agronomy

MDPI

Article

YOLOv5-AC: A Method of Uncrewed Rice Transplanter Working Quality Detection

Yue Wang [1], Qiang Fu [2], Zheng Ma [1], Xin Tian [1], Zeguang Ji [1], Wangshu Yuan [1], Qingming Kong [1], Rui Gao [1,*] and Zhongbin Su [1,*]

[1] Institutions of Electrical and Information, Northeast Agricultural University, Harbin 150038, China
[2] School of Conservancy and Civil Engineering, Northeast Agricultural University, Harbin 150038, China; fuqiang@neau.edu.cn
* Correspondence: rgao@neau.edu.cn (R.G.); suzb001@163.com (Z.S.)

Abstract: With the development and progress of uncrewed farming technology, uncrewed rice transplanters have gradually become an indispensable part of modern agricultural production; however, in the actual production, the working quality of uncrewed rice transplanters have not been effectively detected. In order to solve this problem, a detection method of uncrewed transplanter omission is proposed in this paper. In this study, the RGB images collected in the field were inputted into a convolutional neural network, and the bounding box center of the network output was used as the approximate coordinates of the rice seedlings, and the horizontal and vertical crop rows were fitted by the least square method, so as to detect the phenomenon of rice omission. By adding atrous spatial pyramid pooling and a convolutional block attention module to YOLOv5, the problem of image distortion caused by scaling and cropping is effectively solved, and the recognition accuracy is improved. The accuracy of this method is 95.8%, which is 5.6% higher than that of other methods, and the F1-score is 93.39%, which is 4.66% higher than that of the original YOLOv5. Moreover, the network structure is simple and easy to train, with the average training time being 0.284 h, which can meet the requirements of detection accuracy and speed in actual production. This study provides an effective theoretical basis for the construction of an uncrewed agricultural machinery system.

Keywords: uncrewed rice transplanter; YOLO; least square method; smart agriculture

Citation: Wang, Y.; Fu, Q.; Ma, Z.; Tian, X.; Ji, Z.; Yuan, W.; Kong, Q.; Gao, R.; Su, Z. YOLOv5-AC: A Method of Uncrewed Rice Transplanter Working Quality Detection. *Agronomy* **2023**, *13*, 2279. https://doi.org/10.3390/agronomy13092279

Academic Editor: Juncheng Ma

Received: 13 July 2023
Revised: 20 August 2023
Accepted: 23 August 2023
Published: 29 August 2023

1. Introduction

With the continuous development and advancement of smart agriculture and uncrewed farming technology, uncrewed rice transplanters have become essential in modern agricultural production. Currently, the primary methods of hybrid rice production in agriculture include mechanized direct seeding and rice transplanting [1,2], and the farm machinery is manually driven. The advantage of this method is that the driver can control the speed and accuracy of the operation and monitor the working quality of the machine during operation.

However, in the actual production, the complexity of the working environment and the machine can lead to mechanical failure and deviation, resulting in rice transplanting omission, seedling flotation, and other problems. Actual planting efficiency is about 93.03% [3], and the operation process is affected by many factors, such as the mechanized seeding performance, the number of rice seedlings caught by the seedling picker, the quality of the rice seedling blanket tray, and other factors. Currently, only the machine execution progress on the established route can be detected. The working quality of the uncrewed rice transplanters is not effectively detected and visualized, making monitoring their working quality problematic. Thus, effectively detecting the work quality of uncrewed rice transplanters has become an urgent problem.

Omission detection for uncrewed rice transplanters must determine the intersection of horizontal and vertical crop rows based on crop row detection. Therefore, crop row

detection is integral to rice transplanter omission detection. The research on crop rows is usually divided into two steps: identifying the crop and then fitting the crop rows.

Crop row recognition can be divided into deep learning-based, stereo vision-based, and crop feature-based approaches. The crop feature-based approach is mainly based on the spectral, geometric and color characteristics of the crop and uses technical means, such as image segmentation or classification to separate the crop from the farmland background. Some researchers used the U component of the YUV color space as the optimal component of grayscale images to remove the effect of light intensity on the data, combined with the Fourier transform for crop-row contour extraction, and used the least squares method (LSM) for linear fitting [4]. Other researchers conducted multispectral analysis experiments on wheat leaves, weeds, and soil and obtained different spectral features that can distinguish these three by analyzing different spectral wavelengths [5].

In a complex rice field environment, reflections from the water surface can significantly interfere with the spectral features, making it challenging to distinguish the farmland background effectively from the crop by the characteristic wavelengths. Thus, some researchers used the YCrCb color model to construct a Cg component independent of illumination and used a feature factor to grayscale the image to reduce the effect of illumination on data features and improve the universality of the algorithm [6].

The standard methods for linear fitting include the Hough transform and the LSM. Although the Hough transform can perform complex curve fitting, it has problems such as high complexity and poor real-time performance, so the LSM is more common for simple straight-line fitting tasks. Thus, some researchers estimated crop centroids based on multiple regions of interest and used the LSM for line fitting after removing pseudo points using the classification algorithm [7].

Deep learning has developed rapidly in recent years, and convolutional neural networks (CNNs) with good feature extraction and classification capabilities have been widely used in image recognition, speech recognition, natural language processing, and other fields. Deep learning has also been gradually applied in agriculture, such as obtaining phenotypic traits of crops, including estimating biomass [8]; detecting growth process [9], pests, diseases, and weeds [10]; and obtaining the quantitative traits of crops, including the number of flowers [11], ears [12], and fruits [13,14]. A RCNN-based network model was used in spatial research to detect rice seedlings, using the detection frame centroids as seedling feature points [15,16]. Moreover, one study in 2020 proposed a crop row detection algorithm based on YOLOv3. After identifying and locating rice seedlings using the network model, the smallest univalue segment assimilating nucleus (SUSAN) corner points of seedlings in the detection frame were extracted, and the crop rows were obtained by fitting the SUSAN corner points using the LSM [17].

Due to the limitations of these network models, there are disadvantages in detection accuracy and processing speed. Therefore, uncrewed rice transplanter omission detection with a processing speed based on higher detection accuracy is urgently needed.

This paper assesses the detection of rice seedling omission based on the target detection model YOLOv5. First, we employ neural networks for the preliminary recognition of image information acquired during uncrewed machine operations. Second, the network model is modified to improve the detection accuracy of the model further and balance the problem of model detection accuracy and operation efficiency. Finally, the algorithm is used to fit the crop row and determine the omission locations.

The specific objectives of this study are as follows. First, we propose an uncrewed rice transplanter omission detection method. Second, we design a CNN (YOLOv5-AC, YOLOv5 combines atrous spatial pyramid pooling and a convolutional block attention module) to improve rice seedling recognition. The proposed method is fully validated on self-built datasets, which is vital to improving the efficiency of uncrewed farm machinery operations and building systems.

2. Materials and Methods

2.1. Experimental Site Description and Image Acquisition

The experimental site of this paper is the Agricultural Extension Center of the Northeast Agricultural University in Harbin, Heilongjiang Province, China. It is in the southwest of Heilongjiang Province (about 45°30′ 59.95″ N to 45°31′ 21.75″ N and 127°01′ 34.28″ to 127°01′ 58.39″ E), with an altitude of about 112 m. The center covers an area of 83.4 ha, of which 50.5 ha is cultivated. The region has a cold temperate continental monsoon climate with prominent seasonal characteristics. The annual average precipitation is 591 to 783 mm. The soil is fertile and low-lying, and the soil organic matter content is 3% to 6%, with high natural fertility. Generally, crops are harvested once a year. Paddy fields are the main cultivated land in the experimental station, and rice is the main crop. Figure 1 illustrates the location of the research area.

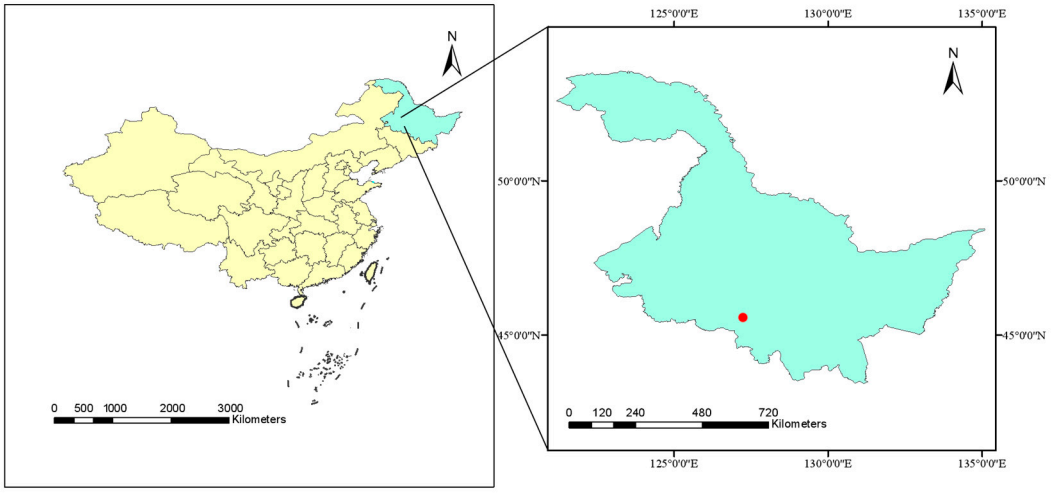

Figure 1. Experimental site diagram (The red dot indicates the location of the experiment area).

The image acquisition system is a rice transplanter (2ZG-6DK riding-type rice transplanter, Nantong FLW Agricultural Equipment Co., Ltd., Jiangsu, China) equipped with a color camera (MV-CS200-10GC, Hangzhou Hikvision Digital Technology Co., Ltd., Hangzhou, China) and image data acquisition device. The transplanter has six seeders in three groups.

The rice variety used in the experiment was Dongfu 112 (DF112), which was sown and planted in mid-April 2022 and transplanted in late May. The fields were planted according to the traditional local rice cultivation pattern. Field trial plots and large areas were managed the same way as by local farmers. In addition, 528 images were collected from 10:30 a.m., and 200 were selected to participate in the experiment. Through the test, considering the image resolution, shooting height, uncrewed transplanter shaking, water surface ripples, and other factors, the red, green, and blue (RGB) image of a rice seedling with a resolution of 3840 × 2160, 1.7 m away from the water surface was selected, which completely covers six rows of rice seedlings during the transplanter operation. All images in the experiment were saved in JPG format. The transplanter worked for 2 h, totaling 1.06 ha, to transplant about 3710 rice seedlings, and the leakage rate was about 8.2%.

In this study, machine-implanted rice seedling images taken in the field environment and artificially enhanced images were used as training sets to train the neural network and improve the accuracy and robustness of the network detection. Figure 2 depicts some representative images. The individual rice seedlings are clear, the distribution of the seedling rows is more uniform, and there is an apparent omission phenomenon. Based on the level of the paddy field, most of the rice seedlings were rooted in deep soil, and a few were immersed in water, but all the rice seedlings were still clearly visible.

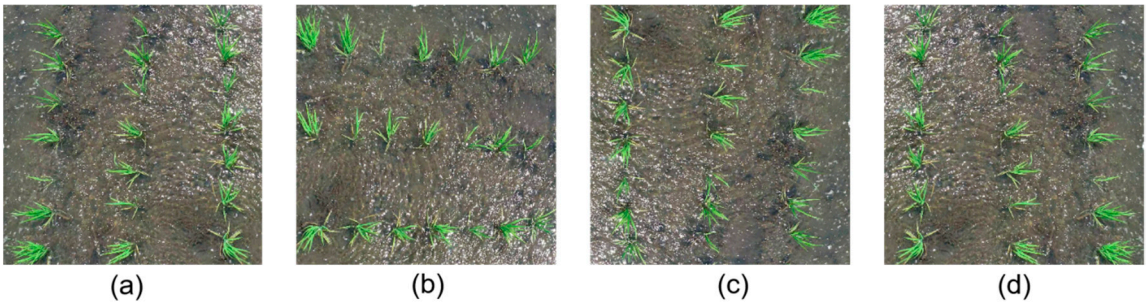

Figure 2. Representative images: (**a**) original, (**b**) 90° rotation, (**c**) 180° rotation, and (**d**) flipped.

The artificial enhancement methods of rice seedling images include the flipped, 90°, and 180° rotations. This method expands the dataset, improves the sample size, and improves the robustness of the network model. The rice seedling datasets were annotated using the open-source tool "LabelImg" and were saved in VOC format. Datasets usually include training, verification, and testing sets. In this paper, the ratio of 9:1 is used to divide the set into training and testing sets and training and verification sets.

2.2. Detection of Rice Seedlings

2.2.1. YOLOv5

The "you only look once" (YOLO) series algorithms are one-stage series algorithms [18]. After continuous improvement and optimization, the YOLO series has achieved the best detection accuracy and processing speed among one-stage algorithms. Compared with the previous generation network model, YOLOv5 has a higher detection accuracy and smaller model size, but such improvement is not achieved at the cost of the number of parameters.

In addition, YOLOv5 is a single-stage target detection algorithm. The model can be divided into four parts: the input, backbone, neck, and head. The main work of the input stage is to preprocess the image and scale the image data prepared by the user to the size of the network entrance. The primary work of the input side of YOLOv5 includes mosaic data enhancement, adaptive anchor frame calculation, adaptive picture scaling, and other work. The size of the network entrance is 640 × 640, and image preprocessing is usually performed before input. That is, the input image is scaled to the input size of the network, and normalization and other operations are conducted.

The backbone network of YOLOv5 comprises the CBS (made up of convolution, batch normalization, and SiLU unit) and CSP (cross stage partial) modules. The backbone extracts feature information from images and combines it to form feature maps of different granularities. The CBS module has three normalization functions: two-dimensional convolution (Conv2D), batch normalization, and activation. The activation function uses a weighted linear unit (sigmoid linear unit) to activate the next layer, and the CSP module contains three standard convolutional layers and several bottleneck modules.

The neck part is responsible for combining feature images and extracting features, improving the robustness of the detection network, and the head outputs the target detection results. The number of branches of the output end differs for different detection algorithms, usually including classification and regression branches. Figure 3 presents the YOLOv5 network structure, where BN stands for batch normalization, Conv stands for convolution, and SPPF stands for spatial pyramid pooling fast.

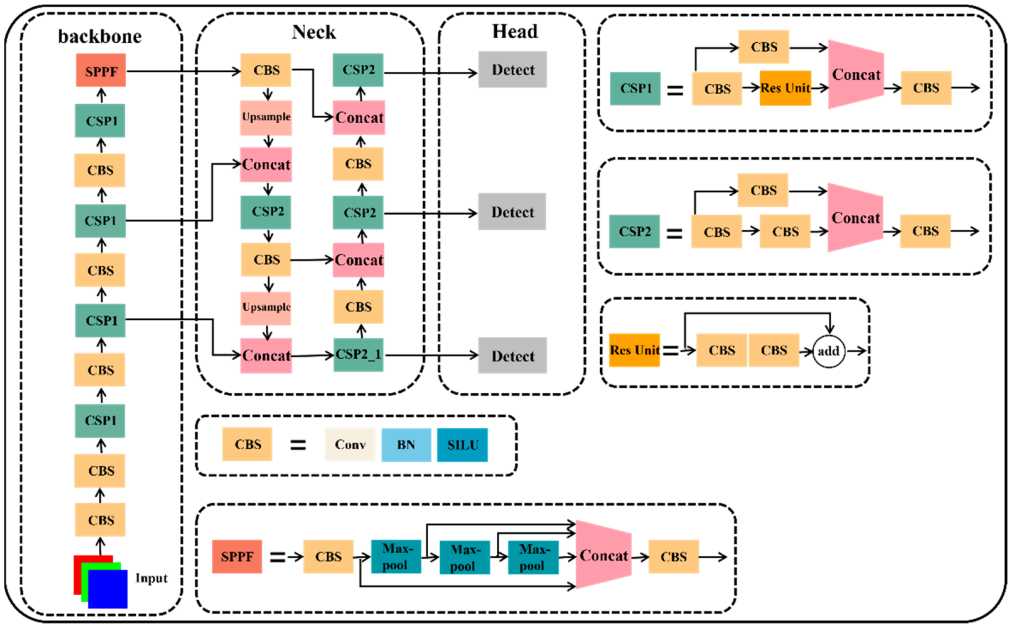

Figure 3. Schematic diagram of the YOLOv5 structure.

2.2.2. Atrous Spatial Pyramid Pooling

The atrous spatial pyramid pooling (ASPP) module is an integrated spatial pyramid structure [19]. This module uses multiple parallel void convolutional layers with different sampling rates. The features extracted for each sampling rate are further processed in a separate branch and fused to generate the result. This module constructs a convolutional kernel with different receptive fields through various voids to obtain multiscale object information. As depicted in Figure 4, the ASPP module consists of a series of cavity convolutions and global average pooling with different expansion rates to extract multiscale object features and ensure a high image resolution.

The ASPP module solves two problems: First, it effectively avoids image distortion caused by clipping and zooming of the image area. Second, it extracts graph-related repetitive features using the CNN, significantly improving the speed of generating candidate frames and reducing the calculation cost.

However, the module also has the following shortcomings. First, an improper setting of the parallel cavity convolutional expansion rate is likely to result in a "gridding effect." Second, the function area of the module is square, which is not conducive to capturing the context information of strips and thin objects. Moreover, when numerous input feature channels exist, the number of module parameters is large, and the memory usage is high.

2.2.3. Convolutional Block Attention Module

The convolutional block attention module (CBAM) combines space and channels, including the channel and spatial attention modules [20]. Figure 5 illustrates the structure of the attention mechanism in the CBAM. The feature map passes through the channel attention mechanism, then goes through the adaptive average pooling and adaptive maximum pooling to obtain two 1×1 channel weight matrices. The single-channel feature map is multiplied by the input feature to generate an intermediate feature map. After the spatial attention mechanism, two 2D vectors are spliced through maximum and average pooling. After the convolutional layer and sigmoid activation function, the output feature map is multiplied by the intermediate feature map to obtain the output feature map.

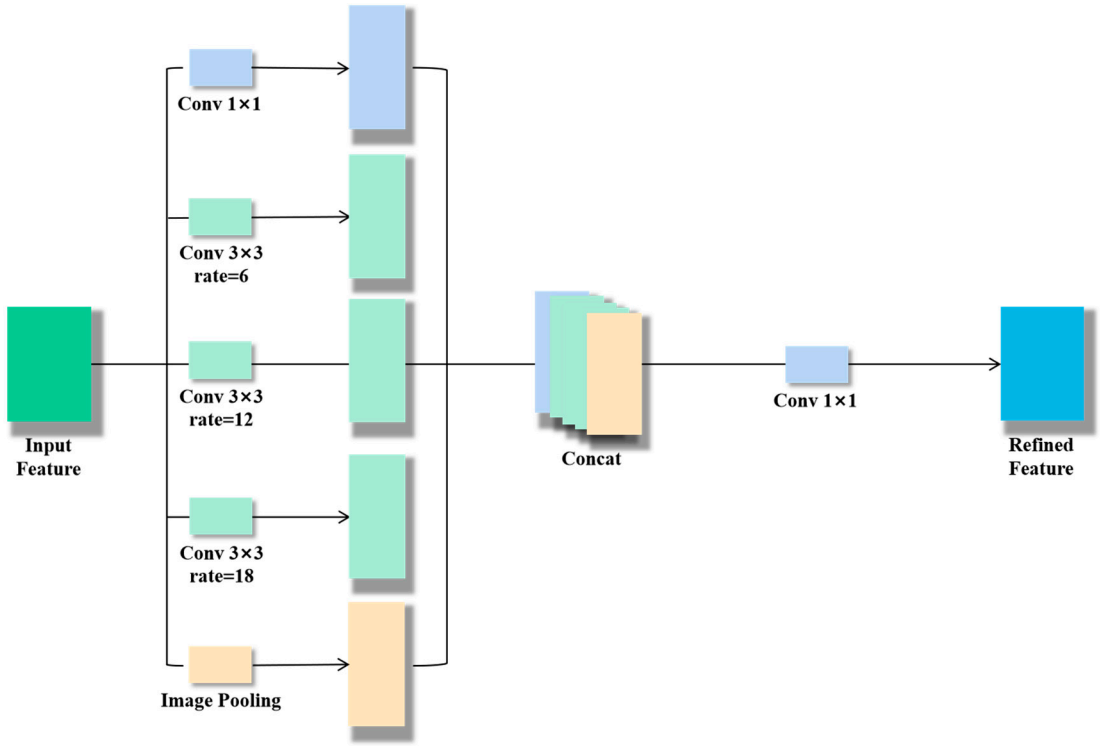

Figure 4. Atrous spatial pyramid pooling module structure diagram.

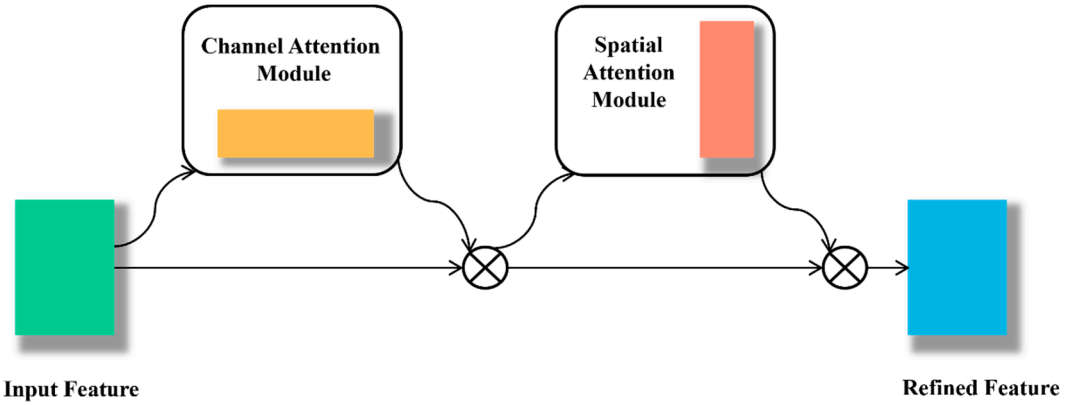

Figure 5. Convolutional block attention module structure diagram.

2.2.4. Network Architecture

This paper aims to detect the working quality of uncrewed farm machinery. As the YOLO series algorithms have made specific achievements in detection accuracy and processing speed, this paper proposes a working quality detection network based on YOLOv5, ensuring the network processing speed to an extent and improving the detection accuracy of targets. Considering the timeliness and equipment integration factors, the network performance should be higher at a lower cost.

ASPP performs parallel sampling of a given input using hollow convolution at different sampling rates, equivalent to capturing the context of the image at multiple scales. Atrous convolution can also increase the receptive field of the filter without increasing the computation. Thus, the model can improve the perception of various scale segmentation targets. The expansion coefficient of each layer of the pool pyramid can be customized to achieve free multiscale feature extraction. In this paper, the ASPP module is added to the CBS layer to enable the model to conduct multiscale extraction and fusion of rice

seedling image features to improve the detection of rice seedlings. We applied and tested the detection performance of the ASPP module on a rice seedling of YOLOv5 in five positions. The ASPP modules were placed behind the CBS module of the backbone, called Y5A1–Y5A5, in order. The application positions with the best performance are selected through comprehensive comparison. Figure 6 depicts the locations of the ASPP module tested.

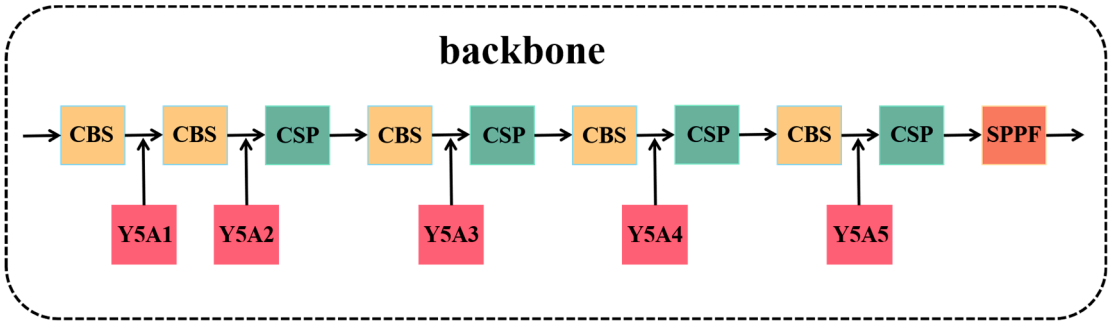

Figure 6. Diagram of the ASPP module test location.

The CBAM is a simple and effective attention module in feedforward CNNs. Given an intermediate feature map, the CBAM infers the attention map once along two independent dimensions (space and channel) and then multiplies the attention map with the input feature map for adaptive feature optimization. Because CBAM is a lightweight, general-purpose module, it can be seamlessly integrated into the CNN architecture without the overhead of the module and trained end-to-end with the base CNN.

This paper employs two integration methods of CBAMs to verify their effectiveness. In the first method, the CBAM is placed at the end of the backbone part to obtain the features of the entire backbone and obtain the global vision (Y5CB is substituted in this paper). In the second method, the CBAM is placed behind each CSP module of the backbone so that the attention mechanism can obtain local features of this part (named Y5C1–Y5C4 in order). Figure 7 presents the test diagram of the CBAM.

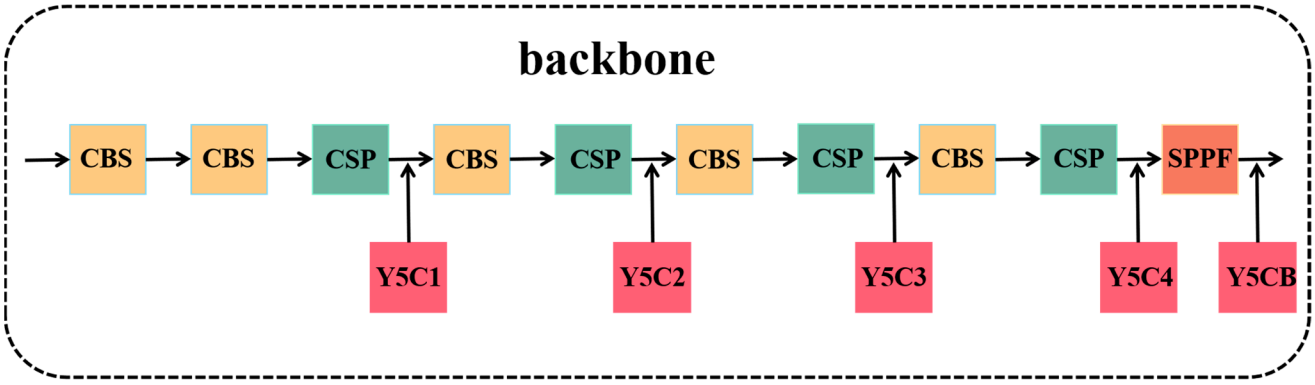

Figure 7. Diagram of the CBAM test location.

Through the above experiments, we obtain the best performing module integration position, and further obtain the structure of the whole convolutional neural network. Figure 8 depicts the network architecture.

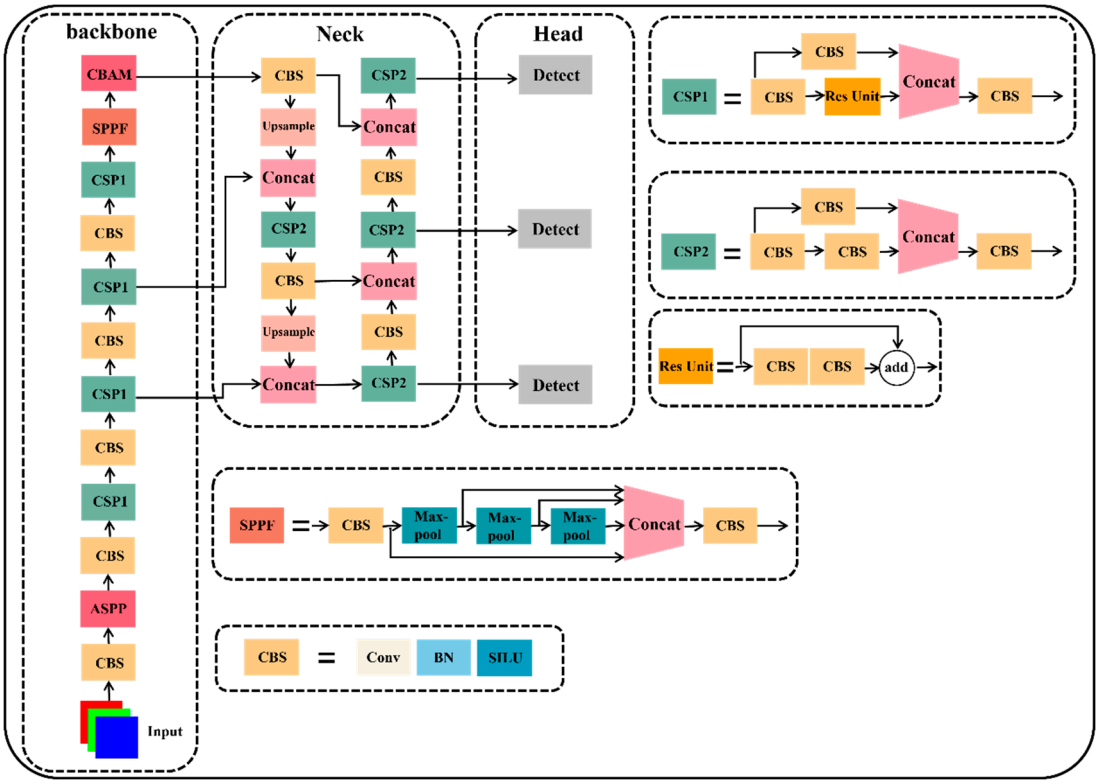

Figure 8. Diagram of the YOLOv5-AC network structure.

2.3. Crop Row Fitting and Working Quality Detection

In the operation of automatic farm machinery, the machinery moves forward at a uniform speed according to the planned route. The format of the rice seedling network is distributed in the operation area. Due to the angle of view of the camera, it presents a trapezoidal grid. The rice seedling coordinates recognized in the image must be fit to the seedling belt coordinates to detect the working quality of the agricultural machinery.

2.3.1. Least Squares Method

This paper uses the LSM for line fitting. After the detection of machine-inserted rice seedlings using the CNN, a bounding box can be obtained, and the central coordinate of the boundary frame is used as the approximate coordinate value of the corresponding rice seedling to fit the straight line of the seedling row. The LSM (ordinary least squares) is a mathematical optimization technique commonly used to solve curve-fitting problems [21]. It searches for the best function match of the data by minimizing the sum of the squares of the errors. The LSM can be used to obtain unknown data easily and minimize the sum of squares of errors between the obtained and actual data:

$$f(x) = a_1\varphi_1(x) + a_2\varphi_2(x) + \cdots + a_m\varphi_m(x),$$

where $\varphi_k(x)$ denotes a group of linearly independent functions selected in advance, and a_k represents the undetermined coefficient. The fitting criterion minimizes the sum of squares of the distance between $y_i(i = 1, 2, \cdots, n)$ and $f(x_i)$, called the least squares criterion.

2.3.2. Working Quality Detection

After target detection and crop row fitting, the following data can be obtained: target bounding box, crop row coordinates, and transverse and longitudinal crop row intersection coordinates. We take the interleaved point of the seedling row as the predicted position of the seedling. If the predicted point falls into any bounding box, it is judged as passing

the working quality detection. If the predicted point does not fall into any boundary box, it is determined that the missing point phenomenon occurs at this point. The missing point position can be accurately calculated and recorded using various parameters (e.g., the geographical coordinates, flight altitude, and angle of view). Figure 9 depicts the position of the missing insertion.

Figure 9. Schematic diagram of the missing insertion positions (The red squares marks the omission positions).

2.4. Experimental Procedure

First, the acquired images of the machine-inserted rice seedlings were manually annotated to obtain training label files. The training and testing sets were divided into a ratio of 9:1, and the training and verification sets were also divided into a ratio of 9:1. Second, the training set with 300 epochs was inputted into the YOLOv5 network model with different improved modes for training. The optimal weights of the network models with different structures were obtained by training. Finally, the testing set was used to test the performance of these network models, and the results were compared with those of the original YOLOv5 network. Based on various evaluation indicators, the network model with the best effect was selected as the detection network model of machine-inserted rice seedlings in the paddy field environment. Figure 10 presents the experimental process.

2.5. Evaluation Indicator and Experimental Equipment

2.5.1. Evaluation Indicator

In deep learning, performance indicators are critical to measuring model performance by measuring the gap between the model output and actual value. Performance indicators, such as the accuracy and recall rate, can intuitively compare the advantages and disadvantages of models through data or charts, often the final goal of model training. Performance indicators also serve as the basis for the verification set to make decisions and are critical standards for evaluating output quality.

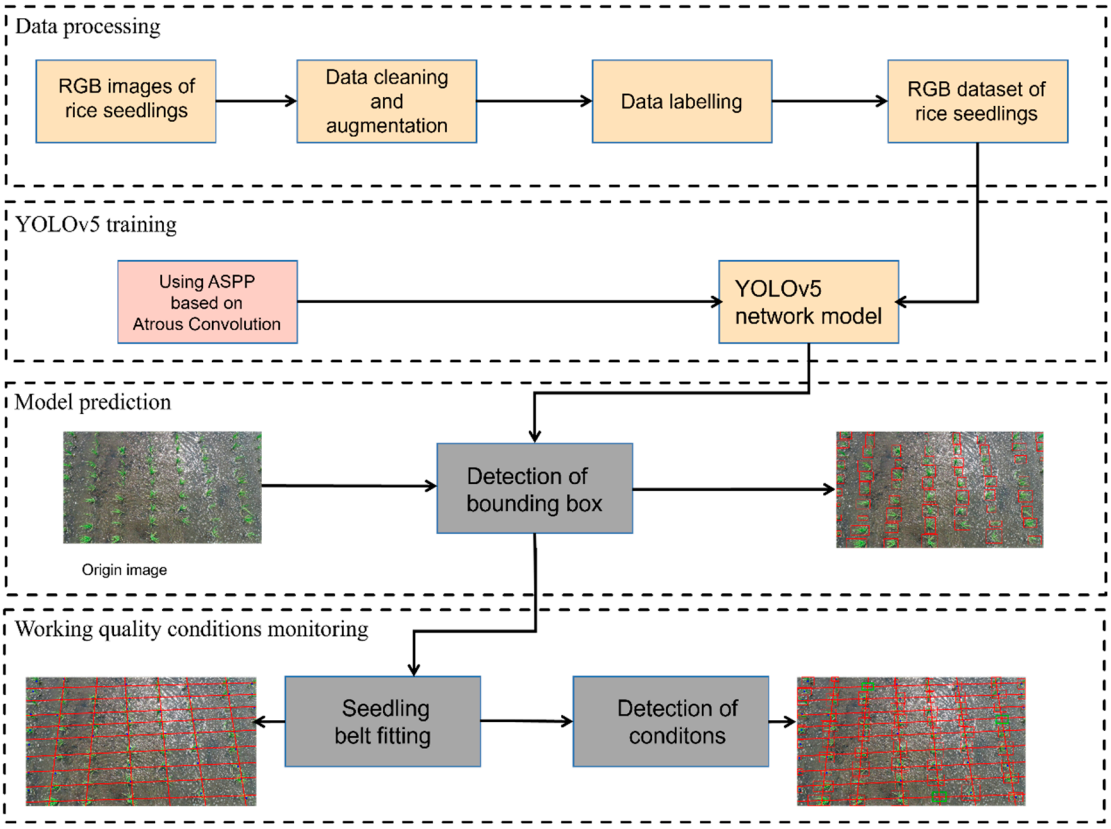

Figure 10. Diagram of the experimental procedure.

Performance indicators are usually divided into regression or classified performance indicators. In this paper, precision (P), recall (R), harmonic average (F_1), and mean average precision (mAP) are selected as evaluation indicators, calculated as follows:

$$P = \frac{TP}{TP+FP} \times 100\%$$

$$R = \frac{TP}{TP+FN} \times 100\%$$

$$AP = \frac{\sum precision}{N} \times 100\%$$

$$mAP = \frac{\sum AP}{N_t} \times 100\%$$

$$F_1 = \frac{2 \times precision \times recall}{precision + recall} \times 100\%$$

where TP represents the number of correctly identified rice seedlings, FP denotes the number of incorrectly classified rice seedlings, FN indicates the number of undetected rice seedlings, N represents the total number, N_t represents the number of detected target categories, AP (average precision) indicates the average p-values on the precision-recall (P-R) curve, and mAP denotes the average of the average accuracy of all categories in the dataset. Only rice seedlings are tested in this paper; thus, mAP and AP are equal.

In the linear fitting process, evaluation indicators are also needed to measure the degree of fitting. Compared with the images, the evaluation indicators can more intuitively show the advantages and disadvantages of the fitting results, and the evaluation index is also an important basis to choose the final results. Therefore, this paper introduces the residual sum of squares (RSS), mean square error (MSE) and mean absolute error (MAE) as

the evaluation criteria, where $\hat{y}_i$ represents the true value of the sample, and y_i represents the estimated value of the sample, calculated as follows:

$$RSS = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|$$

2.5.2. Experimental Equipment

This experiment used a desktop computer as the processing platform, equipped with an Ubuntu 20.04.3 operating system (Canonical Co., Ltd., Mann, British). The algorithm uses PyCharm (JetBrains S.R.O., Prague, Czech Republic) in Python 3.9.7. The PyTorch framework (Meta Platform Inc., Menlo Park, CA, USA) was built and configured in Anaconda 3 (Anaconda Inc., Austin, TX, USA), Compute Unified Device Architecture (CUDA v. 11.4). Table 1 lists the detailed hardware and software configurations.

Table 1. Detailed hardware and software configurations used in the experiment.

Configuration	Parameter
CPU	Intel i9-12900k
RAM	64 G
GPU	Nvidia RTX3090
Operation system	Ubuntu 20.04.3
Deep learning framework	PyTorch 1.11.0
Programming language	Python 3.9.7
GPU computing platform	CUDA 11.4

3. Results

This section presents the effectiveness of the improved model, comparing the performance with the mainstream target detection models, such as YOLOv5, and the evaluation indicators, including accuracy, convergence speed, and training time. The performance of the proposed rice omission detection method is verified in this paper.

3.1. Comparison of YOLOv5-AC with Other NETWORK Models

We tested the detection performance of three network models using the same datasets. Table 2 presents the test results.

Table 2. Performance comparison of networks trained on the same dataset.

Model	Precision (%)	Recall (%)	mAP@.50 (%)	F1-Score (%)	Time (h)
YOLOv5	90.2	87.3	91.5	88.73	0.212
YOLOv5-Ghost	88.2	86.3	88.7	87.24	0.192
YOLOv7	96.5	92.3	93.7	93.42	0.415
Faster-RCNN	92.8	91.9	92.6	92.34	0.426
YOLOv5-AC	95.8	91.1	94.1	93.39	0.284

The accuracy of YOLOv5-AC is 95.8%, significantly higher than the accuracy of 90.2% for YOLOv5 and 88.2% for YOLOv5-Ghost. The recall rate of YOLOv5-AC is 91.1%, which is also higher than 87.3% for the YOLOv5 model and 86.3% for the YOLOv5-Ghost model. The mAP@.5 value of YOLOv5-AC is 94.1%, whereas the values of the YOLOv5 and YOLOv5-Ghost models are 91.5% and 88.7%, respectively. Compared with the original YOLOv5,

mAP@.5 increased by 2.6%, the F1-score increased by 4.66%, and the recall rate increased by 3.8%. The statistical result reveals that the improved model can be applied to identifying rice seedlings and detecting transplanting omissions. Although a slight disadvantage exists in the training time of the model, it can achieve higher detection accuracy and provide strong support for detecting rice seedling omissions.

3.2. Effect of ASPP Module Position on the Model Effect

In this experiment, the ASPP module is integrated into various parts of YOLOv5, and we explored the influence of the ASPP module on the performance of this network. We evaluated the ASPP module with a channel number of 256 at various locations, using the datasets of rice seedlings collected in a real field environment, and the performance of the network model with the ASPP module at different locations is provided in Figure 11 and Table 3.

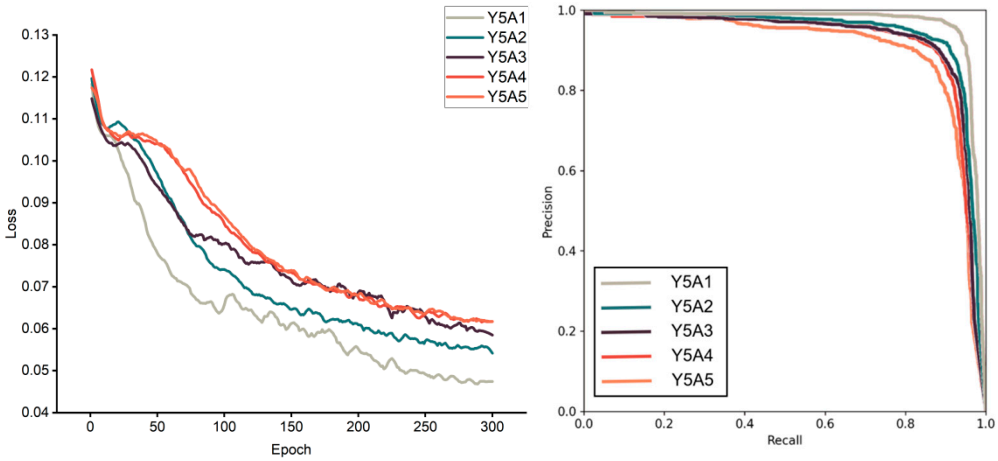

Figure 11. Loss curves and precision-recall plots of the ASPP module integration locations trained on the same datasets.

Table 3. Performance comparison of ASPP module integration locations trained on the same datasets.

Model	Precision (%)	Recall (%)	mAP@.50 (%)	F1-Score (%)	Time (h)
Y5A1	93.7	91.5	95.8	92.59	1.392
Y5A2	91.8	90.3	94.1	91.04	0.435
Y5A3	90.2	88.3	92.5	89.24	0.364
Y5A4	90.3	87.6	92	88.93	0.215
Y5A5	87.3	86.4	90.3	86.85	0.210

As illustrated in Figure 10, when the ASPP module is in position Y5A1, the convergence speed is better than the other four positions, indicating that the module in that position results in a better feature extraction capability. The model performance in Table 3 indicates that when the ASPP module is integrated into the Y5A1 position, the accuracy is 93.7%, and mAP@.5 is 95.8%. Furthermore, the accuracy of Y5A2 is 91.8%, and mAP@.5 is 90.3%. The accuracy of Y5A3 is 90.2%, and mAP@.5 is 88.3%. The accuracy of Y5A4 is 90.3%, and mAP@.5 is 87.6%. Finally, the accuracy of Y5A5 is 87.3%, and mAP@.5 is 86.4%. The indicators of integrating the ASPP module in the Y5A1 position are significantly higher than those in the Y5A2, Y5A3, Y5A4, and Y5A5 positions. However, the training time is 1.392 h for Y5A1, 0.435 h for Y5A2, 0.364 h for Y5A3, 0.215 h for Y5A4, and 0.210 h for Y5A5. Although the experiment at each position is slightly extended compared with the training speed of YOLOv5, the training time for Y5A1 is significantly extended. Therefore, this paper aims to reduce the number of channels of the ASPP module in position Y5A1 to obtain faster processing speed.

3.3. Influence of the Number of ASPP Module Channels on the Model Effect

This section explores the effect of various channel numbers on the improved model with the ASPP module to obtain a higher model training speed. Three channel numbers were selected in this experiment: 64, 128, and 256. Figure 12 presents the loss function image, and Table 4 lists the evaluation indicators.

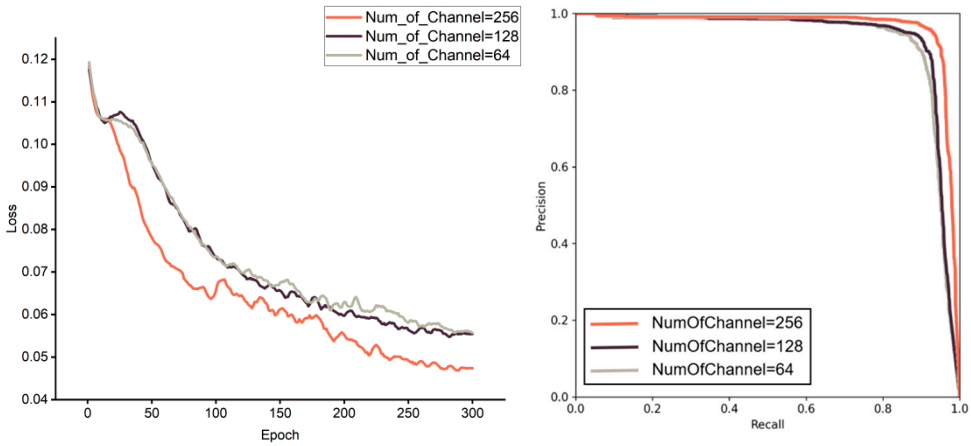

Figure 12. Loss curve and precision-recall diagram of Y5A1 with various channel numbers trained on the same datasets.

Table 4. Performance comparison of Y5A1 with different channel numbers trained on the same datasets.

Channel	Precision (%)	Recall (%)	mAP@.50 (%)	F1-Score (%)	Time (h)
256	93.7	91.5	95.8	92.59	1.392
128	93.6	89.9	93.9	91.71	0.434
64	94.2	87	93.2	90.46	0.283

Table 4 reveals that when the number of channels is 64, the accuracy is 94.2%, the recall rate is 87%, mAP@.5 is 93.2, and the F1-score is 90.46%. When the number of channels is 128, the accuracy is 93.6%, the recall rate is 89.9%, the mAP@.5 is 93.9%, and the F1-score is 91.71%. With 256 channels, the accuracy is 93.7%, the recall rate is 91.5%, mAP@.5 is 95.8%, and the F1-score is 92.59%. Figure 10 indicates that, as the number of channels decreases, the decline in the loss function also slows, indicating that the decrease in the number of channels of ASPP modules slightly weakens the performance of the network and the capability of feature extraction. However, when the number of channels is 64, the training time is 0.283 h. When the number of channels is 128, the training time is 0.434 h. When the number of channels is 256, the training time is 1.392 h. This result indicates that a smaller number of channels significantly improves the training speed of the model. When the number of channels is 64, the training speed of the model is close to that of the network model without the ASPP module.

3.4. Influence of the CBAM on the Model

At this stage, the experiment aims to improve the performance of network models by adjusting the position of CBAMs. The Y5C1–Y5C4 are integrated into the C3 modules of the backbone part. This operation aims to obtain local features of the backbone network and share the pressure of model training. The Y5CB is integrated at the end of the entire backbone to obtain its feature map and provide the global view.

By comparing the model performance in Table 5, when the CBAM is integrated into Y5C1 in the YOLOv5-AC network model, the accuracy is 91.2%, the recall rate is 88.5%, and mAP@.5 is 94.6%. When the CBAM is integrated into Y5C2, the accuracy is 92%, the recall

rate is 90.1%, and mAP@.5 is 94.2%. When the CBAM is integrated into Y5C3, the accuracy rate is 90.5%, the recall rate is 89.3%, and mAP@.5 is 93.5%. When the CBAM is integrated into Y5C4, the accuracy rate is 91.9%, the recall rate is 88.1%, and the mAP@.5 is 92.6%. When the CBAM is integrated into Y5CB, the accuracy rate is 95.8%, the recall rate is 91.1%, and mAP@.5 is 94.1%. These results indicate that enabling the CBAM to obtain the global view of the backbone can improve the overall model performance and provide a guarantee for the experimental effect. Figure 13 presents some examples of rice seedling detection.

Table 5. Network performance of CBAMs integrated into locations trained on the same datasets.

Model	Precision (%)	Recall (%)	mAP@.50 (%)	F1-Score (%)	Time (h)
Y5C1	91.2	88.5	94.6	89.8	0.284
Y5C2	92	90.1	94.2	91	0.284
Y5C3	90.5	89.3	93.5	89.9	0.28
Y5C4	91.9	88.1	92.6	90	0.284
Y5CB	95.8	91.1	94.1	93.39	0.284

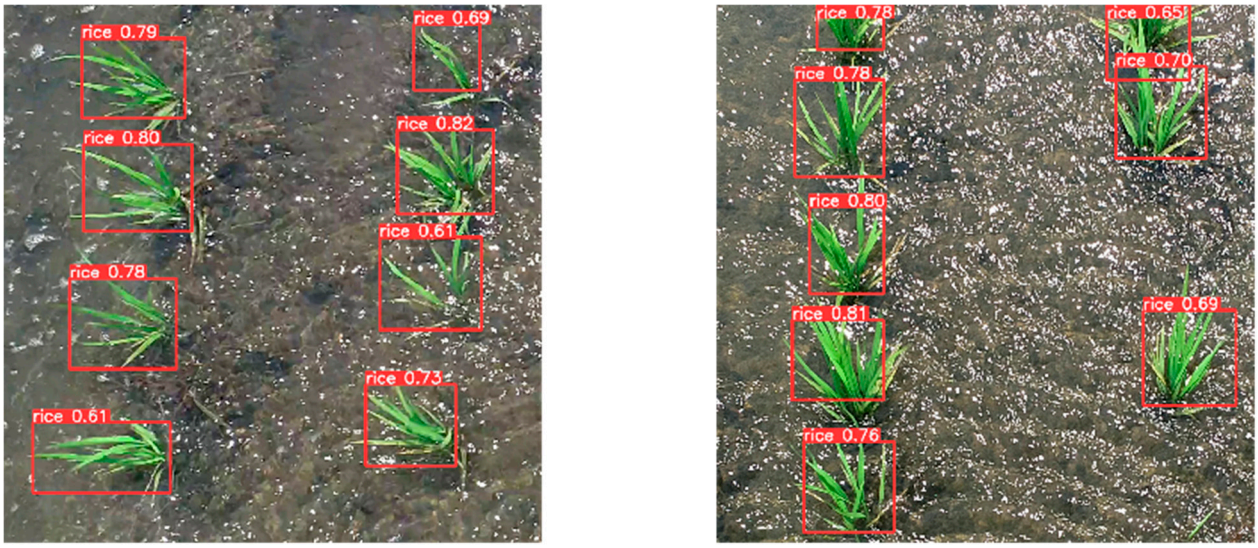

Figure 13. Examples of rice seedling detection.

3.5. Results of Crop Row Fitting and Omission Detection

Different methods using the coordinates in the actual experiment were used to fit the crop row equation, and RSS, MSE, and MAE were calculated. These methods include LSM, Hough transform, total least square (TLS), and random sample consensus (RANSAC). This paper takes one rice row as an example, and the statistical results are provided in Table 6.

Table 6. Results of the same rice row equation for different methods.

Method	Equation	RSS	MSE	MAE
LSM	$y = 0.68x - 1070.08$	259,287.77	37,041.11	139.86
Hough transform	$y = 0.51x - 616.66$	285,191.59	40,741.66	155.22
TLS	$y = 0.79x + 32.57$	294,209.68	39,251.39	161.27
RANSAC	$y = 0.79x + 32.53$	287,188.17	38,242.61	144.14

The LSM has certain advantages in each evaluation indicator. Table 6 reveals that the linear equations fitted using the LSM, Hough transform, TLS, and RANSAC are $y = 0.68x - 1070.08$, $y = 0.51x - 616.66$, $y = 0.79x + 32.57$, and $y = 0.79x + 32.53$, respectively. The RSS of the LSM is 259,287.77, the MSE is 37,041.11, and the MAE is 139.86. The RSS of the

Hough transform is 285,191.59, the MSE is 40,741.66, and the MAE is 155.22. The RSS of the total LSM is 294,209.68, the MSE is 39,251.39, and the MAE is 161.27. The RSS, MSE, and MAE of the random sampling consistency method are 287,188.17, 38,242.61, and 144.14, respectively.

The Hough transformation is limited by the input sequence of the coordinate points and the number of iterations, so the equation of the line is slightly different, but the evaluation indices are all within the general range. In the fitting process, RANSAC randomly selected data points, so the model obtained each time may differ, creating a certain degree of uncertainty in the algorithm. In simple linear fitting tasks, the LSM has advantages, with low algorithm complexity and a short execution time.

The accuracy of omission detection of the LSM combined with different network models is different. The accuracy of the YOLOv5+LSM was 92.9%, and the detection time was 15 ms. The accuracy of the YOLOv5-Ghost+LSM was 92.3%, and the detection time was 12 ms. The accuracy rate of the Faster-RCNN+LSM was 93.6%, and the detection time was 26 ms. The accuracy of the YOLOv5-AC+LSM was 94.9%, and the detection time was 17 ms. Table 7 presents the specific results and Figure 14 presents the final detection results.

Table 7. Comparison of omission detection results of different methods.

Method	Accuracy (%)	Time (ms)
YOLOv5+LSM	92.9	15
YOLOv5-Ghost+LSM	92.3	12
Faster-RCNN+LSM	93.6	26
YOLOv5-AC+LSM	94.9	17

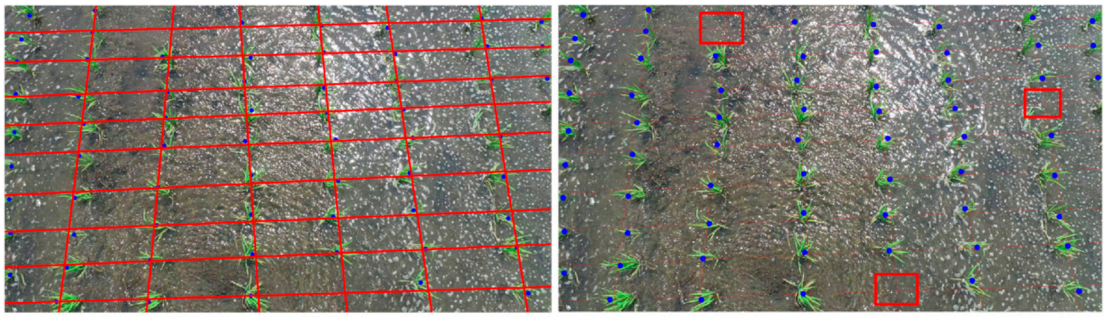

Figure 14. Seedling row fitting and omission detection effect: (**a**) fitting lines, (**b**) detected omission positions (The red squares marks the omission positions).

4. Discussion

4.1. Discussion on the Method of Working Quality Detection of Uncrewed Farm Machinery

Transplanting is one of the most common methods of rice production. An essential requirement for transplanting is straight rows of plants and even row spacing in the paddy field. Many studies have been conducted on the detection of crop rows, but very few have assessed the detection of the working quality of uncrewed rice transplanters. Yu et al. used the Otsu method for ternary classifications and proposed a 2D adaptive clustering algorithm to cluster crops with linear fitting [22]. This method of fitting crop rows with traditional machine learning is ideal. However, the features extracted by the gray map and Otsu methods for clustering are not as easily accepted by the computer as the features extracted by the neural network. Ma et al. also used the traditional machine learning method, but their method used spectral information, conducted binarization with the vegetation index, and obtained a binary graph with a significant difference between the crops and environment [23]. Then, the cluster number of crop rows in the image was obtained using the horizontal banding method at the top of the image. Finally, appropriate parametric regression equations were calculated by crop row to solve the problem. Their

method achieves high accuracy but is cumbersome in operation. Adhikari et al. adopted a deep neural network method to design an encoder-decoder structure to extract semantic information on crops in paddy fields, extended the concept of learning semantic graphs, and used it to extract crop rows to fit seedling belts [24]. In addition, Lin et al. used an R-CNN to locate rice seedlings and grouped rice seedlings with hierarchical clustering to generate reference lines, applying this method to the real-time navigation of a rice transplanter [15]. These methods use deep learning to extract features and clustering methods to generate reference lines, a typical mainstream method in recent years.

In this paper, the rice transplanting situation of an uncrewed rice transplanter is regarded as the problem of target detection, and YOLOv5 was selected as the primary network to solve the problems of complicated operation and poor feature extraction in traditional machine learning problems. Considering that YOLOv5 resizes the input image, this operation makes the incoming image the same size. However, the image size in this study and the size extracted by the feature differ, so distortion inevitably occurs after shrinking. Therefore, the ASPP module is introduced in this experiment to solve this problem, and the influence of the ASPP module on network performance at various positions in the network is explored. The ASPP module typically performs void convolution at multiple scales and combines the results to obtain a perception of objects at various scales. The number of channels is also an essential factor affecting the effectiveness of ASPP modules. This experiment also explores the recognition effect of this factor. Through comparative experiments, the accuracy of the network model proposed in this study is improved compared with that of YOLOv5 without an ASPP module. In addition, compared with mainstream models, such as YOLOv5, this model has a higher execution efficiency and recognition accuracy.

4.2. Discussion on the Influence of ASPP Modules

Moreover, ASPP extracts features by parallel sampling void convolutions as much as possible at different sampling rates for a given input image [19]. Its task is to enlarge the receptive field to obtain more valuable features at different scales. This method first appeared in the deep semantic segmentation model. The ASPP module in DeepLab improves the ability to recognize details, such as object shapes and boundaries, and reduces the dependence on the global context. Experiments reveal that DeepLab using the ASPP module performs better in semantic segmentation tasks than traditional CNNs. Zhang et al. combined an ASPP module with a U-Net model to classify urban land cover in satellite imagery [25]. Experiments have demonstrated that a network model with an ASPP module has a higher classification accuracy than the original CNN model, and the ResASPP-Unet combined with residual structure has the best effect. Wang et al. applied an ASPP module to detect cow estrus behavior [26]. They improved YOLOv5 by adding an ASPP module and introducing a channel attention mechanism and deep asymmetric bottleneck module, providing superior detection for YOLOv5, YOLOv3, Faster R-CNN, and other mainstream models. Wei et al. also introduced an ASPP module instead of spatial pyramid pooling module in the YOLOv5 network [27]. The problem they faced was the detection of edible fungi. When introducing the ASPP module, they introduced the idea of recursion and designed a recursive YOLOv5 network. Although the number of parameters is high, the network can identify 98% of edible fungi, which is 87.5% higher than the accuracy of YOLOv12X.

This research aimed to detect the working quality detection of uncrewed farm machinery. The adopted method is based on the YOLOv5 network with targeted modifications. This study introduced the ASPP module and studied its influence. The experiment found that the F1-score of the original network without the ASPP module is 88.73%, whereas it reaches 92.59% after the ASPP module is added to the appropriate part. In addition, this study also explored the number of channels of the ASPP module, finding that different channel numbers influence the detection results. Increasing the number of channels can improve detection accuracy but increases the training time.

4.3. Discussion of the Influence of CBAM

Woo et al. proposed the CBAM, a simple and effective attention module for feedforward CNNs [20]. They conducted extensive experiments on ImageNet-1K, MS COCO, and VOC 2007 detection datasets to verify the effect of CBAM and demonstrate its broad applicability. Du et al. modified the CNN structure and introduced the CBAM module and EfficientNet-B7-CBAM model [28]. Compared with AlexNet, VGG8, InceptionV88, ResNet20, DenseNet05, and other classical network models, the effectiveness of this model for seedling quality classification was increased by 16.3% to 50.121%. Ma et al. introduced the CBAM to MobileNetV2 to solve the problem of maize seed identification [29]. They improved the CBAM by replacing the cascade connection with a parallel connection. Thus, they constructed an advanced mixed attention module I_CBAM and established a new model I_CBAM_MobileNetV2, achieving high accuracy on the datasets. Wang et al. also added the CBAM to various neural networks [30]. They improved the fine-grained identification of crop diseases and pests to varying degrees compared with the previously unmodified network.

This study also introduced the CBAM and studied its influence on the experimental results at various positions in the network. After introducing the CBAM, the overall recognition effect of the network improved. The effect of the attention mechanism is to make the neural network better able to extract features that are more acceptable by the computer. Therefore, an improvement in the recognition effect after joining is expected. This study conducted experiments on the influence of the CBAM introduced in various positions on the overall effect and found that its influence was also different. The introduction of CBAM at Y5CB can achieve the most accurate effect.

Although the proposed model achieved satisfactory results in large-scale rice seedling identification and working quality detection of uncrewed farm machinery, there is still room for further optimization of the model size and detection speed. The robustness of the seedling row fitting and quality analysis methods is insufficient, and deviation in the detection results may occur under less-than-ideal environments.

In future studies, the data collected under different weather and illumination conditions will be trained to improve the recognition effect of the model. The unsupervised learning method should be assessed to reduce the training cost of this task further and improve the working quality detection task for uncrewed agricultural machinery.

5. Conclusions

This paper proposes a real-time, accurate working quality detection method for uncrewed rice transplanters adapted to complex field environments. The main conclusions are as follows:

(1) The performance test of the model demonstrates that the detection accuracy of machine-inserted rice seedlings in the actual field environment can reach 95.8%, the recall rate is 91.1%, the mAP@.5 is 94.1%, and the F1-score is 93.39%, 4.66% higher than the original YOLOv5 network.

(2) Compared with the test results, the detection performance of the YOLOv5-AC network model compared with mainstream models, such as YOLOv5, has higher recognition accuracy and stable execution efficiency. The overall performance has certain advantages, solving the problem of uncrewed rice transplanter working quality detection in a natural field environment and filling the gap in this direction.

Uncrewed farm machinery working quality detection based on RGB images can meet the actual requirements of faster detection speed and lower computational complexity and is more suitable for real-time detection of rice seedlings in complex environments. Identifying and detecting rice seedlings and working quality for uncrewed farm machinery are of great significance for improving work efficiency and constructing uncrewed machinery systems.

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

Funding: This research was funded by Science and Technology Innovation 2030- "New Generation of Artificial Intelligence" major project grant number 2021ZD0110904 and Scholars Program of Northeast Agricultural University: Young talents grant number 20QC32 and The University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province grant number UNPYSCT-2020091.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, Z.; Ma, X.; Li, X.; Chen, L.; Li, H.; Yuan, Z. Research progress of rice transplanting mechanization. *Nongye Jixie Xuebao* **2018**, *49*, 1–20.
2. Luo, X.; Wang, Z.; Zeng, S.; Zang, Y.; Yang, W.; Zhang, M. Recent advances in mechanized direct seeding technology for rice. *J. South China Agric. Univ.* **2019**, *40*, 1–13.
3. Patil, S.; Shahare, P.; Aware, V.V. Field testing of power operated Paddy transplanter suitable for root washed seedlings. *Int. J. Pure Appl. Biosci.* **2017**, *5*, 1146–1152.
4. Yong, Z.; Ying, X.; Yang, X. Research on rice seedling recognition algorithm based on machine vision in rice field. *J. Hunan Agric. Univ.* **2018**, *44*, 320–325.
5. Wang, N.; Zhang, N.; Dowell, F.E.; Sun, Y.; Peterson, D.E. Design of an optical weed sensor usingplant spectral characteristics. *Trans. ASAE* **2001**, *44*, 409. [CrossRef]
6. Meng, Q.; Zhang, M.; Yang, G.; Qiu, R.; Xiang, M. Guidance line recognition of agricultural machinery based on particle swarm optimization under natural illumination. *Trans. Chin. Soc. Agric. Mach.* **2016**, *47*, 11–20.
7. Jiang, G.; Wang, Z.; Liu, H. Automatic detection of crop rows based on multi-ROIs. *Expert Syst. Appl.* **2015**, *42*, 2429–2441. [CrossRef]
8. Bendig, J.; Yu, K.; Aasen, H.; Bolten, A.; Bennertz, S.; Broscheit, J.; Gnyp, M.L.; Bareth, G. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *39*, 79–87. [CrossRef]
9. Yu, Z.; Cao, Z.; Wu, X.; Bai, X.; Qin, Y.; Zhuo, W.; Xiao, Y.; Zhang, X.; Xue, H. Automatic image-based detection technology for two critical growth stages of maize: Emergence and three-leaf stage. *Agric. For. Meteorol.* **2013**, *174*, 65–84. [CrossRef]
10. Espinoza, K.; Valera, D.L.; Torres, J.A.; López, A.; Molina-Aiz, F.D. Combination of image processing and artificial neural networks as a novel approach for the identification of Bemisia tabaci and Frankliniella occidentalis on sticky traps in greenhouse agriculture. *Comput. Electron. Agric.* **2016**, *127*, 495–505. [CrossRef]
11. Kumar, J.P.; Domnic, S. Image based leaf segmentation and counting in rosette plants. *Inf. Process. Agric.* **2019**, *6*, 233–246.
12. Khaki, S.; Pham, H.; Han, Y.; Kuhl, A.; Kent, W.; Wang, L. Convolutional neural networks for image-based corn kernel detection and counting. *Sensors* **2020**, *20*, 2721. [CrossRef] [PubMed]
13. Häni, N.; Roy, P.; Isler, V. A comparative study of fruit detection and counting methods for yield mapping in apple orchards. *J. Field Robot.* **2020**, *37*, 263–282. [CrossRef]
14. Zabawa, L.; Kicherer, A.; Klingbeil, L.; Töpfer, R.; Kuhlmann, H.; Roscher, R. Counting of grapevine berries in images via semantic segmentation using convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2020**, *164*, 73–83. [CrossRef]
15. Lin, S.; Jiang, Y.; Chen, X.; Biswas, A.; Li, S.; Yuan, Z.; Wang, H.; Qi, L. Automatic detection of plant rows for a transplanter in paddy field using faster r-cnn. *IEEE Access* **2020**, *8*, 147231–147240. [CrossRef]
16. Wang, S.; Yu, S.; Zhang, W.; Wang, X. Detection of rice seedling rows based on Hough transform of feature point neighborhood. *Nongye Jixie Xuebao* **2020**, *51*, 18–25.
17. Zhang, Q.; Wang, J.; Li, B. Extraction method for centerlines of rice seedings based on YOLOv3 target detection. *Trans. Chin. Soc. Agric. Mach.* **2020**, *51*, 34–43.
18. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
19. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef]
20. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
21. Abdi, H. The method of least squares. *Encycl. Meas. Stat.* **2007**, *1*, 530–532.
22. Yu, Y.; Bao, Y.; Wang, J.; Chu, H.; Zhao, N.; He, Y.; Liu, Y. Crop row segmentation and detection in paddy fields based on treble-classification otsu and double-dimensional clustering method. *Remote Sens.* **2021**, *13*, 901. [CrossRef]

23. Ma, Z.; Tao, Z.; Du, X.; Yu, Y.; Wu, C. Automatic detection of crop root rows in paddy fields based on straight-line clustering algorithm and supervised learning method. *Biosyst. Eng.* **2021**, *211*, 63–76. [CrossRef]
24. 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]
25. Zhang, P.; Ke, Y.; Zhang, Z.; Wang, M.; Li, P.; Zhang, S. Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery. *Sensors* **2018**, *18*, 3717. [CrossRef] [PubMed]
26. Wang, R.; Gao, Z.; Li, Q.; Zhao, C.; Gao, R.; Zhang, H.; Li, S.; Feng, L. Detection Method of Cow Estrus Behavior in Natural Scenes Based on Improved YOLOv5. *Agriculture* **2022**, *12*, 1339. [CrossRef]
27. Wei, B.; Zhang, Y.; Pu, Y.; Sun, Y.; Zhang, S.; Lin, H.; Zeng, C.; Zhao, Y.; Wang, K.; Chen, Z. Recursive-YOLOv5 network for edible mushroom detection in scenes with vertical stick placement. *IEEE Access* **2022**, *10*, 40093–40108. [CrossRef]
28. Du, X.; Si, L.; Jin, X.; Li, P.; Yun, Z.; Gao, K. Classification of plug seedling quality by improved convolutional neural network with an attention mechanism. *Front. Plant Sci.* **2022**, *13*, 967706. [CrossRef] [PubMed]
29. Ma, R.; Wang, J.; Zhao, W.; Guo, H.; Dai, D.; Yun, Y.; Li, L.; Hao, F.; Bai, J.; Ma, D. Identification of maize seed varieties using MobileNetV2 with improved attention mechanism CBAM. *Agriculture* **2022**, *13*, 11. [CrossRef]
30. Wang, M.; Wu, Z.; Zhou, Z. Fine-grained identification research of crop pests and diseases based on improved CBAM via attention. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 239–247.