Article

Identification and Counting of Sugarcane Seedlings in the Field Using Improved Faster R-CNN

Yuyun Pan [1,†], **Nengzhi Zhu** [1,2,†], **Lu Ding** [1], **Xiuhua Li** [1,3,*], **Hui-Hwang Goh** [1] (ID), **Chao Han** [1] **and Muqing Zhang** [3,4,5]

1 School of Electrical Engineering, Guangxi University, Nanning 530004, China
2 Anhui Transportation Holding Group Co., Ltd., Hefei 230088, China
3 Guangxi Key Laboratory of Sugarcane Biology, Guangxi University, Nanning 530004, China
4 State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangxi University, Nanning 530004, China
5 IRREC-IFAS, University of Florida, Fort Pierce, FL 34945, USA
* Correspondence: lixh@gxu.edu.cn
† These authors contributed equally to this work.

Abstract: Sugarcane seedling emergence is important for sugar production. Manual counting is time-consuming and hardly practicable for large-scale field planting. Unmanned aerial vehicles (UAVs) with fast acquisition speed and wide coverage are becoming increasingly popular in precision agriculture. We provide a method based on improved Faster RCNN for automatically detecting and counting sugarcane seedlings using aerial photography. The Sugarcane-Detector (SGN-D) uses ResNet 50 for feature extraction to produce high-resolution feature expressions and provides an attention method (SN-block) to focus the network on learning seedling feature channels. FPN aggregates multi-level features to tackle multi-scale problems, while optimizing anchor boxes for sugarcane size and quantity. To evaluate the efficacy and viability of the proposed technology, 238 images of sugarcane seedlings were taken from the air with an unmanned aerial vehicle. Outcoming with an average accuracy of 93.67%, our proposed method outperforms other commonly used detection models, including the original Faster R-CNN, SSD, and YOLO. In order to eliminate the error caused by repeated counting, we further propose a seedlings de-duplication algorithm. The highest counting accuracy reached 96.83%, whilst the mean absolute error (MAE) reached 4.6 when intersection of union (IoU) was 0.15. In addition, a software system was developed for the automatic identification and counting of cane seedlings. This work can provide accurate seedling data, thus can support farmers making proper cultivation management decision.

Keywords: sugarcane seedling; convolutional neural network; object detection; unmanned aerial vehicle; faster RCNN

1. Introduction

Sugarcane accounts for around 86% of global sugar output and 90% of China's total sugar production [1,2]. China cultivated an estimated 1.35 million hectares of sugarcane in 2020, yielding 108.12 million tons (http://www.stats.gov.cn/english/, accessed on 25 September 2022). Due to variety, the sugarcane emergence rate in each field varies. Unsuitable climate and environment, as well as pests and diseases, can damage sugarcane buds and prevent them from germinating, resulting in reduced seedling emergence, and thereby, productivity and profitability [3–6]. Sugarcane emergence as a phenotypic feature is crucial for both sugarcane breeding and planting. Effective evaluation of the emergence rate of sugarcane seedlings can not only contribute to better understanding of the physiological and genetic mechanisms underlying crop growth and development, but also facilitate the implementation of timely crop management strategies to prevent yield losses.

Traditional seedling survey primarily relies on manual sampling, which makes it difficult to implement over a vast area, due to statistical and manual bias [7]. With the progress of information and UAV technology [8–10], high-throughput quick detection and counting of sugarcane seedlings have become feasible by getting high-resolution images of sugarcane field seedlings.

Artificial intelligence approaches, such as computer vision that trains computers to interpret and understand the visual world, have been implemented in precision agriculture [11–14], wherein object detection is crucial for crop monitoring. Traditional object detection primarily focuses on color, texture, form, to differentiate targets. In seedling detection, a crops segmentation method (AP-HI) was presented to automatically detect seedling emergence and the three-leaf stage of corn seedlings with a 96.68% average accuracy [15]. Liu et al. [16] used excessive green and Otsu's approach to extract wheat seedlings data and developed a chain code-based skeleton optimization method for automatic counting of wheat seedlings, achieving an accuracy of 89.98%. In addition, Zhao et al. [17] collected shape characteristics such as area, aspect ratio, and ellipse fit from segmented rape plants to develop a seedling number regression model with an R^2 of 0.846 and a mean relative error of 6.83. Nevertheless, these techniques are noise-sensitive, and vulnerable to environmental factors. Many machine learning techniques, like as Support Vector Machine (SVM), K-Means, and Random Forest (RF), have been applied in the field of seedling detection since the introduction of machine learning. Xia et al. [18] used SVM and maximum likelihood classification to identify cotton plants in images and devised a method based on morphological characteristics of cotton plants to count the number of overlapping cotton seedlings with 91.13% accuracy. Li et al. [19] employed six morphological characteristics of potato plants as variables in a random forest classifier to estimate the number of potato seedlings that emerge, with an R^2 of 0.96. Additionally, Banerjee et al. [20] developed a method to estimate wheat seedlings using multispectral images captured by drones, combining spectral and morphological information extracted from multispectral images and developing machine learning models, regression tree, SVM, and Gaussian process regression, to estimate wheat seedlings, with the Gaussian process regression model performing the best with R^2 of 0.86 and MAE of 3.21. However, the primary issue with conventional machine learning was that the model's expression capability and amount of feature abstraction were limited, resulting in low robustness and recognition and detection accuracy [21].

Deep learning (DL) [22] is a type of machine learning, which is essentially a mechanism for learning features spontaneously. Deep learning networks incorporate more hidden layers than typical machine learning approaches in order to achieve greater learning capabilities and consequently superior performance and precision. Convolutional neural network (CNN), the most representative model of deep learning, is widely utilized in object detection today. CNN's highly hierarchical structure and vast learning capacity enable it to perform well in classification and prediction and to cope with highly complicated situations in a flexible manner [23]. In recent years, deep learning has introduced new concepts for intelligent agriculture [24] and produced positive outcomes. In seedling detection, Villaruz et al. [25] used deep learning models to categorize Philippine indigenous plant seedlings with a validation set accuracy of above 90%. Li et al. [26] proposed an automatic detection method for hydroponic lettuce seedlings based on an improved Faster RCNN framework, taking the dead and double-planting status of seedlings as research objects, with an average accuracy of 86.2%. However, these data were collected in a laboratory with a single background, which was less difficult than the field's complex environment. To count cotton seedlings in the field, Jiang et al. [27] developed a method based on the Faster R-CNN [28] model with the inception ResNet v2 feature extractor and Kalman filter. Quan et al. [29] developed an enhanced Faster R-CNN model for a field robot platform to recognize maize seedlings with a precision of over 97.71%. In contrast to the method of obtaining seedling images using a UAV platform, the sampling area and moving speed using handheld camera and field robot are restricted. Fromm et al. [30] explored the possibility of using three CNN to automatically detect conifer seedlings, of which Faster R-CNN

performed best (mean average precision 81%). Lin et al. [31] employed two deep learning models, the SSD-MobileNet V2 model and CenterNet ResNet50, to detect and count cotton plants at the seedling stage using UAS imaging, with 79% and 88% average precision, respectively. Oh et al. [32] introduced a deep learning-based plant-counting algorithm to reduce the number of input variables and used YOLOv3 to separate, find, and count cotton seedlings. Neupane et al. [33] employed the Faster R-CNN to automatically count banana seedlings in UAV aerial images with a detection accuracy of 97%. Feng et al. [34] utilized the Faster R-CNN model with ResNet18 as the features extractor to recognize and count cotton seedlings in the field; the method could estimate stand count accurately with R^2 of 0.95 in the test dataset. Additionally, the one-stage pretrained object detector EfficientDet-D1 was utilized to detect the faulty paddy rice seedlings with the highest precision and F1-Scores of 0.83 and 0.77, respectively [35]. These reported studies demonstrated the efficiency of CNN-based object detection technology. Compared to the one-stage object detection algorithm, the Faster R-CNN architecture (which is a representative of the two-stage target detection algorithm) introduced a region proposal network that significantly improves the seedling detection accuracy. However, unlike banana, maize, cotton, and other crop seedlings, whose appearance is reasonably regular and intuitive, sugarcane seedlings resemble weeds and are difficult to detect. Therefore, the algorithm has to be updated for detection performance based on the characteristics of sugarcane seedlings.

In this study, RGB images of sugarcane seedlings were acquired and used as input data to an improved Faster R-CNN model, based on the features of sugarcane seedlings. ResNet 50 was employed as the feature extraction network and SN-block as the channel attention method to acquire high-resolution sugarcane seedlings' feature expressions. Feature pyramid network (FPN) was also introduced to fuse multi-level features to solve multi-scale problems, and the anchor boxes were tuned to accommodate the size and number of sugarcane seedlings. Furthermore, a seedlings de-duplication approach based on the image cropping with overlap technique was presented to reduce the inaccuracy caused by repeated counting. Finally, by training and evaluating our dataset, a software tool was operationalized for the automatic detection and counting of sugarcane seedlings with high performance.

2. Experimental Data

2.1. Experimental Sugarcane Field

The experimental sugarcane field as shown in Figure 1 is situated in the Guangxi University Agricultural Science New City in Fusui County, Chongzuo City, Guangxi Zhuang Autonomous Region (22.5121°N, 107.7897°E). Guangxi is the largest sugarcane and sugar producer in China [36]. The region has a subtropical monsoon climate with an average annual temperature of 22 °C and substantial precipitation resources. The average annual rainfall is 1647.7 mm, with the majority of precipitation falling between May and September, which is favorable for sugarcane cultivation. Low-altitude aerial images of sugarcane seedlings were acquired individually for two experimental areas in this investigation. Test fields 1 and 2 were planted on 6 July 2020 and 13 July 2020, with corresponding areas of 0.43 ha and 1.19 ha.

2.2. UAV Imagery Data Acquisition

The RGB imageries of the seedling fields were registered using DJI Phantom 4 Pro (DJI, Shenzhen, China) (Figure 2a) on 2 August 2020, in the morning (8:00 a.m. to 11:00 a.m.) and afternoon (3:00 p.m. to 5:00 p.m.). DJI GO4 software (Version 4.3.36, DJI, Shenzhen, China) was used to design the flight plan. The flight height was 10 m above the ground, the flight speed was 2 m/s, and both the side overlap and front overlap was set to 80%. The camera parameters were set as follows: the ISO was 100, the aperture was F/3.5, the focal length was 8.8 mm, the capture interval was 2 s, and the shooting angle was vertically downward (Figure 2b). A total of 238 RGB images (120 for field 1 and 118 for field 2) with resolution of 5472 pixels × 3648 pixels, and the ground sampling distance (GSD) of 2.7 mm, were

captured. Differences in the color of the soil, the intensity of the light, and the background between the images of field 1 and field 2 were visible.
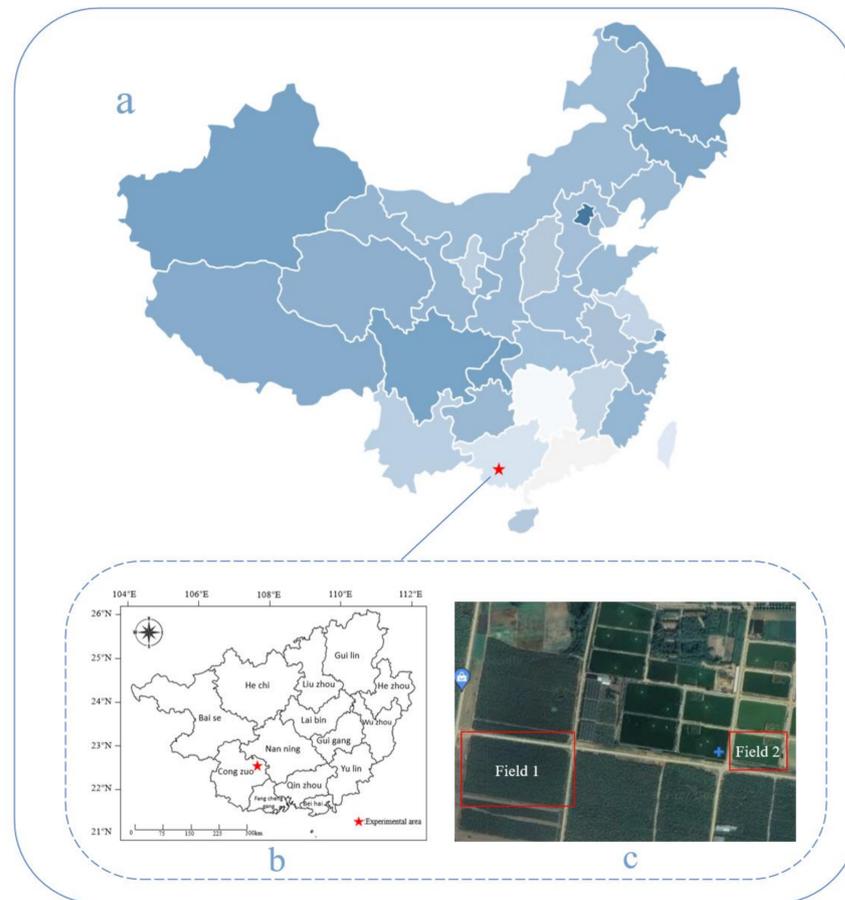

Figure 1. The geographic information of the experimental sugarcane fields: (**a**) location of the study area in China; (**b**) geographical location of the study area; (**c**) satellite image of the study fields (source: google satellite map).
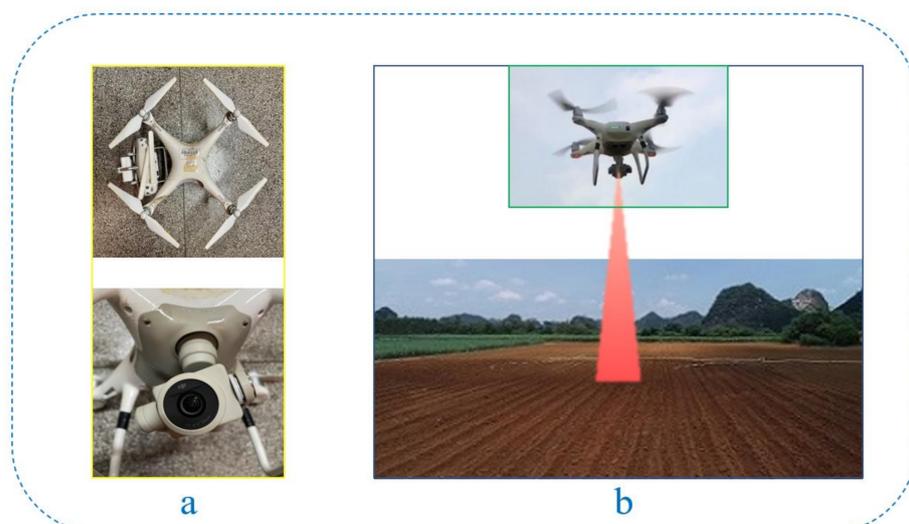

Figure 2. Data-collecting platform. (**a**) UAV and RBG camera. (**b**) Collection of images of sugarcane seedlings.

2.3. COCO Dataset

COCO (Common Objects in Context) [37] is considered one of the most concerned and authoritative datasets of competitions in the field of computer vision. It is a large object detection, segmentation, and character dataset marked by Microsoft in 2014, with 330 k images and 80 categories. During transfer Learning, the plant categories of the COCO dataset can be initialized for the sugarcane seedling feature extraction network, making training processing time and resources efficient for our new detection task.

3. Methodology

3.1. Overall Program

Figure 3 shows the experiment's overall flowchart, which consists of three primary components: dataset preparation, sugarcane seedling detection model, and duplication removal. The primary objective of dataset preparation was to create a dataset for training a deep learning model. This involves mostly image enhancement, image clipping, data enhancement, and sugarcane seedling labelling. The seedling identification model construction module improved the Faster R-CNN model which was then trained and validated on the prepared dataset to obtain the sugarcane seedling detection results for each image tile. The duplication removal component is primarily responsible for splicing the detection results of each image tile back to their original size and de-duplicating the sugarcane seedling detection frame in order to retrieve the statistical data from the original images.
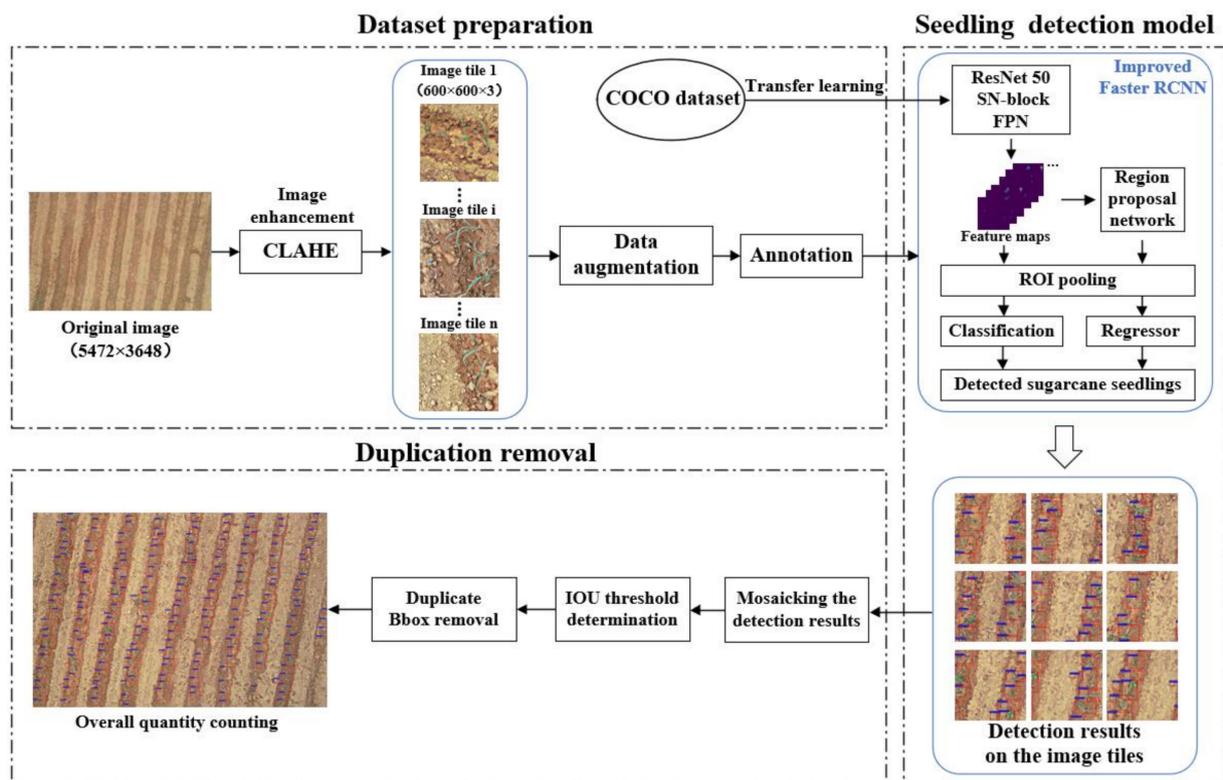
Figure 3. Overall flowchart of the sugarcane seedling detection and counting method.

3.2. Data Preprocessing

Data preprocessing mainly consisted of four components: image enhancement, image cropping, data augmentation, and sugarcane seedling labeling.

The primary objective of image enhancement was to increase the contrast between the plants and the background, and thus, to raise the recognition accuracy of sugarcane seedlings. Contrast limited adaptive histogram equalization (CLAHE) [38], which could enhance the contrast of the image while suppressing noise by employing an inverse sharp-

ening mask technique, was used in this study (Figure 4a) to create enhanced images that highlight high-frequency information such as blade edges.

To assure further the ability of recognition, 238 original images were cropped with overlap into 600 pixels × 600 pixels sub images with 32.3% (194 pixel) and 36.5% (219 pixel) horizontal and vertical overlap, respectively. Among a total of 2790 sub-images containing the seedlings, 2392 were retained for training purposes and 398 for validation.

Data augmentation was achieved through image flipping and rotation (Figure 4b), resulting in 14,352 sub-images in the training set and 2388 sub-images in the validation set. The image labelling tool LabelImg was used for seedling labelling (Figure 4c). This self-prepared dataset is denoted as SC dataset which has 16,740 images.
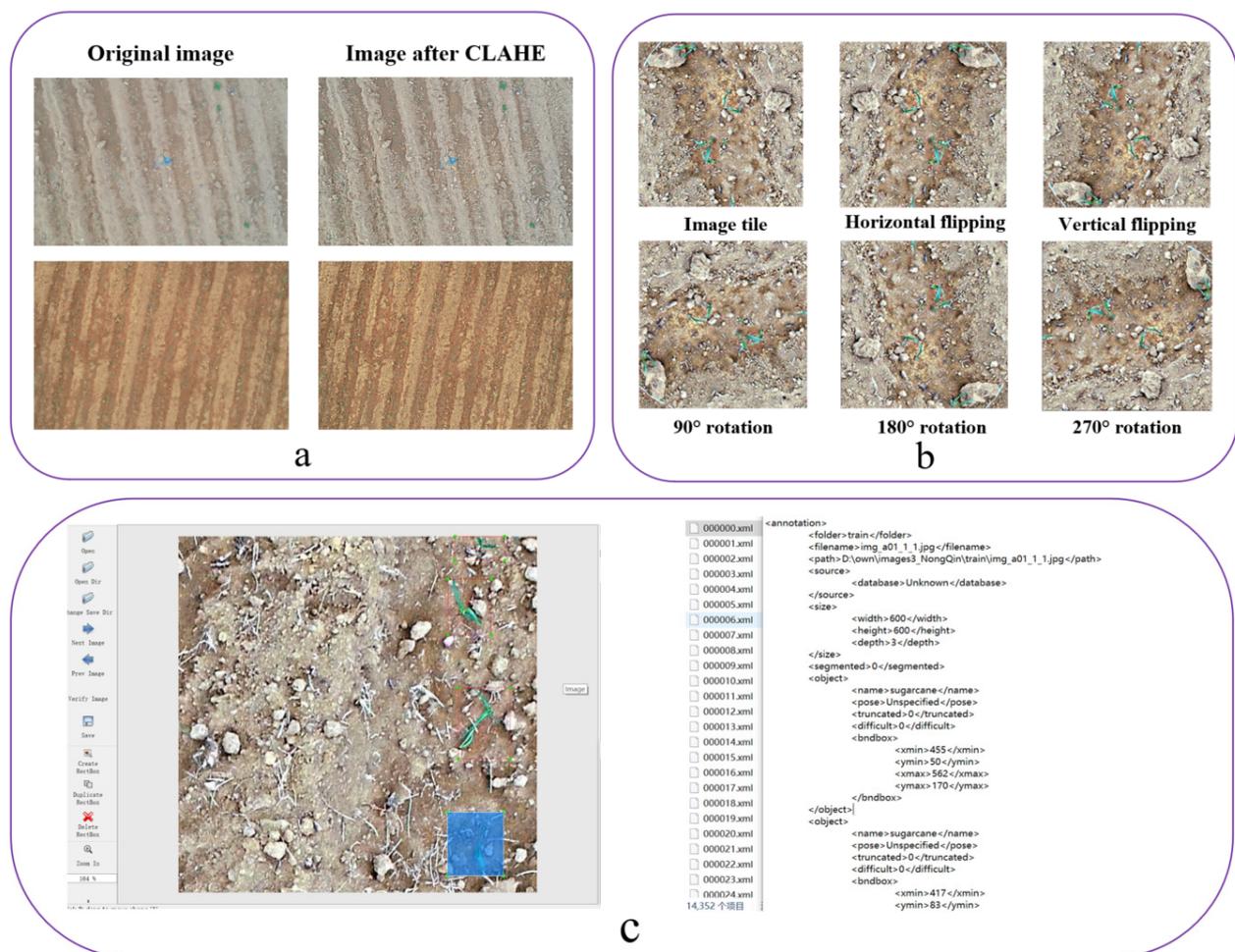
Figure 4. Data preprocessing. (**a**) The CLAHE process. (**b**) Data augmentation. (**c**) Images of sugarcane seedlings with annotations.

3.3. Improved Faster R-CNN Detection Algorithm

Faster R-CNN has two networks: region proposal network (RPN) for generating region proposals and a network using these proposals to detect objects. RPN ranks region boxes (called anchors) and proposes the ones most likely containing objects. In this study, the Faster RCNN object detection framework for sugarcane seedlings detection contains four primary components: feature extraction of the sugarcane seedlings network, region proposal generation, region of interest (RoI) pooling, classification, and bounding box regression. Enhancing the feature representation capability of convolutional neural networks was preferred in this study as a pathway for resolving the issue of false detection and missed detection due to few seedling pixels and mutual sheltered leaves. ResNet50 was used as the feature extraction network and augment CNN's feature extraction capacity

with SN-block, a lightweight attention mechanism module. The technique also proposed a feature pyramid network (FPN) to combine different scale feature maps of the network in order to increase the accuracy of detecting small targets. Finally, the anchor boxes in the region suggestion network (RPN) were optimized to increase the accuracy of sugarcane seedling detection.

3.3.1. Feature Extraction Using ResNet-50

VGG16 is the foundation of the original Faster RCNN. With the advancement of convolutional neural networks, the vanishing gradient problem resulting from the increase in neural network depth is becoming increasingly apparent. ResNet [39] was proposed as a solution to the problem of network degradation. Due to residual structure design, the model's precision can be improved with the increase of the number of convolution layers. Considering the concentration and small object characteristics of sugarcane seedlings, we selected the ResNet50 network, which has a relatively simple structure, as the feature extractor (Figure 5).
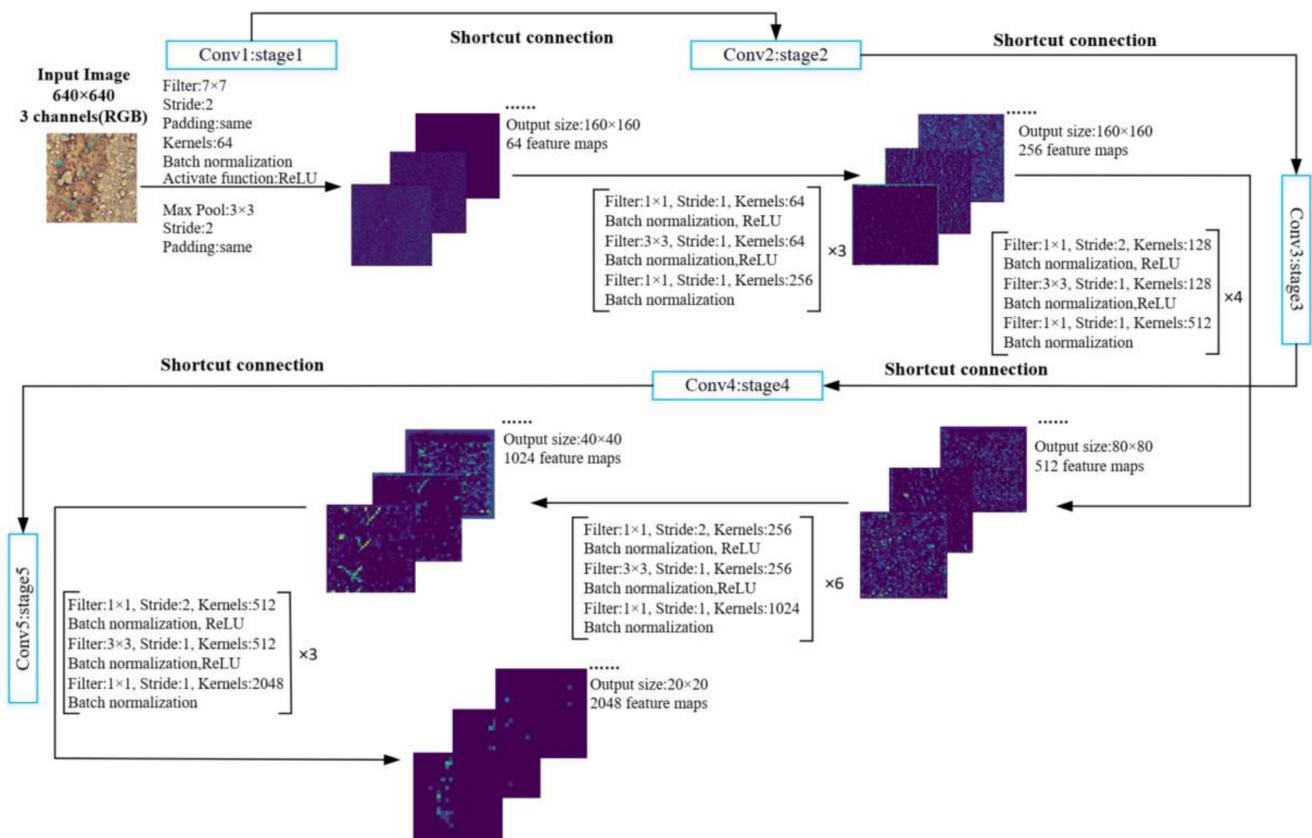
Figure 5. The structure of the ResNet50 network.

The network took advantage of five phases of Conv block and identity block, with Conv block used to alter the network's dimension and identity block used to increase its depth. For feature extraction, each block consists of a distinct number of convolution layer combinations, batch normalization layers, ReLU activation layers and max pooling layers. Additionally, shortcut connection does identity mapping for residual learning. C1~C5 denotes the first to fifth feature maps and their corresponding sizes C1~C5 are (320,320,64), (160,160,256), (80,80,512), (40,40,1024), (20,20,2048), respectively.

3.3.2. Attention Network SN-Block

A lightweight network structure based on the channel attention mechanism SN-Block was proposed to address the issue of sugarcane seedlings with variable sizes and mutual

sheltered leaves. The lower-level features are weighted in the backbone network to enhance the target features. As demonstrated in Figure 6, the SN-Block module can be implemented in ResNet50 to increase the performance of the convolutional neural network by expanding the receptive field of the feature extraction layer.

SN-block performs a convolution operation and global average pooling on the feature map F_{in} (H × W × C feature map) and generates a vector F_{global} with global feature information (1 × 1 × C feature map) to give it a global receptive field while maintaining the same number of channels. The fully connected layer is then utilized to compute weights for each feature channel, and the sigmoid function is employed for normalization. Finally, the normalized weight corresponding to each channel is determined, and after weighting the channel, the feature map F_{out} is obtained. By modulating the weight, the SN block emphasizes the significant aspects and downplays the minor ones, making the extracted features more directive.
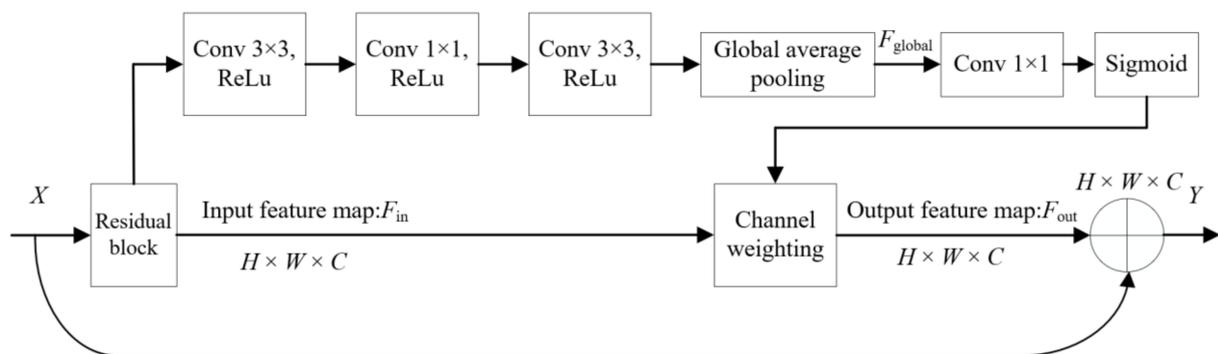

Figure 6. The structure of SN-Block.

The calculation process of the SN-Block module is:

$$F_{out} = H_c(F_{in}) \otimes F_{in} \tag{1}$$

where Fin is the input feature map by backbone network; F_{out} is the output feature map obtained after weighting processing; $H_c(F_{in})$ is the operation of obtaining the channel weight through F_{in}, and $\otimes$ is the operation of weighting each channel.

$$H_c(F_{in}) = \sigma\left[M\left(F_{global}\right)\right] \tag{2}$$

We use Equation (2) to achieve channel weight. "σ" is the sigmoid function; M is 1 × 1 convolution operation; F_{global} is determined by the feature map F_{in} is obtained after global average pooling.

$$F_{global} = \rho\{\delta[O(F_{in})]\} \tag{3}$$

In Equation (3), ρ is the global average pooling operation; δ is the ReLu activation function; O is a series of convolution operations performed on the input feature map.

3.3.3. Multi-Scale Feature Fusion

This study applied the feature pyramid network (FPN) [40] (Figure 7) for feature fusion in order to make full use of the features of each stage extracted by the ResNet50 network. The strong semantic information at the top level is conducive to classification, while the high-resolution information at the bottom level is conducive to positioning. For small objects such as sugarcane seedlings, it is possible to obtain an informative multiscale feature map, considerably enhancing the model's detection precision. Down-top pathways, top-down pathways, and lateral connections comprise the FPN's overall network structure. The down-top pathway is the process of extracting features from the input image via the backbone. The top-down pathway is the process of up sampling the high-level feature map

generated from the down-top pathway to maintain the same size as the low-level feature maps of the previous stage.
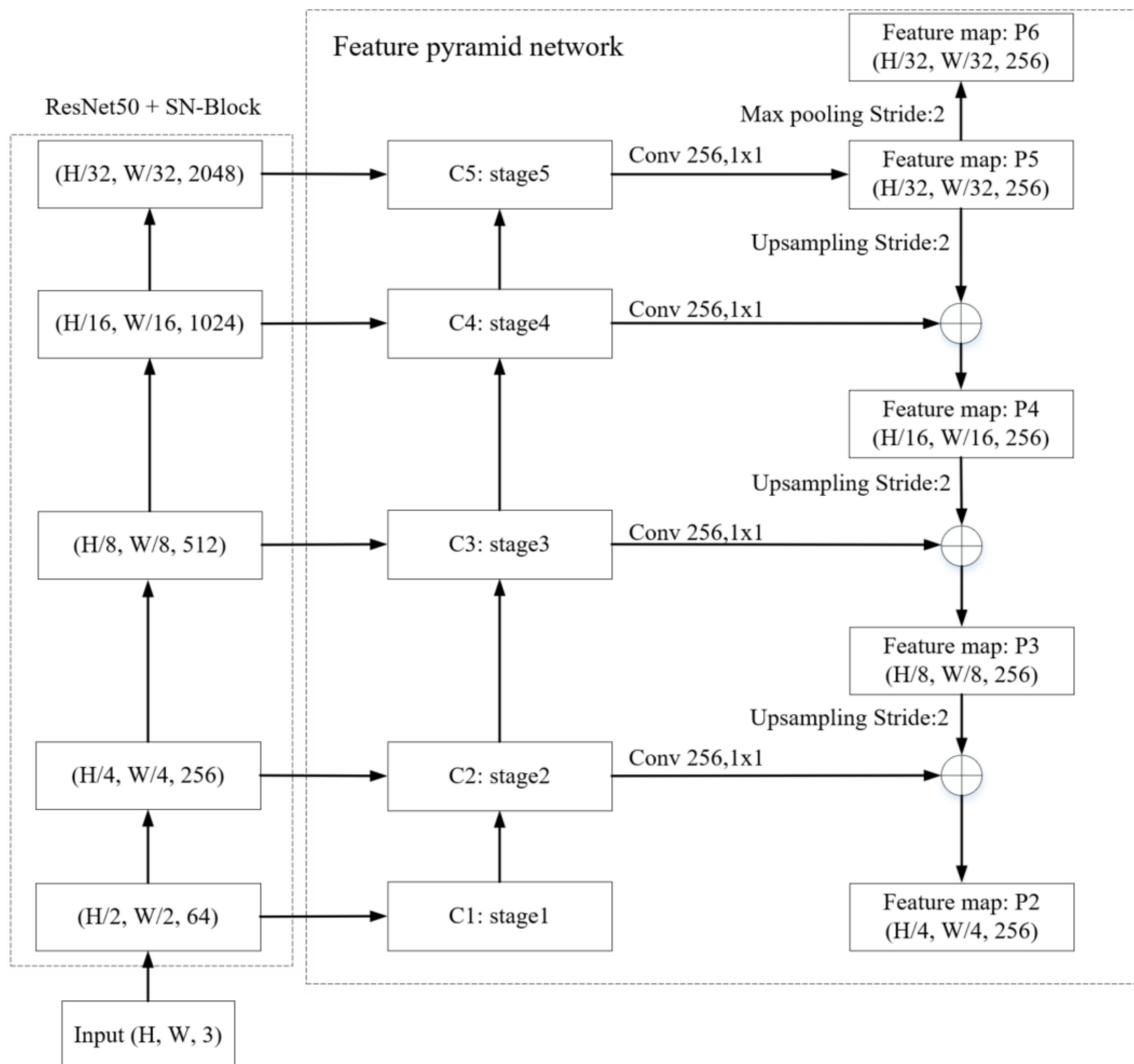
Figure 7. The structure of the feature pyramid network.

3.3.4. Anchor Optimization for the Regional Proposed Network (RPN)

RPN is utilized to produce the Region of Interest (RoI). The output feature maps from the FPN structure are utilized as the inputs of the RPN and convolved with a 3×3 sliding window. Anchor refers to the point corresponding to the center of the sliding window in the original pixel space, which is used to generate many anchor boxes with variable scales and aspect ratios.

The original Faster R-CNN anchor boxes had three aspect ratios (1:2, 1:1, 2:1). To make the model more suitable for the geometry of sugarcane seedlings, the aspect ratios of sugarcane seedling marker frames in the training set were estimated. As a result, the four most common aspect ratios were 1:1, 2:1, 4:5, and 1:2. Therefore, the four aspect ratios derived from statistical analysis were used to replace the aspect ratios of the original anchor boxes and set their size to (16,32,64,128,256).

3.4. Construction of a System for the Automatic Recognition and Counting of Sugarcane Seedlings

In order to display the detection results on the original image, a coordinate transformation of the sub images obtained from model detection back to the corresponding positions of the original image was performed, non-maximum suppression (NMS) algorithm was

then applied to remove the repeatedly identified same frames in different sub images by setting the IoU threshold λ and thus, retain the one with the highest confidence. Using the pyqt5 framework, the study also constructed a software system for the automatic detection and counting of sugarcane seedlings. The system features an intuitive graphical user interface (GUI) that guides users through the process of training a deep learning model with their own datasets and utilizing the model to generate the detection map and counting results of sugarcane seedlings. The system's flowchart and GUI are shown in Figure 8.

(**a**)

(**b**)

Figure 8. (**a**) The flowchart of the software system. (**b**) The graphical user interface (GUI) allows users to choose to train their own models or use existing model.

3.5. Assessment Metrics

3.5.1. Metrics for Detection Accuracy Assessment

Precision, Recall, Average Precision (AP), Average Recall (AR), and F1 score were employed as indices to evaluate the precision of sugarcane seedling detection. Precision

and recall are the most often employed metrics for evaluating object detection techniques. Precision indicates the ratio of truly detected seedlings over the sum of truly and falsely detected seedlings in the datasets. Recall describes the ration between the truly detected seedlings over actually existed seedling (including therefore those not detected). AR is the average recall rate across a particular number of data sets, while AP is the area of the PR curve used to illustrate the accuracy rate under different recall rates. The F1-score seeks to achieve a balance between the two metrics, meaning that the F1 score becomes high only when both precision and recall are high. The equations for calculating Precision, Recall, AP, AR, and the F1 score are Equations (4)–(8), respectively.

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1 - score = \frac{2Precision \times Recall}{Precision + Recall} \tag{6}$$

$$AP = \int_0^1 PrecisiondRecall \tag{7}$$

$$AR = \frac{1}{Q} \sum_{q=1}^{Q} Recall(q) \tag{8}$$

where:

- True Positive (*TP*)—the number of correctly detected sugarcane seedlings
- False Positive (*FP*)—the number of weed or other features incorrectly detected as sugarcane seedlings
- False Negative (*FN*)—the number of sugarcane seedlings that are not detected
- *Q*—the total number of images to be verified
- *Recall (q)*—the recall rate of an image q in the data set

3.5.2. Counting Accuracy Assessment Metrics

The algorithms were developed to evaluate the performance of identification and quantification of sugarcane seedlings in the test datasets. Identified sugarcane seedlings were labelled by bounding boxes. Therefore, the number of bounding boxes in each test image represented the number of identified sugarcane seedlings.

Mean Absolute Error (MAE), Classification Accuracy (CA), and R Squared (R^2) were employed as evaluation indicators for determining the counting performance of this application. The MAE quantifies the average of the absolute error between the predicted value and the actual value. CA reflects the fraction of sugarcane seedlings that are correctly identified out of the total number of sugarcane seedlings. R^2 (the coefficient of determination) expresses how well the independent variables in the model explain the variability in the dependent variable or how well the model fits the dependent variable. The lower the value of MAE, the greater the values of CA and R^2, and the higher the counting performance, determined as in Equations (9)–(11):

$$\text{MAE} = \frac{1}{n} \sum_{1}^{n} |a_i - c_i| \tag{9}$$

$$\text{CA} = \left(1 - \frac{1}{n} \sum_{1}^{n} \frac{|a_i - c_i|}{a_i}\right) \times 100\% \tag{10}$$

$$\text{R}^2 = 1 - \frac{\sum_{1}^{n} (a_i - c_i)^2}{\sum_{1}^{n} (a_i - a_i^*)^2} \tag{11}$$

where:

- n = the total number of test images
- a_i = the ground truth number of sugarcane seedlings in the i-th aerial image
- c_i = the model prediction number (the total number of detected target frames)
- a^*_i = the average ground truth number

3.6. Training Platform and Parameter Settings

Our experimental platform was programmed using Python 3.6 (Python Software Foundation, Wilmington, DE, USA) in Windows 10 64-bit environment powered by Intel® Core i9-9900X as CPU and GeForce GTX 2080Ti (NVIDIA, Santa, Clara, CA, USA) as GPUs. TensorFlow was selected as the deep learning framework.

As the ground datasets for annotated sugarcane seedlings were scarce, transfer learning was used to produce higher training performance [41]. This method used the weight parameters obtained by training the Faster RCNN on the COCO dataset as a pre-trained model and applied it to the SC dataset of this study for model fine-tuning in order to decrease training time and improved the accuracy of the newly trained model without more extensive training data. The experiment's training settings are listed in Table 1.

Table 1. Training parameters of Faster R-CNN.

Parameters	Value
Learning rate	0.001
Weight decay	0.0005
Momentum	0.9
Epoch	150
Regularization factor	0.0001
optimizer	Momentum Optimizer

4. Results and Discussion

4.1. Model Performance Comparison

4.1.1. Comparison of Detection Performance before and after Model Improvement

To confirm the performance of the model based on Faster R-CNN, a series of ablation experiments were carried out and the outcomes were compared (Table 2). First, adjusting the number of anchor boxes raised the average precision and recall by roughly 1.6% and 1.35%, respectively. Then, the replacement of the original feature extraction networks VGG16 by ResNet50 demonstrated that the residual network was more suited to serve as the backbone network of the sugarcane seedling detection model. In addition, the attention mechanism SN-block proposed in this research increased the average accuracy by 0.71% and the average recall by 0.43%, thereby resolving the problem of shielding between sugarcane seedling leaf blades. Finally, the use of FPN structure may fuse the hierarchical semantic properties of sugarcane seedlings, which was helpful for recognizing small targets, hence enhancing the algorithm's detection precision.

Table 2. Comparison of detect results under different improvement methods.

Method	The Number of Anchor Boxes	AP/%	AR/%
Original Faster R-CNN	9	84.98	83.56
Faster R-CNN	20	86.58	84.91
Faster R-CNN+ResNet50	20	89.88	87.22
Faster R-CNN+ResNet50+SN-Block	20	90.59	87.65
Faster R-CNN+ResNet50+SN-Block+FPN	20	93.67	89.78

The detection models trained by the original Faster R-CNN network and the modified Faster R-CNN network were used to identify sugarcane seedlings in the field in the test dataset; the results are illustrated in Figure 9. It can be observed that the improved Faster R-CNN network detects smaller sugarcane seedlings and seedlings shaded by leaves more

accurately than the original Faster R-CNN network. In addition, as the color of small weeds was very similar to that of sugarcane seedlings, the distinction of these two was not very obvious, and the original Faster R-CNN would misidentify this portion of the weeds as sugarcane seedlings, whereas the improved Faster R-CNN had fewer missed detections and successfully differentiates them.

Figure 9. Comparison of sub-images test results. Subfigure (**a1**–**a3**) represent the test images, subfigure (**b1**–**b3**) represent the detection results of the test images in the original Faster R-CNN, and subfigure (**c1**–**c3**) represent the detection results of the test images in the improved Faster R-CNN. The green rectangle box represents the detection result of the original Faster R-CNN, the solid black ellipse box represents the missed detection, and the dashed black ellipse box represents the false detection. The red rectangle box represents the detection result of the improved Faster R-CNN.

4.1.2. Performance Comparison with Other Object Detection Models

The improved Faster R-CNN model was compared with the original Faster R-CNN model, and some other state-of-the-art (SOTA) models such as YOLO v2, YOLO v3, and SSD on the SC dataset. As shown in Table 3, the results indicated that the obtained detection accuracy here was at least 11.21 percent higher than that of the three commonly used networks, indicating that the method proposed in this study was more reliable for detecting and counting sugarcane seedlings in the field environment. An observation can also be made (Table 3) that the improved network based on Faster R-CNN performed the best on

the validation set, with an averaged precision of 93.67%, which was 8.69% higher than the original network, and an average recall rate close to 90%. It can be concluded that our proposed Faster R-CNN network can significantly improve the detection accuracy of sugarcane seedlings in the field and can better adapt to the changing environmental conditions such as varying lighting conditions, shading degrees, and weed densities, demonstrating the model's robust generalizability. This model with the highest performance was named SGD-D and retained for further investigation.

Table 3. Comparison with some SOTA models.

Models	AP%	AR%
YOLO v2	79.75	76.84
YOLO v3	82.46	80.33
SSD	77.50	74.21
Faster R-CNN	84.98	83.56
Improved Faster R-CNN	93.67	89.78

4.2. The Influence of IoU Threshold on Detection and Counting Results of Sugarcane Seedlings in Original Aerial Images

4.2.1. Analysis of Detecting Performance of Sugarcane Seedlings in Original Images

The IoU threshold (λ) was utilized as a de-duplication algorithm to reduce the low confidence bounding boxes with high overlap with high confidence bounding boxes. To demonstrate that the threshold affected the detection accuracy of all sugarcane seedlings in large-scale aerial images, several IoU thresholds (0.05, 0.1, 0.15, 0.2, 0.25 and 0.3) were tested to compare the detection performance using 20 test images.

The results of the selected SGN-D model with various λ parameter values (Table 4) demonstrated that the model trained with this method could accurately detect sugarcane seedlings in large-scale aerial images.

Table 4. Performance of field sugarcane seedling population counting method using different IOU thresholds.

IOU Thresholds	Precision (%)	Recall (%)	F1-score (%)
λ = 0.05	95.32	86.93	90.93
λ = 0.10	93.14	92.76	92.95
λ = 0.15	92.70	94.65	93.66
λ = 0.20	91.34	95.20	93.23
λ = 0.25	89.93	96.24	92.98
λ = 0.30	88.02	96.66	92.14

In the extreme case of λ = 0, the model discarded all overlapping detection results and only retained disjoint detection outcomes. This led to an underestimation, as some sugarcane seedlings in close proximity would have overlapping portions. When λ approached 0, the number of false-negative detections grew, as shown by the decline in recall and F1 scores. As the value of λ increased, the portion of the intersecting detection results would be progressively preserved, and therefore, the number of false negative detection. Using λ = 0.1 and 0.15 as examples, the F1 score of sugarcane seedling detection reached 92%, and the recall rate was around 5 to 7% higher than when λ = 0.05, indicating that increasing λ could produce more precise detection results. However, when λ progressed around 0.2 and 0.3, the number of false-positive samples increased, leading the F1 score and detection accuracy of sugarcane seedlings to drop. When λ = 0.3, there was no discernible increase in recall rate, and the precision fell below 90%. Particularly when λ = 1, all detection results were retained. In this instance, the NMS mechanism lost the ability to reduce duplication, resulting in a surge in the false positive rate and thereby, a drop in accuracy and F1 scores. On the basis of the aforementioned experimental findings, it can be concluded that setting the IoU threshold appropriately can improve the performance of the

splicing and de-duplication algorithm of large-scale aerial image detection results, as well as the accuracy of field sugarcane seedling population identification. Figure 10 illustrates the actual detection results with a λ setting of 0.15.
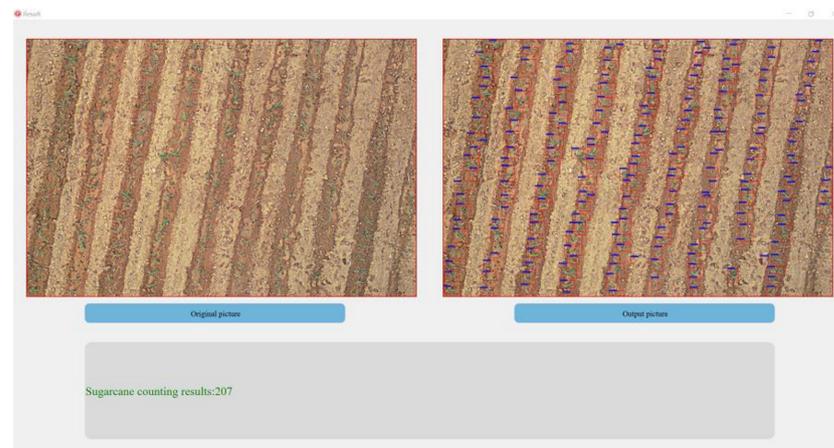

Figure 10. The software system's graph of counting results with the original image displayed on the left, the detection results displayed on the right, and the number of cane seedlings displayed below.

4.2.2. Analysis of Counting Performance

To verify the counting accuracies obtained using different values of λ, correlations were made between the number of sugarcane seedlings obtained by manual counting and the model predictions with no de-duplication algorithm ($\lambda = 0$) and six different λ values ($\lambda = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3$) (Figure 11). As shown in Figure 11a, the correlation between the number of predictions without de-duplication and manual counting was extremely low, deviating significantly from the line y = x, indicating that the prediction result obtained using the approach without de-duplication was untrustworthy. All trend lines moved closer to the axis y = x after using the de-duplication technique in Figure 11b. The values of R^2 ranged from 0.9687 and 0.9905, strongly suggesting that the method described here is effective for counting sugarcane seedlings in the field.
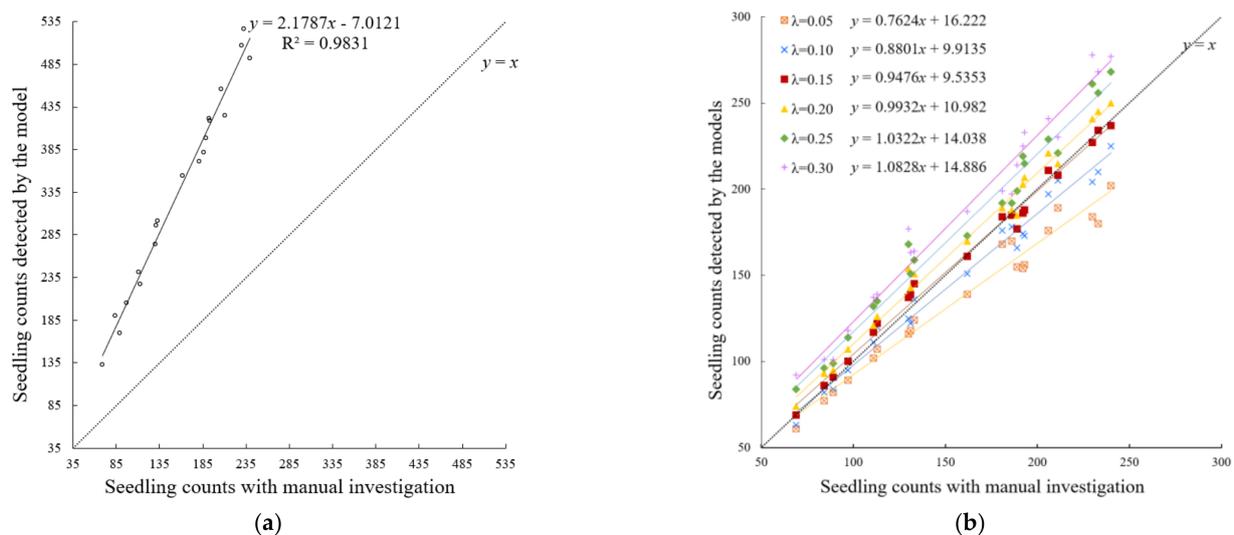

Figure 11. Correlation between the seedling counts predicted by the models and manual investigation. X axis represents the ground truth, whereas y-axis represents the count result anticipated by the matching models. (**a**) Correlation of the model without de-duplication. (**b**) Correlation of the models with de-duplication.

When using λ of 0.15, best prediction results could be achieved with a higher accuracy (96.83%) and a smaller MAE (4.60), indicating the applicability of the model predicting the number of sugarcane seedlings in large-scale field operation (Table 5).

Table 5. Results when the model uses different IoU thresholds for detection.

IOU Thresholds	CA (%)	MAE	R^2
$\lambda = 0.05$	87.63	21.55	0.9687
$\lambda = 0.10$	94.08	10.15	0.9824
$\lambda = 0.15$	96.83	4.60	0.9905
$\lambda = 0.20$	92.73	10.30	0.9880
$\lambda = 0.25$	86.75	19.15	0.9784
$\lambda = 0.30$	81.06	28.05	0.9746

5. Conclusions

In this study, an improved Faster R-CNN-based object detection model for sugarcane seedlings SGN-D was described. The results demonstrated that the improved Faster R-CNN has an average detection accuracy of 93.67%, 8.69% higher than the original network training model. The improved Faster R-CNN outperformed other commonly used object identification algorithms, including YOLO v2, YOLO v3, and SSD, and can improve the detection accuracy of small-sized sugarcane seedlings and shielded seedlings in the field. In addition, a Stitching and de-duplication algorithm-based NMS has been proposed for the detection and counting of sugarcane seedlings in large-scale aerial images. The experimental results demonstrated that using the IoU threshold of 0.15, the detection could achieve the best performance with an average detection accuracy of 93.66% and counting accuracy of 96.83%. The average absolute error MAE of 4.60 indicated also a high positive correlation with the ground truth.

Despite the fact that the developed system is able to conduct autonomous identification of sugarcane seedlings with high accuracy, improvement could be found on the further differentiation and distinction between weeds and sugarcane seedlings for a better labeling outcome.

References

1. Li, Y.R.; Song, X.P.; Wu, J.M.; Li, C.N.; Liang, Q.; Liu, X.H.; Yang, L.T. Sugar industry and improved sugarcane farming technologies in China. *Sugar Tech* **2016**, *18*, 603–611. [CrossRef]
2. Zhang, M.; Govindaraju, M. Sugarcane production in China. In *Sugarcane-Technology and Research*; IntechOpen: London, UK, 2018; p. 49. [CrossRef]
3. Elsharif, A.A.; Abu-Naser, S.S. An Expert System for Diagnosing Sugarcane Diseases. *Int. J. Acad. Eng. Res. (IJAER)* **2019**, *3*, 19–27. Available online: https://ssrn.com/abstract=3369014 (accessed on 23 November 2020).

4. Flack-Prain, S.; Shi, L.; Zhu, P.; da Rocha, H.R.; Cabral, O.; Hu, S.; Williams, M. The impact of climate change and climate extremes on sugarcane production. *GCB Bioenergy* **2021**, *13*, 408–424. [CrossRef]

5. Bhatt, R. Resources management for sustainable sugarcane production. In *Resources Use Efficiency in Agriculture*; Springer: Singapore, 2020; pp. 647–693. [CrossRef]

6. Linnenluecke, M.K.; Nucifora, N.; Thompson, N. Implications of climate change for the sugarcane industry. *Wiley Interdiscip. Rev. Clim. Chang.* **2018**, *9*, e498. [CrossRef]

7. Stein, M.; Bargoti, S.; Underwood, J. Image based mango fruit detection, localisation and yield estimation using multiple view geometry. *Sensors* **2016**, *16*, 1915. [CrossRef]

8. Velusamy, P.; Rajendran, S.; Mahendran, R.K.; Naseer, S.; Shafiq, M.; Choi, J.G. Unmanned Aerial Vehicles (UAV) in precision agriculture: Applications and challenges. *Energies* **2021**, *15*, 217. [CrossRef]

9. Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A review on UAV-based applications for precision agriculture. *Information* **2019**, *10*, 349. [CrossRef]

10. Radoglou-Grammatikis, P.; Sarigiannidis, P.; Lagkas, T.; Moscholios, I. A compilation of UAV applications for precision agriculture. *Comput. Netw.* **2020**, *172*, 107148. [CrossRef]

11. Pathan, M.; Patel, N.; Yagnik, H.; Shah, M. Artificial cognition for applications in smart agriculture: A comprehensive review. *Artif. Intell. Agric.* **2020**, *4*, 81–95. [CrossRef]

12. Ponti, M.; Chaves, A.A.; Jorge, F.R.; Costa, G.B.; Colturato, A.; Branco, K.R. Precision agriculture: Using low-cost systems to acquire low-altitude images. *IEEE Comput. Graph. Appl.* **2016**, *36*, 14–20. [CrossRef]

13. Montibeller, M.; da Silveira, H.L.F.; Sanches, I.D.A.; Körting, T.S.; Fonseca, L.M.G. Identification of gaps in sugarcane plantations using UAV images. In Proceedings of the Brazilian Symposium on Remote Sensing, Santos, Brazil, 28–31 May 2017.

14. Sanches, G.M.; Duft, D.G.; Kölln, O.T.; Luciano, A.C.D.S.; De Castro, S.G.Q.; Okuno, F.M.; Franco, H.C.J. The potential for RGB images obtained using unmanned aerial vehicle to assess and predict yield in sugarcane fields. *Int. J. Remote Sens.* **2018**, *39*, 5402–5414. [CrossRef]

15. Yu, Z.; Cao, Z.; Wu, X.; Bai, X.; Qin, Y.; Zhuo, W.; 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]

16. Liu, T.; Wu, W.; Chen, W.; Sun, C.; Zhu, X.; Guo, W. Automated image-processing for counting seedlings in a wheat field. *Precis. Agric.* **2016**, *17*, 392–406. [CrossRef]

17. Zhao, B.; Zhang, J.; Yang, C.; Zhou, G.; Ding, Y.; Shi, Y.; Liao, Q. Rapeseed seedling stand counting and seeding performance evaluation at two early growth stages based on unmanned aerial vehicle imagery. *Front. Plant Sci.* **2018**, *9*, 1362. [CrossRef] [PubMed]

18. Xia, L.; Zhang, R.; Chen, L.; Huang, Y.; Xu, G.; Wen, Y.; Yi, T. Monitor cotton budding using SVM and UAV images. *Appl. Sci.* **2019**, *9*, 4312. [CrossRef]

19. Li, B.; Xu, X.; Han, J.; Zhang, L.; Bian, C.; Jin, L.; Liu, J. The estimation of crop emergence in potatoes by UAV RGB imagery. *Plant Methods* **2019**, *15*, 1–13. [CrossRef]

20. Banerjee, B.P.; Sharma, V.; Spangenberg, G.; Kant, S. Machine learning regression analysis for estimation of crop emergence using multispectral UAV imagery. *Remote Sens.* **2021**, *13*, 2918. [CrossRef]

21. Saleem, M.H.; Potgieter, J.; Arif, K.M. Automation in agriculture by machine and deep learning techniques: A review of recent developments. *Precis. Agric.* **2021**, *22*, 2053–2091. [CrossRef]

22. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

23. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]

24. Patrício, D.I.; Rieder, R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* **2018**, *153*, 69–81. [CrossRef]

25. Villaruz, J.A.; Salido, J.A.A.; Barrios, D.M.; Felizardo, R.L. Philippine indigenous plant seedlings classification using deep learning. In Proceedings of the 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Baguio City, Philippines, 29 November–2 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–5. [CrossRef]

26. Li, Z.; Li, Y.; Yang, Y.; Guo, R.; Yang, J.; Yue, J.; Wang, Y. A high-precision detection method of hydroponic lettuce seedlings status based on improved Faster RCNN. *Comput. Electron. Agric.* **2021**, *182*, 106054. [CrossRef]

27. Jiang, Y.; Li, C.; Paterson, A.H.; Robertson, J.S. DeepSeedling: Deep convolutional network and Kalman filter for plant seedling detection and counting in the field. *Plant Methods* **2019**, *15*, 1–19. [CrossRef]

28. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]

29. Quan, L.; Feng, H.; Lv, Y.; Wang, Q.; Zhang, C.; Liu, J.; Yuan, Z. Maize seedling detection under different growth stages and complex field environments based on an improved Faster R–CNN. *Biosyst. Eng.* **2019**, *184*, 1–23. [CrossRef]

30. Fromm, M.; Schubert, M.; Castilla, G.; Linke, J.; McDermid, G. Automated detection of conifer seedlings in drone imagery using convolutional neural networks. *Remote Sens.* **2019**, *11*, 2585. [CrossRef]

31. Lin, Z.; Guo, W. Cotton stand counting from unmanned aerial system imagery using mobilenet and centernet deep learning models. *Remote Sens.* **2021**, *13*, 2822. [CrossRef]

32. Oh, S.; Chang, A.; Ashapure, A.; Jung, J.; Dube, N.; Maeda, M.; Landivar, J. Plant counting of cotton from UAS imagery using deep learning-based object detection framework. *Remote Sens.* **2020**, *12*, 2981. [CrossRef]

33. Neupane, B.; Horanont, T.; Hung, N.D. Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV). *PLoS ONE* **2019**, *14*, e0223906. [CrossRef]

34. Feng, A.; Zhou, J.; Vories, E.; Sudduth, K.A. Evaluation of cotton emergence using UAV-based imagery and deep learning. *Comput. Electron. Agric.* **2020**, *177*, 105711. [CrossRef]

35. Anuar, M.M.; Halin, A.A.; Perumal, T.; Kalantar, B. Aerial imagery paddy seedlings inspection using deep learning. *Remote Sens.* **2022**, *14*, 274. [CrossRef]

36. Li, Y.R.; Yang, L.T. Sugarcane agriculture and sugar industry in China. *Sugar Tech* **2015**, *17*, 1–8. [CrossRef]

37. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L. Microsoft COCO: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 740–755. [CrossRef]

38. Zuiderveld, K. Contrast limited adaptive histogram equalization. *Graph. Gems* **1994**, *4*, 474–485. Available online: https://ci.nii.ac.jp/naid/10031105927/ (accessed on 25 September 2022).

39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]

40. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125. [CrossRef]

41. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? *arXiv* **2014**, arXiv:1411.1792.