Article

Real-Time Detection of Seedling Maize Weeds in Sustainable Agriculture

Siqi Liu, Yishu Jin, Zhiwen Ruan, Zheng Ma, Rui Gao * and Zhongbin Su *

Institutions of Electrical and Information, Northeast Agricultural University, Harbin 150030, China
* Correspondence: rgao@neau.edu.cn (R.G.); suzb001@163.com (Z.S.)

Abstract: In recent years, automatic weed control has emerged as a promising alternative for reducing the amount of herbicide applied to the field, instead of conventional spraying. This method is beneficial to reduce environmental pollution and to achieve sustainable agricultural development. Achieving a rapid and accurate detection of weeds in maize seedling stage in natural environments is the key to ensuring maize yield and the development of automatic weeding machines. Based on the lightweight YOLO v4-tiny model, a maize weed detection model which combined an attention mechanism and a spatial pyramid pooling structure was proposed. To verify the effectiveness of the proposed method, five different deep-learning algorithms, including the Faster R-CNN, the SSD 300, the YOLO v3, the YOLO v3-tiny, and the YOLO v4-tiny, were compared to the proposed method. The comparative results showed that the mAP (Mean Average Precision) of maize seedlings and its associated weed detection using the proposed method was 86.69%; the detection speed was 57.33 f/s; and the model size was 34.08 MB. Furthermore, the detection performance of weeds under different weather conditions was discussed. The results indicated that the proposed method had strong robustness to the changes in weather, and it was feasible to apply the proposed method for the real-time and accurate detection of weeds.

Keywords: computer vision; attentional mechanism; weeds detection; sustainable agriculture

1. Introduction

With the development of animal husbandry and maize processing industry, and the application of ethanol gasoline using maize as raw material, maize has become an important food crop, fodder crop, and industrial raw material [1], occupying an increasingly important position in agricultural production. The level of maize output and its economic benefits directly affect food security and the development of agricultural production. Weeds are one of the biggest threats to the growth and yield of such crops as maize. They compete with crops for nutrients, sunlight, space, and water [2]. Weed prevention, control and removal are still based on traditional full-drench spraying pesticides [3], but current agricultural spraying practices have been considered unsustainable. This method of evenly spraying herbicides cannot distinguish between crops and weeds [4,5], so it will result in considerable pesticide waste. Moreover, this method is costly, causes soil and water source pollution [6], and affects farmland productivity and crop growth [7,8].

To reduce environmental pollution from chemicals, the European Commission is advocating strategies, such as the Integrated Pest Management (IPM). One of the principle aims of IPM is that chemicals applied will be as specific as possible to the target. The realization of this method is beneficial to the sustainable development of agriculture. The current research trend is to use smart equipment for the precise application of pesticides or intelligent mechanical weeding to reduce pesticide waste and improve the utilization rate of pesticides. Computer vision is the core technology for the precise and rapid identification and positioning of weeds [9], and is also the technical prerequisite for the precise control and...
management of weeds in the field. Therefore, intelligent equipment for precise pesticide application or intelligent mechanical weeding has broad application prospects.

Traditional target detection methods primarily start with image features obtained using machine vision and employ the differences in color, shape, or texture of crops, weeds, and even farmland backgrounds to perform image processing and manually extract features to recognize weed targets. Alchanatis et al. [10] selected two spectral channels from 100 channels of a hyperspectral sensor for soil–crop segmentation based on the acousto-optic tunable hyperspectral sensor and detection algorithm. Based on the texture features, weed detection of cotton plants was performed on the segmented image. The test results revealed that the accuracy rate could reach 86.00% under mixed conditions. Raja et al. [11] proposed a topical marker-based crop signaling technique by applying signaling compounds to lettuce seedlings prior to transplanting, and then using machine vision to recognize crop signals to distinguish lettuce from weeds. Rani et al. [12] proposed an intelligent weed detection system for sustainable agriculture. First, the features of the weeds and crops were extracted using speeded-up robust features and histogram of gradients, then logistic regression and support vector machine (SVM) algorithms were used for the classification of weeds and crops. The test results demonstrated that the accuracy of this method could reach 83.00%. These studies have effectively improved the level of weed detection and weed recognition. However, the image feature extraction based on the traditional method is cumbersome and the requirements for the extracted features are higher, which severely restricts the improvement of the automation level of weeding in maize fields. The realization of a rapid and accurate detection of weeds in maize fields in natural environments is of great significance for developing intelligent weeding equipment.

In recent years, machine learning has had a growing impact on science, health, and sustainable development, and deep-learning technology is considered to be the future development trend due to its strong learning ability, good adaptability, and high performance. It has begun to be applied in agriculture [13], such as crop classification [14], weed segmentation [15], weed detection [16,17], and other tasks [18]. Therefore, deep-learning method is the key technology and future trend to achieve an accurate and rapid response to weed detection in maize fields.

Milioto et al. [19] fed RGB images of sugar beets and near infra-red (NIR) images into an artificial neural network and segmented the plants based on their appearance, which did not require manually defining features of the classified crops, and greatly reduced manual labeling costs. Yu et al. [20] used VGGNet and DetectNet to detect weeds of specified classes, and they achieved an extremely high accuracy rate of 99% on the weeds’ specific dataset. In order to achieve higher accuracy in the target area segmentation of complex natural environment, Dias et al. [21] used convolutional neural network (CNN) and SVM methods to automatically extract crop features in a complex background. This image segmentation method obtained relatively accurate crop region segmentation results. At present, weed detection method based on deep learning has a good effect in practical application, but there are still some problems. The research mainly focuses on three common problems: first, the appearance of crops and weeds is often quite different, and for early weed detection in farmland, the morphological characteristics of seedlings and weeds may be similar and cannot be distinguished well; secondly, the specific network only has a good result on specific weed species, but the network generalization performance is not good, and it is usually not applicable in complex natural environment; and finally, even in natural environment weed detection, the detection effect is also limited by the population density, that is, there are occlusion and overlap problems between overgrown weeds (either under the segmentation task or under the detection task). In the framework based on deep learning, how to improve the network to solve the above problems has become the beginning of research innovation.

Bakhshipour et al. [22] designed successive steps in a discrimination algorithm to determine the wavelet texture features for each image subdivision to be fed to an artificial neural network. The recognition algorithm was optimized by studying the two methods of
single-machine wavelet transform and image segmentation to distinguish weeds from main crops. The results demonstrated that the recognition accuracy of this method for sugar beet and four types of weed could reach 96.00%. However, a small mobile terminal cannot complete such a huge amount of calculation. Farjon et al. [23] used a deep-learning network based on two-stage detection and faster R-CNN algorithm to detect apple blossoms, and the average detection accuracy of this method was 68%. Faster R-CNN comprises two parts: a classification network and a region proposal network (RPN). However, the amount of calculation is large, the detection speed is not fast, and real-time detection cannot be realized. Most of the deep learning-based methods take up a lot of memory, which affects the deployment and application of these models in intelligent equipment and limits the development of sustainable agriculture. The development of agricultural intelligent equipment is an inevitable development trend of sustainable agriculture, and the detection network based on server-side applications often cannot achieve a high detection speed. Therefore, how to make lightweight modifications to the network to adapt to mobile applications has become a recent research topic.

The You Only Look Once (YOLO) [24–26] method unifies target classification and positioning into a regression problem. It is a one-stage object detection method, so it does not require an RPN. It only performs regression to detect the target, so it has faster detection speed. The most advanced version, which is named YOLO v4, has high detection accuracy, high speed, and good detection performance for small targets, and the newly improved lightweight model YOLO v4-tiny reduces the memory and improves the detection speed while ensuring high accuracy.

However, the YOLO v4-tiny model has not been applied to weed detection. After image data are passed into the convolutional neural network and after the convolution and downsampling operations, the feature map will gradually shrink, which makes it difficult for the model to extract the features of small objects. Due to the variable light conditions of maize fields in the natural environment, maize and weeds are densely distributed, and the phenomenon of overlapping leaves is also serious. Moreover, the size, shape, and cluster density of plants at different growth stages are also different. These problems have brought great challenges to detecting weeds in maize fields in the natural environment. In a complex and changeable environment, traditional detection methods are unsuitable for detecting weeds under different weather conditions.

In addition, it is challenging to balance the accuracy and real-time performance of deep-learning methods. In order to solve the problems of change in light conditions in the natural environment, the overlapping phenomenon of corn seedlings and weed leaves, and the difference in the appearance of plants at different growth stages, the effect of the lightweight model for weed detection is poor. In this research object, based on the lightweight YOLO v4-tiny model, a corn weed detection model combining an attention mechanism and a spatial pyramid hybrid structure is proposed. This model is used to train, test, and verify the collected data, and the detection results are compared to other mainstream target detection models. The purpose is to enable the improved model to be applied to the detection of corn seedlings and weeds in the natural environment, and to promote the application of deep-learning methods in the detection of farmland targets so as to provide an accurate identification method for weeds in unmanned agricultural machinery field operations.

2. Materials and Methods

The workflow of this paper was as follows: Firstly, data images of corn and weeds acquired by the camera were used, and the data were manually labeled. These data images were randomly divided into a training set, a validation set, and a test set with a ratio of 3:1:1, and the training images were preprocessed. Then, the training set was input into different, improved YOLO v4-tiny network models for training, and the optimal weights of the network models with different structures were obtained after training. Finally, the test set was used to test the network models with different structures. The model with
the best detection effect was selected, and the detection effects of other mainstream object detection models were compared. The content of this section is organized as follows. Section 2.1 introduces the collection of the image data, including the process of collecting images and tapes of weeds. Section 2.2 describes the preprocessing of the image datasets. Section 2.3 introduces the process of labeling objects. Section 2.4 introduces the YOLO v4-tiny algorithm. Section 2.5 introduces the improved YOLO v4-tiny algorithm. Section 2.6 introduces the structure of the improved YOLO v4-tiny model. Afterwards, the loss function used by the improved YOLO v4-tiny model is introduced. Section 2.7 introduces the loss function of the algorithm proposed in this paper. Section 2.8 introduces the model training in detail.

2.1. Image Data Acquisition

The experimental data were collected in an experimental maize field of the Northeast Agricultural University in Harbin, Heilongjiang Province, China, from May to July 2021. The name of the maize variety is Dongnong275 (the breeder is the Northeast Agricultural University, and the variety source is DN2710 × DN4206). The targets in this study were grown in natural conditions without human intervention, and the soil was also natural. The data collection was divided into 10 stages (with a difference of 3 to 5 days in each stage). The camera used for data acquisition was a 12-megapixel resolution color camera (the model is Sony IMX503, the unit pixel area is 1.4 micron, and the focal length is 26 mm). The camera was fixed on a horizontal telescopic pole of an uncrewed vehicle at 50 cm away from the ground, and the camera lens was perpendicular to the ground for data capturing. The image resolution was 3024 pixels (horizontal) by 4032 pixels (vertical), and all images were saved in JPG format. The image data were taken on sunny, cloudy, and rainy days. The collection times were 10 a.m. and 2 p.m., with 50 pictures taken in each period. The Latin names and abbreviation of the weeds that were photographed are Abutilon theophrasti Medicus (Abutilon), Chenopodium album Linn (Chenopedium), Amaranthus blitum Linnaeus (Amaranthus), and Ipomoea purpurea Lam. seedlings (Morning glory seedlings). The data collection location, the appearance characteristics of the four weeds, the uncrewed vehicle, and the data collection process are illustrated in Figure 1.

Figure 1. Data and the schematic diagram of the acquisition process.
2.2. Data Analysis and Preprocessing

In the collected images, the growth status of the weeds included weeds only, maize seedlings and weeds, densely distributed weeds, sparse weed distribution, and so on. A total of 1000 effective images were taken during this period. The growth and distribution of maize seedlings and weeds in the experimental field are presented in Figure 2.

![Figure 2](image)

Figure 2. Growth and distribution of maize seedlings and weeds: (a) maize seedlings are accompanied by weeds and the weeds are sparse; (b) maize seedlings are accompanied by weeds and the weeds are dense; and (c) weeds only.

The data set was divided into a training set, a validation set, and a testing set with a ratio of 3:1:1. Rotation preprocessing was performed on the images of the training set, and the training set was expanded to enhance the richness of the experimental data set, as depicted in Figure 3. The expanded training data set has 2400 images, and the number of expanded types of data is listed in Table 1.

![Figure 3](image)

Figure 3. Image augmentation methods: (a) original image; (b) 90° clockwise rotation; (c) 180° clockwise rotation; and (d) 270° clockwise rotation.

| Type                  | Maize | Abutilon | Chenopodium | Amaranth | Morning glory Seedlings |
|-----------------------|-------|----------|-------------|----------|------------------------|
| Number                | 3836  | 5887     | 7110        | 3786     | 3113                   |

2.3. Data Tag

Accurately labeling the target of the detection task can enable the weed detection model to better learn the features of the target. The model targeted maize seedlings and various weeds in a test maize field, and the targets were marked with a minimum bounding box on each image. Each minimum bounding box should only contain a maize seedling or a weed and should minimize background pixels. An example of image tagging is illustrated in Figure 4, Different objects are marked with boxes of different colors in the figure.
2.3. Data Tag
Accurately labeling the target of the detection model is important to ensure its detection performance. The model targeted maize seedlings and weeds, and other data. As shown in Figure 4b, the numbers in the first red box represent its class, and the numbers in the second red box represent the center coordinates of the box and the width and height of the box, respectively. Objects with incomplete contours or unclear features were also marked with minimum bounding boxes to ensure the detection performance of the model.

2.4. YOLO v4-Tiny Model
The YOLO v4-tiny is a lightweight target detection model designed on the basis of the YOLO v4. It occupies less memory and runs faster, which greatly increases the possibility of the realization of intelligent weeding equipment. The backbone network of the YOLO v4-tiny is CSPDarknet53-tiny, which is a lightweight version of the CSPDarknet53 network. The network structure of CSPDarknet53-tiny is shown in Figure 5. The CSPDarknet53-tiny network adopts the CSPBlock module in the cross-stage partial network instead of the residual block module (Resblock) module in the residual network. The CSPBlock module divides the feature map into two parts and combines the two parts through the cross-stage residual edge, allowing the gradient flow to propagate on two different network paths and increasing the relative difference in gradient information. Compared to the residual block module, the learning ability of the CSPBlock module is stronger. Although this increases the amount of computation, it improves feature extraction ability and detection effect. This network eliminates the computational bottleneck of the CSPBlock module, which has considerable calculation, and reduces the calculation. The accuracy of the YOLO v4-tiny method is improved, whereas the amount of calculation remains unchanged or even reduced. To further simplify the calculation process and increase the calculation speed of the network, the YOLOv 4-tiny method uses a leaky rectified linear unit (Leaky ReLU) function in the CSPDarknet53-tiny network, instead of the Mish function in the YOLO v4, as the activation function.

Figure 4. Example of image tagging: (a) is the labeling result of maize and weeds, and (b) is the ID and coordinate information for the marked targets in (a).
In terms of feature fusion, the YOLO v4-tiny model uses a feature pyramid network (FPN) structure to extract feature maps of different scales, perform feature fusion on the two effective feature layers, and improve target detection speed. Moreover, the YOLO v4-tiny uses two feature graphs with different proportions, $13 \times 13$ and $26 \times 26$, to predict the detection results.

2.5. YOLO v4-Weeds

2.5.1. Improvement of the Feature Extraction Network

The accuracy of weed detection is often interfered by various factors, such as light and occlusion, so it is critical to overcome these interferences. Given erroneous detection, missed detection, and low confidence of weed detection caused by partial occlusion and the influence of light on the YOLO v4-tiny model, the detection effect can be effectively improved by enhancing the receptive field and enhancing the feature extraction method.

The SE-Block [27] was designed based on the above ideas. As depicted in Figure 6, the SE-Block obtains a weight for each channel through learning. According to the weights, the important channel features are strengthened, and the non-important channel features are suppressed to improve the network model accuracy.

Figure 5. Yolov4-tiny network structure diagram.

Figure 6. SE-Block.
Moreover, the SE-Block primarily comprises the squeeze and excitation operations. The excitation operation generates a weight for each feature channel, different colors in Figure 6 represent different values, which are used to measure the importance of channels. For the squeeze operation, global average pooling of the input feature maps is performed. Each two-dimensional feature channel in the spatial dimension becomes a real number. The global distribution characteristics of the characteristic channels are obtained, providing a global receptive field. The dimensionality of the output matches the number of characteristic channels of the input.

Finally, the scale operation \( f_{\text{scale}} \) multiplies the weight generated by the previous corresponding channel feature, and recalibrates the original features in the channel dimension to strengthen effective features and weaken inefficient or ineffective features. The extracted features have stronger directivity, which improves the detection results. In Figure 6, \( X \) is the input feature map, and \( X' \) is the output feature map. In addition, \( W, H, \) and \( C \) denote the length, width, and channel number of the feature map, respectively.

This paper combines the Resblock and SE-Block to form a lightweight, feature-extraction basic Res-SE structure to ensure the front-end basic network has high accuracy while quickly extracting features, as displayed in Figure 7. Two Resblock structures were added at the end of the feature extraction network. Each Resblock consists of \( 1 \times 1 \) and \( 3 \times 3 \) convolutional layers, and the activation function is Mish. In addition, the SE-Block was added to the second Resblock to obtain a series of new feature outputs. The SE-Block acts as a bypass unit for output characteristics. After a series of operations, such as global average pooling, through the two fully connected layers, ReLu and sigmoid normalization, the weight of each characteristic channel is obtained. Then, the weight is multiplied by each channel element corresponding to the original feature information to complete the channel information intensity recalibration.

![Figure 7. Res-SE structure diagram.](image)

2.5.2. Dense-SPP Structure

In most CNN architectures, the fully connected layers are usually connected after the convolutional layers, so it can only accept a fixed size as input. Thus, the input image is often cropped and zoomed to satisfy the input requirements. However, these operations cause a loss of information and distort the image, resulting in poor detection results. He et al. [28] designed SPP, a multiscale feature fusion model, to solve the problem of image distortion caused by different input image scales in the computer vision field. The SPP structure uses multilevel spatial windows to extract features of different scales from the same feature map and perform pooling operations, producing the same scale output for input images of any size.

Drawing lessons from the idea of SPP, this paper proposes a brand new dense SPP structure, where the multiscale SPP module has four branches, three of them are max pooling layer and convolutional layer with different window sizes, and the other branch is a jump connection. The sizes of the largest pooling layer windows in the three branches are \( 13 \times 13 \), \( 9 \times 9 \), and \( 5 \times 5 \), respectively, with a step size of 1 and passing through a \( 1 \times 1 \) convolutional layer. Given the design idea of feature reuse, the feature map output by the branch of the \( 5 \times 5 \) maximum pooling window is spliced with the feature map before pooling and being concatenated with the \( 9 \times 9 \) pooling window, and the feature map output through the \( 9 \times 9 \) window performs the same operation before entering the \( 13 \times 13 \)
maximum pooling window, as shown by the arrows between the branches in Figure 8. Finally, the feature map output of each branch are also spliced as the output of the module.

Figure 8. Dense-SPP structure diagram.

The SPP module combines the dense connection method [29], which increases the extracted multiscale local feature information and integrates it into subsequent global features to obtain richer feature representations and promote feature reuse. The number of parameters is reduced, and the detection accuracy is ultimately improved.

2.5.3. Spatial Attention Module

The spatial attention module (SAM) [30], as illustrated in Figure 9, performs global average pooling and global maximum pooling operations in the channel dimension to generate two feature maps representing different information. It performs feature fusion operations on the feature maps, reducing the dimensionality through a 7 × 7 convolutional layer with a larger receptive field. Finally, the sigmoid operation generates a weight map and superimposes the original input feature map to enhance the target area, as indicated in Equation (1), where $\sigma$ represents the sigmoid operation and $F$ is the input feature map.

$$M_S(F) = \sigma\left(f^7 \times 7([\text{AvgPool}(F); \text{MaxPool}(F)])\right)$$  \hspace{1cm} (1)

Figure 9. Spatial attention module structure.

In this paper, the residual block [31] is combined with the SAM to form the Res-SAM structure, as depicted in Figure 10. A Resblock structure was added after the dense SPP structure, comprising two 1 × 1 convolutional layers, and the activation function is the Leaky ReLU. After adding the SAM module to the second convolutional layer, a series of new feature outputs were obtained.
Figure 10. Res-SAM structure diagram.

2.6. YOLO v4-Weeds Network Structure

The complete network structure of the YOLO v4-weeds is presented in Figure 11. The YOLO v4-weeds network uses 608 × 608 × 3 as the input image size to improve the model’s ability to handle high-resolution images. First, the backbone feature extraction network uses two 3 × 3 convolutional layers with a stride of 2 to extract global features and then performs down-sampling through the CSPBlock in the network. The resulting output feature map is added to the more efficient output feature map extracted by the channel attention, and the combined feature is used as the output feature of the backbone feature extraction network. Then, two effective feature layers from the backbone are selected and added into the feature fusion module. The feature fusion module consists of the dense SPP structure, the Res-SAM, and the FPN structure. In this way, a richer feature representation can be obtained, and feature reuse can be promoted. Two enhanced features can be obtained using the feature fusion module: 38 × 38 feature maps for small target detection and 19 × 19 feature maps for large target detection. Finally, these two features are passed to the YOLO head to obtain prediction result, and the output prediction result contains the labels of different kinds of weeds and maize seedlings.

![Network structure of the YOLO v4-weeds.](image)
2.7. Loss Function

The loss function of the YOLO v4-weeds consists of bounding box location loss ($\text{Loss}_{\text{ClIoU}}$), confidence loss ($\text{Loss}_{\text{confidence}}$), and classification loss ($\text{Loss}_{\text{class}}$), as indicated in Equations (2)–(9). In addition, the YOLO v4-weeds uses ClIoU loss as the bounding box regression loss function to judge the distance between the predicted box (PB) and the ground truth (GT). Moreover, ClIoU addresses the situation in which PB or GT is completely surrounded by the other party, the convergence speed is too slow, and the accuracy of the predicted box is low. The overlap area, the center point distance, the length and width ratio of PB and GT should be considered simultaneously, as noted in Equations (2) and (3):

$$\text{ClIoU} = \text{IoU} - \frac{b^2 (b_{\text{gt}}^2)}{c^2} - \alpha v,$$  \hspace{1cm} (2)

$$\text{Loss}_{\text{ClIoU}} = 1 - \text{ClIoU},$$  \hspace{1cm} (3)

where $b$ and $b_{\text{gt}}$ represent the center points of PB and GT; $\rho^2$ finds the Euclidean distance; $c$ denotes the shortest diagonal length of the smallest bounding box of PB and GT; $\alpha$ represents a positive balance parameter; and $v$ denotes the consistency of the aspect ratio of PB and GT. The definitions of $v$ and $\alpha$ are given in Equations (4) and (5), where $w_{\text{gt}}$ and $h_{\text{gt}}$ represent the width and height of GT and PB, respectively:

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w_{\text{gt}}}{h_{\text{gt}}} - \arctan \frac{w}{h} \right)^2,$$  \hspace{1cm} (4)

$$\alpha = \frac{v}{(1 - \text{IoU}) + v},$$  \hspace{1cm} (5)

$$\text{Loss}_{\text{confidence}} = \sum_i \sum_{j=0}^{B} K \left[ -\log(p) + \text{BCE} \left( \hat{n}, n \right) \right],$$  \hspace{1cm} (6)

$$\text{Loss}_{\text{class}} = \sum_i \sum_{j=0}^{B} 1_{i,j}^{\text{noobj}} \left[ -\log(1 - p_c) \right],$$  \hspace{1cm} (7)

$$\text{BCE} \left( \hat{n}, n \right) = -n \log(n) - (1 - \hat{n}) \log(1 - n),$$ \hspace{1cm} (8)

$$K = 1_{i,j}^{\text{obj}},$$ \hspace{1cm} (9)

$$\text{Loss} = \text{Loss}_{\text{ClIoU}} + \text{Loss}_{\text{confidence}} + \text{Loss}_{\text{class}}.$$  \hspace{1cm} (10)

In Equations (5) and (6), $S$ is the number of grids, and $B$ is the anchor number corresponding to each grid. In Equation (8), $K$ represents the weight, and its value is 1 if there is an object in the $j$th anchor of the $i$th grid; otherwise, its value is 0. In addition, $n$ represents the actual and predicted classes of the $j$th anchor in the $i$th grid, respectively, and $p$ is the probability that the object is a weed or maize.

2.8. Model Training Detail

The YOLO v4-weeds detection model used in this research was modified using the Darknet framework. In model training and testing, the software environment was Ubuntu 20.04 (Ubuntu is a product of Canonical Corporation, Ubuntu is a Debian Linux-based operating system founded by Mark Shuttleworth, Canonical Corporation is registered in the Isle of Man) and Python 3.8.5 (The author of Python is Guido van Rossum, and the copyright of Python currently belongs to the Python Software Foundation). Table 2 lists the training platform parameters of the improved lightweight weed detection model, YOLO v4-weeds.
Table 2. Training platform parameters.

| Processor         | Basic Frequency | Memory | Graphics Card    | Video Memory |
|-------------------|-----------------|--------|------------------|--------------|
| Intel i9 12900K   | 3733 MHz        | 64 GB  | RTX3090 (GPU)    | 24 GB        |

The model initialization parameters are presented in Table 3. Considering the server’s memory constraints, this article sets the batch size to 16. In order to better analyze the training process, the number of iterations was chosen to be 30,000. For parameters such as momentum, initial learning rate, and weight decay regularization, these refer to the original parameters in the YOLO v4-tiny model, and the model is trained after defining the training parameters. The learning rate drops to 0.0001 after 20,000 steps and 0.00001 after 25,000 steps. Then, the image data in the test set are passed into the trained YOLO v4-weeds model for testing to verify the performance of the model.

Table 3. Network initialization parameters.

| Size of Input Images | Batch | Momentum | Decay | Initial Learning Rate | Iteration Number |
|----------------------|-------|----------|-------|-----------------------|------------------|
| 608 × 608            | 16    | 0.9      | 0.0005| 0.001                 | 30,000           |

3. Results and Analysis

3.1. Evaluation Indicators and Results

For binary classification problems, the samples can be divided into four types: true positive (TP), false positive (FP), true negative (TN), and false negative (FN), according to the combinations of the true and predicted classes of the learner. The confusion matrix of classification results is presented in Table 4. This research uses four indicators for evaluation to verify the model performance: precision, recall, mean average precision (mAP), and detection speed. When the intersection over union (IoU) \( \geq 0.5 \), it is a TP case. When the IoU < 0.5, it is a FP case, and when the IoU = 0, it is a FN case.

Table 4. Confusion matrix of classification results.

| Labeled | Predicted | Confusion Matrix |
|---------|-----------|-----------------|
| Positive| Positive  | TP              |
| Positive| Negative  | FN              |
| Negative| Positive  | FP              |
| Negative| Negative  | TN              |

The \( mAP \) is the mean value of average accuracy when weeds and maize are detected. A higher \( mAP \) value indicates better detection results for weeds. The calculation formulas for \( IoU \), precision, recall, and \( mAP \) are presented in Equations (11)–(14). Taking the precision as the vertical axis and the recall as the horizontal axis, the precision curve, referred to as the P–R curve, can be obtained.

\[
IoU(R, R') = \frac{|R \cap R'|}{|R \cup R'|} \tag{11}
\]

\[
\text{precision} = \frac{TP}{TP + FP} \times 100\%, \tag{12}
\]

\[
\text{recall} = \frac{TP}{TP + FN} \times 100\%, \tag{13}
\]

\[
mAP = \frac{\sum_{c=1}^{C} \text{AP}(c)}{C} \tag{14}
\]

where \( R \) denotes the detection area of the target bounding box; \( R' \) is the actual area of the target bounding box; and \( C \) is the number of detection types. Because there are five detection targets in this paper, \( C = 5 \).
The training loss curve during training is displayed in Figure 12a. It can be seen from Figure 12a that, in the early stage of model training, the model learning is more efficient and the training curve converges faster. The slope of the training curve gradually decreases as the training progresses. Finally, when the number of training iterations reaches about 21,000, the loss reaches a plateau. In addition, the total time for model training is 17.2 h.

Figure 12. Experimental results: (a) is the training loss curve, and (b) is an example of the test results.

The improved lightweight weed detection model was tested to verify its detection capability. The trained model YOLO v4-weeds was used to process the testing set, and the test took 3.49 s. The results are provided in Table 5.

Table 5. Weed detection results of the YOLO v4-weeds.

| Evaluation Indices | Precision (%) | Recall (%) | Average IoU (%) | mAP (%) | Detection Speed (f/s) |
|--------------------|---------------|------------|-----------------|---------|-----------------------|
| Test results       | 76.64         | 83.33      | 61.42           | 86.69   | 57.33                 |

The precision of the model was 76.64%, the recall was 83.33%, the mAP was 86.69%, and the detection speed was 57.33 f/s. An example of the test result is shown in Figure 12b.

3.2. Comparison of Different Target Detection Algorithms

Researchers have used detection methods based on deep learning in many target detection tasks [32–35]. This study compared five target detection algorithms, including the Faster R-CNN [36], SSD 300 [37], YOLO v3, YOLO v3-tiny, and YOLO v4-tiny, to the proposed YOLO v4-weeds target detection algorithm to verify the effectiveness of the proposed method for weed detection in maize fields. During the test, the P–R curves of the different models are listed in Figure 13.

The testing set was used to evaluate the performance of the different detection algorithms. The results are provided in Table 6.

| Evaluation Indices | Precision (%) | Recall (%) | Average IoU (%) | mAP (%) | Detection Speed (f/s) |
|--------------------|---------------|------------|-----------------|---------|-----------------------|
| Test results       | 68.55         | 69.22      | 85.55           | 75.72   | 80.33                 |
|                    | 86.55         | 75.72      | 80.33           | 86.69   | 34.08                 |
|                    | 25.16 f/s     | 47.93 f/s  | 49.14 f/s       | 57.42 f/s | 57.84 f/s             |

The rapid and accurate detection of weeds in maize fields helps protect the growth of crops and provides a technical reference for developing field weeding robots. This research compared five different target detection algorithms (i.e., Faster R-CNN, SSD 300, YOLO v3, YOLO v3-tiny, and YOLO v4-tiny) with the proposed method and analyzed their weed detection performance.

The test results demonstrate that the mAP values of the six target detection algorithms are 68.55%, 69.22%, 85.55%, 75.72%, 80.33%, and 86.69%; the sizes of the models are 108.33 MB, 92.65 MB, 234.98 MB, 33.13 MB, 22.47 MB, and 34.08 MB; and the detection speeds are 25.16 f/s, 47.93 f/s, 49.14 f/s, 57.42 f/s, 57.84 f/s, and 57.33 f/s.
The improved lightweight weed detection model was tested to verify its detection capability. The trained model YOLO v4-weeds was used to process the testing set, and the test took 3.49 s. The results are provided in Table 5.

Table 5. Weed detection results of the YOLO v4-weeds.

| Evaluation Indices | Precision (%) | Recall (%) | Average IoU (%) | mAP (%) | Detection Speed (f/s) |
|--------------------|---------------|------------|-----------------|---------|----------------------|
| Test results       | 76.64         | 83.33      | 61.42           | 86.69   | 57.33                |

The precision of the model was 76.64%, the recall was 83.33%, the mAP was 86.69%, and the detection speed was 57.33 f/s. An example of the test result is shown in Figure 12b.

3.2. Comparison of Different Target Detection Algorithms

Researchers have used detection methods based on deep learning in many target detection tasks [32–35]. This study compared five target detection algorithms, including the Faster R-CNN [36], SSD 300 [37], YOLO v3, YOLO v3-tiny, and YOLO v4-tiny, to the proposed YOLO v4-weeds target detection algorithm to verify the effectiveness of the proposed method for weed detection in maize fields. During the test, the P–R curves of the different models are listed in Figure 13.

Figure 13. P–R curves for the detection models: (a) Faster R-CNN; (b) SSD 300; (c) YOLO v3; (d) YOLO v3-tiny; (e) YOLO v4-tiny; and (f) YOLO v4-weeds.

In terms of detection accuracy, the proposed method has higher detection accuracy than the other five detection algorithms. In terms of detection speed, although the YOLO v3-tiny and the YOLO v4-tiny are slightly faster than the proposed method, the detection speed of the proposed method can still meet real-time requirements. In terms of model size, the YOLO v4-weeds model is slightly larger than the YOLO v3-tiny model and the YOLO v4-tiny model, but is much smaller than the YOLO v3 model.
Table 6. Weed detection results of the different target detection algorithms.

| Algorithms     | Precision (%) | Recall (%) | mAP (%)  | Model Size (MB) | Detection Speed (fps) |
|---------------|---------------|------------|----------|-----------------|-----------------------|
| Faster R-CNN  | 64.27         | 69.42      | 68.55    | 108.33          | 25.16                 |
| SSD 300       | 72.57         | 70.98      | 69.22    | 92.65           | 47.93                 |
| YOLO v3       | 74.77         | 86.34      | 85.55    | 234.98          | 49.14                 |
| YOLO v3-tiny  | 77.93         | 63.47      | 75.72    | 33.13           | 57.42                 |
| YOLO v4-tiny  | 75.13         | 75.18      | 80.33    | 22.47           | 57.84                 |
| YOLO v4-weeds | 76.64         | 83.33      | 86.69    | 34.08           | 57.33                 |

In summary, by comparing the detection performance of the above five detection models with the proposed method on weeds in natural environments, the model proposed in this paper not only ensures high accuracy, but also reduces the size of the model and improves the detection speed. Therefore, the proposed model can achieve accurate and rapid detection of weeds in maize fields in natural environments and is easy to apply to embedded devices.

3.3. Comparison of Algorithms in Varying Weather

The model efficiency was further tested according to actual conditions to test its adaptability and effectiveness in natural scenarios. The sunny, cloudy, and rainy images taken in natural scene in the testing set were divided into two parts, each forming a testing set for detection, and were compared to the YOLO v3, YOLO v3-tiny, and YOLO v4-tiny models. Precision, recall and mAP were used to evaluate model performance. Among the images, 437 images of sunny days were denoted as Type 1, and 563 images of cloudy or rainy days were denoted as Type 2. The YOLO v3, YOLO v3-tiny, and YOLO v4-tiny models were used for comparison, and the evaluation indicators and results are presented in Table 7. The detection results of the four models maintain high accuracy, but the proposed model’s detection effect is better than the other three models in terms of the precision and mAP indicators. Table 5 reveals that the mAP of the proposed model is 1.3% and 0.4% higher than those of YOLO v3 in Type 1 and Type 2, respectively.

Table 7. Evaluation results of the weed test set of each algorithm under different weather.

| Algorithms     | Type 1 Precision (%) | Type 1 Recall (%) | Type 1 mAP (%) | Type 2 Precision (%) | Type 2 Recall (%) | Type 2 mAP (%) |
|---------------|----------------------|-------------------|----------------|----------------------|-------------------|----------------|
| YOLO v3       | 74.77                | 86.79             | 85.91          | 74.77                | 85.52             | 85.63          |
| YOLO v3-tiny  | 78.54                | 66.72             | 77.22          | 78.54                | 62.79             | 75.17          |
| YOLO v4-tiny  | 75.52                | 75.02             | 75.26          | 75.52                | 75.26             | 81.36          |
| YOLO v4-weeds | 76.61                | 83.38             | 87.21          | 76.61                | 83.30             | 86.03          |

Two images were selected for comparison between sunny and rainy days. The detection results and confidence levels of the four models under varying weather conditions are illustrated in Figure 14 and Table 8. As listed in Table 8, for sunny days, the proposed model can detect some occluded or partially missing objects while other models do not, and the proposed model has a high average confidence. The proposed model also performs better on cloudy and rainy days, although slightly worse than the YOLO v3. In summary, the proposed model can accurately and rapidly detect weeds in maize fields under different weather conditions and is more suitable for application in intelligent devices in complex natural environments.
4. Discussion

The situation in maize field in the natural environment is complex, and the interference is substantial. The mAP of the proposed model is 86.69%, and the detection speed can reach 57.33 f/s. This study compared the proposed model with other crop detection
models to more accurately verify the robustness and effectiveness of the algorithm. In addition, the main factors affecting the accuracy of weed detection in a natural environment were analyzed. The comparative analysis of some weed detection algorithms is shown in Table 9. Comparative studies are based on object type, detection methods, and detection speed and accuracy. The model structure, especially the depth of the feature extraction network, greatly influences weed detection results. The reason is that a deeper model has better nonlinear expression ability. Furthermore, deeper models can learn more complex transformations and can fit more complex feature inputs [38].

Table 9. Comparison of various weed detection techniques.

| Study                | Object                  | Detection and Classification | Method    | Speed (f/ms)         | Accuracy (%) |
|----------------------|-------------------------|-----------------------------|-----------|----------------------|--------------|
| Ying et al. [39]     | Carrot and weeds        |                             | CNNs      | $0.79 \times 10^{-1}$ | 88.46        |
| McCool et al. [40]   | Carrot and weeds        |                             | DCNNs     | $0.12 \times 10^{-3}$ | 93.9         |
| Wu et al. [41]       | Apple blossoms          |                             | CNNs      | $0.72 \times 10^{-1}$ | 97.31        |
| Lin et al. [42]      | Maize and weeds         |                             | CNNs      | n/a                  | 95           |
| Milioto et al. [19]  | Sugar beets and weeds   |                             | CNNs      | $50 \pm 10$ ms per blob | 95           |
| Tian et al. [43]     | Apple                   |                             | CNNs      | $0.33 \times 10^{-2}$ | n/a          |
| Proposed method      | Maize and weeds         |                             | CNNs      | $0.57 \times 10^{-1}$ | 86.69        |

4.1. Analysis of Model Structure Influence on Weed Detection Results

Ying et al. [39] replaced the backbone network of the YOLO v4 with MobileNetV3-Small, combined depth separable convolutional and inverted residual structures, and introduced a lightweight attention mechanism to detect carrot seedlings and weeds. The model accuracy in the two testing sets with only carrot seedling images and images including carrots and weeds can reach 89.11% and 87.80%, respectively. The average accuracy is 1.77% higher than that of the proposed model, but the model size is 124.92 MB larger than that of the proposed model. The MobileNetV3-Small backbone network used by Ying et al. consists of 11 bneck structures. The bneck structure is composed of four parts: Firstly, the dimension is enhanced by $1 \times 1$ convolution and then by $3 \times 3$ depth-separable convolution. It then passes through the SE-Block and, finally, $1 \times 1$ convolution is used to reduce the dimension. The use of bneck structure can reduce the computational burden of the model, enhance the attention of key channels, and improve the accuracy of detection.

McCool et al. [40] applied a DCNN (deep convolutional neural network) to classify crops and weeds in a carrot dataset and achieved an accuracy of 93.9%. However, the speed is $0.12 \times 10^{-3}$ f/ms, and it is not applicable for real-time application. The network used by McCool et al. is a new version of GoogleLeNet, Inception-v3. The network uses nine Inception modular structures with a total of 22 floors. The Inception structure consists of multiple $1 \times 1$ and $3 \times 3$ convolution layers combined with max pooling. At the end of the network, the average pooling is adopted to replace the fully connected layer, and, in order to avoid gradient disappearance, the network adds two additional SoftMax for forward conduction gradient.

Compared to the two feature extraction networks introduced above, the feature extraction network proposed in this paper has a simpler structure. Firstly, the backbone feature extraction network uses two $3 \times 3$ convolutional layers, each with a step size of 2, to extract global features, which then pass through the CSPBlock in the network to downsample. In addition, the SE-Block is used at the end of the backbone network to enhance the attention of key channels. The feature extraction network proposed in this paper has shallower layers. Therefore, the feature extraction ability is slightly insufficient, which may be the reason for the low accuracy of the model. As shown in Figure 14i, for the test set, weeds with smaller size and incomplete shape have lower confidence or missed detection.
4.2. Analysis of Influence of Natural Environmental Factors on Weed Detection Results

Natural environmental factors have a great influence on weed detection results. On the one hand, insufficient light will affect the image quality of the data. On the other hand, the appearance of different types of weeds is similar, and the overlap and occlusion between weeds are more serious, so it is difficult to obtain a clear and complete target image, resulting in difficulty in feature extraction.

Wu et al. [41] used the channel pruning-based YOLO v4 model to detect apple blossoms. Wu stated that missed detection was because the mobile phone that automatically took the image blurred the background and the details in the center, making it difficult for the detection model to extract target features. In addition, the illumination also affected data collection. Lin et al. [42] proposed a decision tree-based machine learning technique based on hyperspectral images to classify maize and weeds. They achieved an accuracy of 95%. However, this algorithm had not been tested in areas with dense weeds and could not confirm that the algorithm could be applied to areas with dense weeds. Milioto et al. [19] fed RGB images of sugar beets and near infra-red (NIR) images into an artificial neural network and segmented the plants based on their appearance. Although the classification accuracy achieved was 95%, it took 50 ms $\pm$ 10 ms per blob in the image, which is very high compared to previous methods.

Tian et al. [43] used DenseNet combined with the YOLO v3 model to detect apples. He believed that the main aspect affecting the detection effect was that partial occlusion by branches and leaves and overlaps between apples often occur in orchards. These factors have a certain impact on the test results. The missed detection in this paper is presented in Figure 14. For sunny days with sufficient light, the color and texture characteristics of the plants are obvious, and the contrast with the surrounding environment is clear. However, when the light is insufficient on cloudy and rainy days, the quality of the images captured by the digital devices significantly reduces, resulting in unclear edges, color distortions, and a lack of texture features for the targeted objects, hindering detection. The colors of the targeted plants are relatively similar, and it is easy to miss detection when overlapping or occlusion occurs. There is a gap between the test results of the proposed model and the YOLO v3 model under varying weather conditions, but the gap is small, indicating that the method can detect weeds in maize fields in complex environments. In addition, the results demonstrate that this algorithm has better robustness to data in different situations.

The visual attention mechanism is a method to simulate human attention perception. It can effectively integrate multiscale features and improve the performance of feature extraction. Therefore, other feature extraction algorithms can be developed to fuse multi-scale features to improve the situation where feature extraction is difficult due to overlapping or occlusion by weeds. Under the premise of ensuring speed, the network can be appropriately deepened to improve the ability of feature extraction.

The detection performance of the proposed model for a few objects is constrained by the following. Certain weed types in the data set images have relatively few targets; thus, the training samples are imbalanced. Therefore, during the model training process, the features of some targets were not sufficiently extracted, which can be solved in future research. The next work will consider the detection of weeds in the later stage of maize growth and attempt to apply the model to other types of crop detection.

5. Conclusions

Aiming at identifying various weeds in maize fields in a natural environment, this paper proposes a lightweight maize weed detection model based on the YOLO v4-tiny CNN combining an attention mechanism and an improved SPP (dense SPP) structure. The model can be used to detect Maize seedlings, Abutilon, Chenopodium, Morning glory seedlings, and Amaranth. The mAP of the proposed YOLO v4-weeds model is 86.69%, and the size is 34.08 MB. The detection speed is 57.33 f/s, which is better than the Faster R-CNN, SSD 300, YOLO v3, YOLO v3-tiny, and YOLO v4-tiny models. This research indicates that the model can effectively detect maize seedlings and various weeds in complex scenes of
maize fields and can be easily applied to embedded equipment, providing the foundation and basis for developing intelligent weeding equipment.

**Author Contributions:** Conceptualization, Z.S. and R.G.; methodology, R.G.; software, S.L.; validation, Z.S., R.G. and S.L.; formal analysis, Y.J.; investigation, Z.M. and Z.R.; data curation, S.L.; writing—original draft preparation, S.L.; writing—review and editing, R.G.; visualization, Y.J.; funding acquisition, Z.S. and R.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by the Science and Technology Innovation 2030—“New Generation of Artificial Intelligence” major project (2021ZD0110904), and the Scholars Program of Northeast Agricultural University: Young talents (No. 20QC32).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

**References**

1. Li, X.; Dai, B.; Sun, H.; Li, W. Corn Classification System based on Computer Vision. *Symmetry* **2019**, *11*, 591. [CrossRef]
2. Boydston, R.A.; Williams, M.M. Sweet corn hybrid tolerance to weed competition under three weed management levels. *Renew. Agric. Food Syst.* **2015**, *31*, 281–287. [CrossRef]
3. Pérez-Ortiz, M.; Peña, J.M.; Gutiérrez, P.A.; Torres-Sánchez, J.; Herrvás-Martínez, C.; López-Granados, F. Selecting patterns and features for between-and within-crop-row weed mapping using UAV-imagery. *Expert Syst. Appl.* **2016**, *47*, 85–94. [CrossRef]
4. Fernández-Quintanilla, C.; Peña, J.M.; Andújar, D.; Dorado, J.; Ribeiro, A.; López-Granados, F.; Smith, R. Is the current state of the art of weed monitoring suitable for site-specific weed management in arable crops? *Weed Res.* **2018**, *58*, 259–272. [CrossRef]
5. Oerke, E.-C.; Gerhards, R.; Menz, G.; Sikora, R.A. *Precision Crop Protection—The Challenge and Use of Heterogeneity*; Springer: Berlin/Heidelberg, Germany, 2010; Volume 5.
6. Tang, J.; Wang, D.; Zhang, Z.; He, L.; Xin, J.; Xu, Y. Weed identification based on K-means feature learning combined with convolutional neural network. *Comput. Electron. Agric.* **2017**, *135*, 63–70. [CrossRef]
7. Pantazi, X.-E.; Moshou, D.; Bravo, C. Active learning system for weed species recognition based on hyperspectral sensing. *Biosyst. Eng.* **2016**, *146*, 193–202. [CrossRef]
8. Sabzi, S.; Abbaspour-Gilandeh, Y.; García-Mateos, G. A fast and accurate expert system for weed identification in potato crops using metaheuristic algorithms. *Comput. Ind.* **2018**, *98*, 80–89. [CrossRef]
9. Peteinatos, G.G.; Weis, M.; Andújar, D.; Rueda Ayala, V.; Gerhards, R. Potential use of ground-based sensor technologies for weed detection. *Pest Manag. Sci.* **2014**, *70*, 190–199. [CrossRef]
10. Alchanatis, A.; Moshou, D.; Bravo, C. Active learning system for weed species recognition based on hyperspectral sensing. *Biosyst. Eng.* **2016**, *146*, 193–202. [CrossRef]
11. Sabzi, S.; Abbaspour-Gilandeh, Y.; García-Mateos, G. A fast and accurate expert system for weed identification in potato crops using metaheuristic algorithms. *Comput. Ind.* **2018**, *98*, 80–89. [CrossRef]
12. Rani, S.V.J.; Kumar, P.S.; Priyadarsini, R.; Srividya, S.J.; Harshana, S. Automated weed detection system in smart farming for developing sustainable agriculture. *Int. J. Environ. Sci. Technol.* **2021**, *19*, 9083–9094. [CrossRef]
13. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
14. Lee, S.; Chang, J.H. Deep learning ensemble with asymptotic techniques for oscillometric blood pressure estimation. *Comput. Methods Programs Biomed.* **2017**, *151*, 1–13. [CrossRef] [PubMed]
15. Arribas, J.I.; Sánchez-Ferrero, G.V.; Ruiz-Ruiz, G.; Gómez-Gil, J. Leaf classification in sunflower crops by computer vision and neural networks. *Comput. Electron. Agric.* **2011**, *78*, 9–18. [CrossRef]
16. Gai, J.; Tang, L.; Steward, B.L. Automated crop plant detection based on the fusion of color and depth images for robotic weed control. *J. Field Robot.* **2019**, *37*, 35–52. [CrossRef]
17. Yamamoto, K.; Guo, W.; Yoshioka, Y.; Ninomiya, S. On plant detection of intact tomato fruits using image analysis and machine learning methods. *Sensors* **2014**, *14*, 12191–12206. [CrossRef]
18. Saravi, B.; Nejadhashemi, A.P.; Tang, B. Quantitative model of irrigation effect on maize yield by deep neural network. *Neural Comput. Appl.* **2019**, *32*, 10679–10692. [CrossRef]
19. Milioto, A.; Lottes, P.; Stachniss, C. Real-Time Blob-Wise Sugar Beets Vs Weeds Classification for Monitoring Fields Using Convolutional Neural Networks. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *4*, 41–48. [CrossRef]
20. Yu, J.; Sharpe, S.M.; Schumann, A.W.; Boyd, N.S. Detection of broadleaf weeds growing in turfgrass with convolutional neural networks. *Pest Manag. Sci.* **2019**, *75*, 2211–2218. [CrossRef]
21. Dias, P.A.; Tabb, A.; Medeiros, H. Apple flower detection using deep convolutional networks. *Comput. Ind.* **2018**, *99*, 17–28. [CrossRef]

22. Bakhshipour, A.; Jafari, A.; Nassiri, S.M.; Zare, D. Weed segmentation using texture features extracted from wavelet sub-images. * Biosyst. Eng.* **2017**, *157*, 1–12. [CrossRef]

23. Farjon, G.; Krikeb, O.; Hillel, A.B.; Alchanatis, V. Detection and counting of flowers on apple trees for better chemical thinning decisions. *Precis. Agric.* **2019**, *21*, 503–521. [CrossRef]

24. 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.

25. Redmon, J.; Farhadi, A.YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.

26. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* 2018, arXiv:1804.02767.

27. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

28. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef]

29. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.

30. 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.

31. Zhang, Z.; Tang, Z.; Wang, Y.; Zhang, Z.; Zhan, C.; Zha, Z.; Wang, M. Dense Residual Network: Enhancing global dense feature flow for character recognition. *Neural Netw.* **2021**, *139*, 77–85. [CrossRef]

32. Hou, Q.; Xing, J. KSSD: Single-stage multi-object detection algorithm with higher accuracy. *IET Image Process.* **2020**, *14*, 3651–3661. [CrossRef]

33. Nakahara, H.; Sasao, T. A High-speed Low-power Deep Neural Network on an FPGA based on the Nested RNS: Applied to an Object Detector. In Proceedings of the 2018 IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 27–30 May 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–5.

34. Rezaei, M.; Ravanbakhsh, E.; Namjoo, E.; Haghighat, M. Assessing the effect of image quality on SSD and Faster R-CNN networks for face detection. In Proceedings of the 2019 27th Iranian Conference on Electrical Engineering (ICEE), Yazd, Iran, 30 April–2 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1589–1594.

35. Wei, B.; Hao, K.; Gao, L.; Tang, X.-S. Detecting textile micro-defects: A novel and efficient method based on visual gain mechanism. *Inf. Sci.* **2020**, *541*, 60–74. [CrossRef]

36. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [CrossRef]

37. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.

38. Raghu, M.; Poole, B.; Kleinberg, J.; Ganguli, S.; Sohl-Dickstein, J. On the expressive power of deep neural networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; PMLR: New York, NY, USA, 2017; pp. 2847–2854.

39. Ying, B.; Xu, Y.; Zhang, S.; Shi, Y.; Liu, L. Weed Detection in Images of Carrot Fields Based on Improved YOLO v4. *Traitement Du Signal* **2021**, *38*, 341–348. [CrossRef]

40. McCool, C.; Perez, T.; Upcroft, B. Mixtures of lightweight deep convolutional neural networks: Applied to agricultural robotics. *IEEE Robot. Autom. Lett.* **2017**, *2*, 1344–1351. [CrossRef]

41. Wu, D.; Lv, S.; Jiang, M.; Song, H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Comput. Electron. Agric.* **2020**, *178*, 105742. [CrossRef]

42. Lin, F.; Zhang, D.; Huang, Y.; Wang, X.; Chen, X. Detection of corn and weed species by the combination of spectral, shape and textural features. *Sustainability* **2017**, *9*, 1335. [CrossRef]

43. Tian, Y.; Yang, G.; Wang, Z.; Wang, H.; Li, E.; Liang, Z. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput. Electron. Agric.* **2019**, *157*, 417–426. [CrossRef]