv3 to address the problem of visual navigation path extraction in this scene. The residual film recovery field images at different locations, time periods, and camera depression angles were selected for stubble detection experiments. The results show that the improved YOLOv3 algorithm proposed in this article can effectively accomplish the stubble detection task. By comparing the algorithm in this article with the original YOLOv3 algorithm, the detection effect is better in normal scenes, complex scenes, and jittery images. This shows that the algorithm in this article is more efficient and accurate.

The algorithm presented in this article detects stubble satisfactorily, but there are still shortcomings regarding navigation line fitting. When the distance between two rows of stubble in the image is too large or the stubble rows are partially missing, it will cause path extraction errors. In addition, when different planting patterns and stubble rows have large deflection angles, the article's noise point removal threshold and ROI selection need to be further considered. In the next step, we will focus on applying the path extraction algorithm to the navigation field test. Optimize the parameters in the article to make the visual navigation algorithm more stable and reliable in actual use.

Author Contributions: Conceptualization, Y.Y. and J.L.; methodology, Y.Y.; software, Y.Y.; validation, Y.Y. and J.L.; formal analysis, Y.Y. and J.L.; investigation, S.Y.; resources, J.L.; data curation, J.T.; writing—original draft preparation, Y.Y.; writing—review and editing, Y.Y., J.L. and J.N.; visualization, Y.Y.; supervision, J.L.; project administration, J.L.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully appreciate the financial support provided by the National Natural Science Foundation of China (52175240).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the privacy policy of the organization.

Conflicts of Interest: The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this article.

References

- Hu, C.; Wang, X.F.; Chen, X.G.; Tang, X.Y.; Zhao, Y.; Yan, C. Current situation and control strategies of residual film pollution in Xinjiang. *Trans. Chin. Soc. Agric. Eng. (Trans. CSAE)* **2019**, *35*, 223–234. (In Chinese with English Abstract).
- García-Santillán, I.; Guerrero, J.M.; Montalvo, M.; Pajares, G. Curved and straight crop row detection by accumulation of green pixels from images in maize fields. *Precis. Agric.* **2018**, *19*, 18–41. [\[CrossRef\]](#)
- Li, Y.; Hong, Z.; Cai, D.; Huang, Y.; Gong, L.; Liu, C. A SVM and SLIC Based Detection Method for Paddy Field Boundary Line. *Sensors* **2020**, *20*, 2610. [\[CrossRef\]](#)
- Yuhao, B.; Zhang, B.; Xu, N.; Zhou, J.; Shi, J.; Diao, Z. Vision-based navigation and guidance for agricultural autonomous vehicles and robots: A review. *Comput. Electron. Agric.* **2023**, *205*, 107584. [\[CrossRef\]](#)
- Cheng, H.D.; Jiang, X.H.; Sun, Y.; Wang, J. Color image segmentation: Advances and prospects. *Pattern Recognit.* **2001**, *34*, 2259–2281. [\[CrossRef\]](#)
- Liao, J.; Wang, Y.; Yin, J.; Lu, L.; Zhang, S.; Zhu, D. Segmentation of Rice Seedlings Using the YCrCb Color Space and an Improved Otsu Method. *Agronomy* **2018**, *8*, 269. [\[CrossRef\]](#)
- Tang, L.; Tian, L.; Steward, B.L. Color image segmentation with genetic algorithm for in-field weed sensing. *Trans. ASAE* **2000**, *43*, 1019–1027. [\[CrossRef\]](#)
- Girolamo Neto, C.; Sanches, I.; Neves, A.; Prudente, V.; Körting, T.; Picoli, M.; Aragão, L. Assessment of Texture Features for Bermudagrass (*Cynodon dactylon*) Detection in Sugarcane Plantations. *Drones* **2019**, *3*, 36. [\[CrossRef\]](#)
- Jiang, B.; Wang, P.; Zhuang, S.; Li, M.; Li, Z.; Gong, Z. Detection of maize drought based on texture and morphological features. *Comput. Electron. Agric.* **2018**, *151*, 50–60. [\[CrossRef\]](#)
- Sabzi, S.; Pourdarbani, R.; Ignacio Arribas, J. A Computer Vision System for the Automatic Classification of Five Varieties of Tree Leaf Images. *Computers* **2020**, *9*, 6. [\[CrossRef\]](#)
- Calixto, R.; Neto, L.; Cavalcante, T.; Facundo Aragão, M.; Silva, E. A computer vision model development for size and weight estimation of yellow melon in the Brazilian northeast. *Sci. Hortic.* **2019**, *256*, 108521. [\[CrossRef\]](#)
- Rabab, S.; Badenhorst, P.; Chen, Y.-P.P.; Daetwyler, H.D. A template-free machine vision-based crop row detection algorithm. *Precis. Agric.* **2021**, *22*, 124–153. [\[CrossRef\]](#)
- Soleimanipour, A.; Chegini, G.R. A vision-based hybrid approach for identification of Anthurium flower cultivars. *Comput. Electron. Agric.* **2020**, *174*, 105460. [\[CrossRef\]](#)
- Chen, J.; Qiang, H.; Xu, G.; Liu, X.; Mo, R.; Huang, R. Extraction of navigation line based on improved grayscale factor in corn field. *Ciência Rural* **2020**, *50*, 5. [\[CrossRef\]](#)
- Choi, K.; Han, S.; Park, K.-H.; Kim, S. Vision based Guidance Line Extraction for Autonomous Weed Control Robot in Paddy Field. In Proceedings of the 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), Zhuhai, China, 6–9 December 2015.
- Li, B.; Yang, Y.; Qin, C.; Bai, X.; Wang, L. Improved random sampling consensus algorithm for vision navigation of intelligent harvester robot. *Ind. Robot. Int. J. Robot. Res. Appl.* **2020**, *47*, 881–887. [\[CrossRef\]](#)
- Tian, Y.; Yang, G.; Wang, Z.; Wang, H.; Li, E.; Liang, Z. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput. Electron. Agric.* **2019**, *157*, 417–426. [\[CrossRef\]](#)
- Bhargava, A.; Bansal, A. Automatic Detection and Grading of Multiple Fruits by Machine Learning. *Food Anal. Methods* **2020**, *13*, 751–761. [\[CrossRef\]](#)
- Liu, G.; Mao, S.; Kim, J. A Mature-Tomato Detection Algorithm Using Machine Learning and Color Analysis. *Sensors* **2019**, *19*, 2023. [\[CrossRef\]](#) [\[PubMed\]](#)
- Alam Siddiquee, K.N.E.; Islam, M.S.; Dowla, M.Y.U.; Rezaul, K.M.; Grout, V. Detection, quantification and classification of ripened tomatoes: A comparative analysis of image processing and machine learning. *IET Image Process.* **2020**, *14*, 2442–2456. [\[CrossRef\]](#)
- Khan, S.; Tufail, M.; Khan, M.; Ahmad, Z.; Anwar, S. Deep learning-based identification system of weeds and crops in strawberry and pea fields for a precision agriculture sprayer. *Precis. Agric.* **2021**, *22*, 1711–1727. [\[CrossRef\]](#)
- Wu, Z.; Chen, Y.; Zhao, B.; Kang, X.; Ding, Y. Review of Weed Detection Methods Based on Computer Vision. *Sensors* **2021**, *21*, 3647. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bah, M.; Hafiane, A.; Canals, R. CRowNet: Deep network for Crop row detection in UAV images. *IEEE Access* **2019**, *8*, 5189–5200. [\[CrossRef\]](#)
- Adhikari, S.P.; Kim, G.; Kim, H. Deep Neural Network-Based System for Autonomous Navigation in Paddy Field. *IEEE Access* **2020**, *8*, 71272–71278. [\[CrossRef\]](#)
- Mora-Fallas, A.; Goeau, H.; Joly, A.; Bonnet, P.; Mata-Montero, E. Instance segmentation for automated weeds and crops detection in farmlands. *Tecnol. Marcha* **2020**, *33*, 13–17. [\[CrossRef\]](#)

26. Menshchikov, A.; Shadrin, D.; Prutyaynov, V.; Lopatkin, D.; Sosnin, S.; Tsykunov, E.; Iakovlev, E.; Somov, A. Real-Time Detection of Hogweed: UAV Platform Empowered by Deep Learning. *IEEE Trans. Comput.* **2021**, *70*, 1175–1188. [[CrossRef](#)]
27. Afonso, M.; Fonteijn, H.; Fiorentin, F.S.; Lensink, D.; Mooij, M.; Faber, N.; Polder, G.; Wehrens, R. Tomato Fruit Detection and Counting in Greenhouses Using Deep Learning. *Front. Plant Sci.* **2020**, *11*, 571299. [[CrossRef](#)]
28. Alzadjali, A.; Alali, M.; Veeranampalayam Sivakumar, A.N.; Deogun, J.; Scott, S.; Schnable, J.; Shi, Y. Maize Tassel Detection From UAV Imagery Using Deep Learning. *Front. Robot. AI* **2021**, *8*, 600410. [[CrossRef](#)]
29. Osorio, K.; Puerto, A.; Pedraza, C.; Jamaica, D.; Rodriguez, L. A Deep Learning Approach for Weed Detection in Lettuce Crops Using Multispectral Images. *AgriEngineering* **2020**, *2*, 471–488. [[CrossRef](#)]
30. Aguiar, A.S.; Magalhaes, S.A.; dos Santos, F.N.; Castro, L.; Pinho, T.; Valente, J.; Martins, R.; Boaventura-Cunha, J. Grape Bunch Detection at Different Growth Stages Using Deep Learning Quantized Models. *Agronomy* **2021**, *11*, 1890. [[CrossRef](#)]
31. Liu, Y.; Sun, P.; Wergeles, N.; Shang, Y. A Survey and Performance Evaluation of Deep Learning Methods for Small Object Detection. *Expert Syst. Appl.* **2021**, *172*, 114602. [[CrossRef](#)]
32. Mazzia, V.; Salvetti, F.; Khaliq, A.; Chiaberge, M. Real-Time Apple Detection System Using Embedded Systems with Hardware Accelerators: An Edge AI Application. *IEEE Access* **2020**, *8*, 9102–9144. [[CrossRef](#)]
33. Zhang, X.; Kang, X.; Feng, N.; Liu, G. Automatic recognition of dairy cow mastitis from thermal images by a deep learning detector. *Comput. Electron. Agric.* **2020**, *178*, 105754. [[CrossRef](#)]
34. Li, Y.; Li, M.; Qi, J.; Zhou, D.; Zou, Z.; Liu, K. Detection of typical obstacles in orchards based on deep convolutional neural network. *Comput. Electron. Agric.* **2021**, *181*, 105932. [[CrossRef](#)]
35. Xu, W.; Zhao, L.; Li, J.; Shang, S.; Ding, X.; Wang, T. Detection and classification of tea buds based on deep learning. *Comput. Electron. Agric.* **2022**, *192*, 106547. [[CrossRef](#)]
36. Yang, Y.; Nie, J.; Kan, Z.; Yang, S.; Zhao, H.; Li, J. Cotton stubble detection based on wavelet decomposition and texture features. *Plant Methods* **2021**, *17*, 113. [[CrossRef](#)] [[PubMed](#)]
37. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
38. Lu, Z.; Zhao, M.; Luo, J.; Wang, G.; Wang, D. Design of a winter-jujube grading robot based on machine vision. *Comput. Electron. Agric.* **2021**, *186*, 106170. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.