agronomy

Special Issue Reprint

“Smart Agriculture” Information Technology and Agriculture Cross-Discipline Research and Development

Edited by
Jian Zhang, Randy G. Goebel and Zhihai Wu

mdpi.com/journal/agronomy

MDPI

“Smart Agriculture” Information Technology and Agriculture Cross-Discipline Research and Development

“Smart Agriculture” Information Technology and Agriculture Cross-Discipline Research and Development

Editors

Jian Zhang
Randy G. Goebel
Zhihai Wu

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Editors

Jian Zhang
Department of Biology
University of British Columbia
Kelowna
Canada

Randy G. Goebel
Department of Computer Science
University of Alberta
Edmonton
Canada

Zhihai Wu
Faculty of Agronomy
Jilin Agricultural University
Changchun
China

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Agronomy* (ISSN 2073-4395) (available at: www.mdpi.com/journal/agronomy/special_issues/Smart_Agriculture_Technology).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-0818-2 (Hbk)
ISBN 978-3-7258-0817-5 (PDF)
doi.org/10.3390/books978-3-7258-0817-5

Contents

Preface

The aim of this Special Issue is to explore and support the evolution of emerging digital technology applications in agriculture and biology, including but not limited to agriculture, data collection, data mining, bioinformatics, genomics, and phenomics, as well as applications of machine learning and artificial intelligence. The development of a community to support this goal requires the cross-linking and integration of multiple sources of agricultural research across 3S technologies (remote sensing—RS; geographic information systems—GIS; global positioning systems—GPS). This provides a basis for the detection of crop pathogens, weeds, and pests (insects) using multi-spectrum techniques and the exploitation of remote sensing technology to create and analyze multiple heterogeneous-structured data sets, which enables effective cross-linking and phenomic classification. It is essential to study growth models of plants and crops and utilize expert support to develop production and smart management decision systems to achieve real-time, quantified, and precise decisions. Topics of high interest include the capture and curation of biological "big data" research on multi-spectrum analysis, the assembly of complex genetic sequencing fragments, and structural gene predictions coupled with intermediate structures to predict phenotypes. In this context, novel data structures are required to capture predictive structures in the path from genome type to phenotype, together with new techniques to identify the regularity of biological data. Finally, multiple-sources-based monitoring and decision-making for plants, water, and nutrients are required, with a research focus on the utilization of remote sensing and drone sensing to compute and predict plant water usage. This will lead to the development of precision models of crop water/nutrient management systems and form the foundation for the digitalization of agricultural water/nutrient applications.

Jian Zhang, Randy G. Goebel, and Zhihai Wu
Editors

agronomy

MDPI

Editorial

"Smart Agriculture" Information Technology and Agriculture Cross-Discipline Research and Development

Jian Zhang [1,2,*], Randy G. Goebel [3] and Zhihai Wu [1]

[1] Faculty of Agronomy, Jilin Agricultural University, Changchun 131018, China; wuzhihai@jlau.edu.cn
[2] Department of Biology, University of British Columbia Okanagan, Kelowna, BC V1V 1V7, Canada
[3] Centre for Machine Learning, Department of Computer Science, University of Alberta, Edmonton, AB T6G 2R3, Canada; rgoebel@cs.ualberta.ca
* Correspondence: jian.zhang@ubc.ca

Citation: Zhang, J.; Goebel, R.G.; Wu, Z. "Smart Agriculture" Information Technology and Agriculture Cross-Discipline Research and Development. *Agronomy* **2023**, *13*, 2536. https://doi.org/10.3390/agronomy13102536

Received: 25 September 2023
Accepted: 28 September 2023
Published: 30 September 2023

1. Introduction

Digital technology applications in agriculture and biology are a dynamic area of research interest, with topics including, but not limited to, agriculture, data collection, data mining, bioinformatics, genomics and phenomics, as well as applications of machine learning and artificial intelligence.

The development of a community to support this goal requires the cross linking and integration of multiple sources of agricultural research, including 3S technologies (remote sensing—RS, geographic information systems—GIS, global positioning systems—GPS). This broad framework provides a basis for the detection of crop pathogens, weeds and pests (insects) using multi-spectrum techniques and the exploitation of remote sensing technology to create and analyze multiple heterogeneous structured data sets, which subsequently enables effective cross linking and phenomics classification. It is essential to study the growth models of plants/crops and exploit expert support to develop smart production and management decision systems to achieve real-time, quantified and precise decisions.

This SI's topics of high interest included the capture and curation of biological "big data" research on multi-spectral data analysis, the assembly of complex genetic sequencing fragments, and structural gene predictions coupled with intermediate structures to predict phenotypes. In this context, novel data structures are required to capture predictive structures in the path from genotype to phenotype, concurrently with new techniques to capture and identify the regularity of biological data.

Finally, multiple-sources-based monitoring and decision making for plants, water and nutrients are required, with a research focus on the utilization of remote sensing and drone sensing to compute and predict plant water usage. This framework will lead to the development of precision models of crop water/nutrient management systems and form the foundation for the digitalization of agricultural water/nutrient applications.

The topic is expending rapidly and has a potential high impact. Detection, sensing and decision making in the context of modern agriculture practice is benefiting from smart/digital agriculture research and will contribute to the improvement of sustainable development. Many colleagues have contributed this Special Issue by submitting their research and studies.

2. Results

A total of 31 manuscripts were submitted to this Special Issue, with 20 manuscripts accepted and published. The content of these manuscripts includes artificial intelligence and decision making, sensor and sensing, imaging, and geographic information technology (https://www.mdpi.com/journal/agronomy/special_issues/Smart_Agriculture_Technology (accessed on 26 August 2023)). Using imaging and decision-making technologies, Bi et al. reported a novel method of corn seed identification [1]. Cai et al. determined

the varying maturity levels of strawberries by utilizing image segmentation methods from the improved DeepLabV3+ [1]. Crop seed vigor monitoring is revolutionized by optical sensor technology, which enhances performance and ensures consistent production. This technology supports reliable and non-destructive calibration techniques that allow for accurate seed use through agronomic evaluation [2]. A combination of deep learning with machine vision and Swin transformer-based models is capable of achieving a high assortment accuracy through attention to specific characteristics and multi-scale feature fusion networks [3]. To facilitate automated weed management systems, a faster R-CNN network model for weed detection in cropping regions has been introduced, increasing recognition accuracy to over 95%. Additionally, a Swin-DeepLabv3+ model has been utilized for weed recognition in soybean fields, incorporating a Swin transformer and a convolution block attention module. It resolves border contour identification concerns, ameliorates accuracy by 2.94% and achieves an average intersection ratio of 91.53% [4,5]. A deep learning grading approach has been used to identify ginseng, a crucial component of Chinese medicine, with an accuracy of 97.39% and a loss value of 0.035, making it a valuable tool for ginseng appearance quality identification. Using an ATmega 328 microcontroller and an SIM900A GSM modem, an automated system is being developed to mass produce panchagavya, a traditional organic fertilizer used in India. This system will benefit farmers and society [6,7]. A smartphone-based machine vision system is able to accurately simulate the quantities of sand, clay and silt in soil, featured in this Special Issue. This system is using learning models to determine soil texture [8]. In order to remotely monitor atmospheric CO_2, CH_4 and N_2O above rice paddy fields, two high-resolution laser heterodyne radiometers (LHRs) were set up in Hefei. The emissions from rice fields increased the amount of CO_2. Due to the negative correlations between CH_4 and N_2O emissions, rice fields play a part in carbon sequestration during the rice growth season. The LHRs are promising for monitoring emissions from agricultural fields due to their accuracy in observing air concentrations [9]. The network-based automatic algorithm for classifying and detecting rice pests has a recognition accuracy of 98.28%. The model illustrates how deep learning-based classification methods perform better at detection, while insect images are segmented effectively. With a CA attention module and STR detection head, the YOLOv5n model, CTR YOLOv5n, will detect maize disease in mobile applications. It decreases memory to 5.1 MB with an average recognition accuracy of 95.2%. With detection errors of less than 2.0% for both large- and small-diameter soybeans, a photoelectric sensor-based real-time monitoring system assesses the efficiency of soybean seed metering devices. The technique benefits from the system's assistance for device evaluation and seeding monitoring. An Efficient-Net-B5 network has been employed to describe citrus leaf diseases and foster the sustainability of the citrus industry. The network can handle small, unevenly distributed samples [10–13] and is able to obtain high accuracy rates, precision, recall and F1 scores. Hyperspectral data are used to determine plant diseases like *Septoria tritici* and *Stagonospora nodorum* blotch in cereal crops. A trained neural network can classify between healthy and damaged plants with high accuracy. This technique might help with crop estimation before harvest [14]. The development of ICT has had a huge impact on the world, especially on India's sugarcane harvest. The goal is to anticipate soil moisture and categorize sugarcane production using two-level ensemble classifiers and an ensemble model that combines support vector machines, convolutional neural networks and the Gaussian probabilistic technique. The suggested approach outperforms existing classifiers by 89.53%, enabling farmers and agricultural authorities to enhance sugarcane farming productivity, thus increasing production. The phenotype monitoring technology uses an internal gradient algorithm to accurately measure the target region and diameter of maize stems. The technology uses color images captured during the small bell stage, extracting color information and applying a morphological gradient algorithm [15,16]. A simple detection system is created for automated yield estimation and picking in small-target apple orchards using the public MinneApple dataset. The program employs 829 photos with difficult weather circumstances and reduces the size by 15.81%, while attaining an

mAP of 80.4%. A single-leaf labelling technique and a seedling detection network based on YOLOv5 and a transformer mechanism are presented to identify agricultural seedlings in challenging field situations [17]. The transformer mechanism module increased the detection capability by 1.5%, while the single-leaf labelling approach increased the model's mAP0.5 by 1.2%. By the 23 ms/frame, the optimized model increased computing speed. In recognizing early fruit body diseases in edible fungus, the study introduces a ShuffleNetV2 + SE model. This model improves disease classification performance, accuracy, precision, recall and Macro-F1 value, making it acceptable for devices with low resources [18,19]. A unique algorithm for void detection in yield maps is presented. The algorithm exhibits 100% sensitivity, 91% specificity and 82% accuracy. The method allows for smooth incorporation into real-time big data quality assessment systems based on different dimensions by mapping geographical mistakes to two common data quality dimensions [20].

3. Future Perspectives

This Special Issue, entitled ""Smart Agriculture" Information Technology and Agriculture Cross-Discipline Research and Development", has come to a close. We are very grateful for the efforts of the journal editors, peer reviewers and hard-working authors. We also would like to thank all colleagues who contributed to this section; without their efforts, this research topic would not have extended to Agronomy readers. This Special Issue has initiated much attention from research communities towards this topic and, we believe, has been completed during a remarkable time of development. We believe there will be more presentations that could be shared in the future.

Author Contributions: J.Z. and R.G.G. conception and draft. Z.W. review. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Jilin Agricultural University high level researcher grant (JLAUHLRG20102006).

Data Availability Statement: Not applicable.

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

References

1. Bi, C.; Hu, N.; Zou, Y.; Zhang, S.; Xu, S.; Yu, H. Development of Deep Learning Methodology for Maize Seed Variety Recognition Based on Improved Swin Transformer. *Agronomy* **2022**, *12*, 1843. [CrossRef]
2. Cai, C.; Tan, J.; Zhang, P.; Ye, Y.; Zhang, J. Determining Strawberries Varying Maturity Levels by Utilizing Image Segmenttion Methods of Improved DeepLabV3+. *Agronomy* **2022**, *12*, 1875. [CrossRef]
3. Mu, Y.; Feng, R.; Ni, R.; Li, J.; Luo, T.; Liu, T.; Li, X.; Gong, H.; Guo, Y.; Sun, Y.; et al. A Faster R-CNN-Based Model for the Identification of Weed Seedling. *Agronomy* **2022**, *12*, 2867. [CrossRef]
4. Yu, H.; Che, M.; Yu, H.; Zhang, J. Development of Weed Detection Method in Soybean Fields Utilizing Improved DeepLabv3+ Platform. *Agronomy* **2022**, *12*, 2889. [CrossRef]
5. Li, D.; Piao, X.; Lei, Y.; Li, W.; Zhang, L.; Ma, L. A Grading Method of Ginseng (Panax ginseng C. A. Meyer) Appearance Quality Based on an Improved ResNet50 Model. *Agronomy* **2022**, *12*, 2925. [CrossRef]
6. Sumathi, V.; Mohamed, A.J. Smart Automation for Production of Panchagavya Natural Fertilizer. *Agronomy* **2022**, *12*, 3044. [CrossRef]
7. Zhao, Z.; Feng, W.; Xiao, J.; Liu, X.; Pan, S.; Liang, Z. Rapid and Accurate Prediction of Soil Texture Using an Image-Based Deep Learning Autoencoder Convolutional Neural Network Random Forest (DLAC-CNN-RF) Algorithm. *Agronomy* **2022**, *12*, 3063. [CrossRef]
8. Li, J.; Xue, Z.; Li, Y.; Bo, G.; Shen, F.; Gao, X.; Zhang, J.; Tan, T. Real-Time Measurement of Atmospheric CO_2, CH_4 and N_2O above Rice Fields Based on Laser Heterodyne Radiometers (LHR). *Agronomy* **2023**, *13*, 373. [CrossRef]
9. Gong, H.; Liu, T.; Luo, T.; Guo, J.; Feng, R.; Li, J.; Ma, X.; Mu, Y.; Hu, T.; Sun, Y.; et al. Based on FCN and DenseNet Framework for the Research of Rice Pest Identification Methods. *Agronomy* **2023**, *13*, 410. [CrossRef]
10. Ma, L.; Yu, Q.; Yu, H.; Zhang, J. Maize Leaf Disease Identification Based on YOLOv5n Algorithm Incorporating Attention Mechanism. *Agronomy* **2023**, *13*, 521. [CrossRef]
11. Zhang, J.; Hou, Y.; Ji, W.; Zheng, P.; Yan, S.; Hou, S.; Cai, C. Evaluation of a Real-Time Monitoring and Management System of Soybean Precision Seed Metering Devices. *Agronomy* **2023**, *13*, 541. [CrossRef]

12. Dai, Q.; Guo, Y.; Li, Z.; Song, S.; Lyu, S.; Sun, D.; Wang, Y.; Chen, Z. Citrus Disease Image Generation and Classification Based on Improved FastGAN and EfficientNet-B5. *Agronomy* **2023**, *13*, 988. [CrossRef]
13. Zhelezova, S.; Pakholkova, E.; Veller, V.; Voronov, M.; Stepanova, E.; Zhelezova, A.; Sonyushkin, A.; Zhuk, T.; Glinushkin, A. Hyperspectral Non-Imaging Measurements and Perceptron Neural Network for Pre-Harvesting Assessment of Damage Degree Caused by Septoria/Stagonospora Blotch Diseases of Wheat. *Agronomy* **2023**, *13*, 1045. [CrossRef]
14. Zhang, J.; Fang, W.; Xu, C.; Xiong, A.; Zhang, M.; Goebel, R.; Bo, G. Current Optical Sensing Applications in Seeds Vigor Determination. *Agronomy* **2023**, *13*, 1167. [CrossRef]
15. Bhattacharyya, D.; Joshua, E.; Rao, N.; Kim, T. Hybrid CNN-SVM Classifier Approaches to Process Semi-Structured Data in Sugarcane Yield Forecasting Production. *Agronomy* **2023**, *13*, 1169. [CrossRef]
16. Zhou, J.; Cui, M.; Wu, Y.; Gao, Y.; Tang, Y.; Chen, Z.; Hou, L.; Tian, H. Maize (*Zea mays* L.) Stem Target Region Extraction and Stem Diameter Measurement Based on an Internal Gradient Algorithm in Field Conditions. *Agronomy* **2023**, *13*, 1185. [CrossRef]
17. Ma, L.; Zhao, L.; Wang, Z.; Zhang, J.; Chen, G. Detection and Counting of Small Target Apples under Complicated Environments by Using Improved YOLOv7-tiny. *Agronomy* **2023**, *13*, 1419. [CrossRef]
18. Kong, S.; Li, J.; Zhai, Y.; Gao, Z.; Zhou, Y.; Xu, Y. Real-Time Detection of Crops with Dense Planting Using Deep Learning at Seedling Stage. *Agronomy* **2023**, *13*, 1503. [CrossRef]
19. Xu, X.; Zhang, Y.; Cao, H.; Yang, D.; Zhou, L.; Yu, H. Recognition of Edible Fungi Fruit Body Diseases Based on Improved ShuffleNetV2. *Agronomy* **2023**, *13*, 1530. [CrossRef]
20. Byabazaire, J.; O'Hare, G.; Collier, R.; Kulatunga, C.; Delaney, D. A Comprehensive Approach to Assessing Yield Map Quality in Smart Agriculture: Void Detection and Spatial Error Mapping. *Agronomy* **2023**, *13*, 1943. [CrossRef]

agronomy

MDPI

Article

A Comprehensive Approach to Assessing Yield Map Quality in Smart Agriculture: Void Detection and Spatial Error Mapping

John Byabazaire [1,*], Gregory M. P. O'Hare [1,2], Rem Collier [1], Chamil Kulatunga [1] and Declan Delaney [3]

[1] School of Computer Science, University College Dublin, D04 V1W8 Dublin, Ireland; gregory.ohare@tcd.ie (G.M.P.O.); rem.collier@ucd.ie (R.C.); chamil.kulatunga@ucd.ie (C.K.)
[2] School of Computer Science and Statistics, Trinity College Dublin, D02 PN40 Dublin, Ireland
[3] School of Electrical and Electronic Engineering, University College Dublin, D04 V1W8 Dublin, Ireland; declan.delaney@ucd.ie
* Correspondence: john.byabazaire@ucdconnect.ie

Abstract: Smart agriculture relies on accurate yield maps as a crucial tool for decision-making. Many yield maps, however, suffer from spatial errors that can compromise the quality of their data, while several approaches have been proposed to address some of these errors, detecting voids or holes in the maps remains challenging. Additionally, the quality of yield datasets is typically evaluated based on root mean squared errors after interpolation. This evaluation method relies on weighbridge data, which can occasionally be inaccurate, impacting the quality of decisions made using the datasets. This paper introduces a novel algorithm designed to identify voids in yield maps. Furthermore, it maps three types of spatial errors (GPS errors, yield surges, and voids) to two standard data quality dimensions (accuracy and completeness). Doing so provides a quality score that can be utilized to assess the quality of yield datasets, eliminating the need for weighbridge data. The paper carries out three types of evaluations: (1) evaluating the algorithm's efficacy by applying it to a dataset containing fields with and without voids; (2) assessing the benefits of integrating void detection and other spatial error identification techniques into the yield data processing chain; and (3) examining the correlation between root mean squared error and the proposed quality score before and after filtering out spatial errors. The results of the evaluations demonstrate that the proposed algorithm achieves a 100% sensitivity, 91% specificity, and 82% accuracy in identifying yield maps with voids. Additionally, there is a decrease in the root mean squared error when various spatial errors, including voids after applying the proposed data pre-processing chain. The inverse correlation observed between the root mean squared error and the proposed quality score (-0.577 and -0.793, before and after filtering spatial errors, respectively) indicates that the quality score can effectively assess the quality of yield datasets. This assessment enables seamless integration into real-time big data quality assessment solutions based on various data quality dimensions.

Keywords: spatial data quality; smart agriculture; data quality assessment; data quality dimensions; interpolation; classification

Citation: Byabazaire, J.; O'Hare, G.M.P.; Collier, R.; Kulatunga, C.; Delaney, D. A Comprehensive Approach to Assessing Yield Map Quality in Smart Agriculture: Void Detection and Spatial Error Mapping. *Agronomy* **2023**, *13*, 1943. https://doi.org/10.3390/agronomy13071943

Academic Editors: Jian Zhang, Randy G. Goebel and Zhihai Wu

Received: 1 June 2023
Revised: 18 July 2023
Accepted: 19 July 2023
Published: 22 July 2023

1. Introduction

Technological advancements have revolutionized the agricultural industry and have significantly improved agricultural practices. The use of information and communication technologies in agriculture is collectively referred to as smart or precision agriculture [1]. Technology has been integrated into various domains of the agricultural ecosystem, and examples include automated irrigation systems that use sensors to monitor soil moisture levels and weather patterns [2] and the use of robotics to perform tasks such as harvesting crops, planting seeds, and weeding fields [3]. These technological innovations have increased efficiency, reduced labour costs, and minimized the impact of farming on the environment [3].

Yield mapping is the other such example that has gained widespread adoption due to advancements in harvesting equipment. The combine harvester is equipped with a data acquisition system that enables the collection of crop yield data during the harvesting process, including location, grain flow, and area [4]. Data from yield maps usually contain thousands of data points which have to be interpolated to create continuous yield maps that can be used for decision-making, diagnosing production issues, and optimizing management practices, as well as in research applications [4].

Data from combines, however, often contain errors arising from systematic and operator actions. Previous research [5,6] has identified several of such errors, including yield map smoothing errors, unknown crop width entering the header during harvest, time lag of grain through the threshing mechanism, positional errors, surging grain through the combine transport system, and voids/holes. As yield is important for decision-making, it is important to devise a means to mitigate the negative effects of such errors. To this end, several approaches have been defined.

Numerous solutions have been proposed to address the spatial errors [4,7,8], with some focused on a single error while others aimed to tackle multiple issues. Nonetheless, a solution for identifying and filtering voids/holes in yield maps remains yet to be defined. Voids typically result from topographical features such as waterlogging, rendering it impossible to till and consequently leaving no data points for such areas. Currently, interpolation is utilized to fill voids by incorporating neighbouring data points. However, this method is prone to errors as voids increase in size. Identifying voids in yield maps could significantly enhance the performance of downstream processes, such as interpolation.

The other limitation is that, currently, data quality of yield datasets in based the calculation of root mean square error after interpolation. This requires weighbridge data which can sometimes be unavailable, and in some cases unreliable [9]. This can affect the quality of decisions made from such unreliable data. This paper proposes a mapping of spatial errors to form data quality dimensions (DQDs), which can be used to assess the quality of yield data without the need for a gold standard (weighbridge data). DQDs provide an acceptable way to measure data quality [10]. DQDs have been used in many fields to standardise the description of quality errors so that quality improvements processes can be evaluated on a comparative basis [11]. As data from multiple sources is increasingly being integrated for decision-making, mapping spatial errors to DQDs would allow for a unified data quality assessment framework that is based on similar metrics across multiple data sources.

This paper, therefore, implements a solution to achieve two main objectives, namely: (1) To develop a novel algorithm to identify voids in yield maps. This uses yield data (location and yield) and field boundary data. (2) Create a mapping of three spatial errors, including GPS errors, yield surges, and voids, to two common DQDs of accuracy and completeness. This allows the use of DQDs as a means to assess the quality of yield datasets without the need of a gold standard and also enable seamless integration with other IoT applications that are based on DQDs. Inverse distance weighting was used as an example, as it is one of the most common downstream process for yield map data.

The rest of the paper is structured as follows; Section 2 provides an in-depth analysis of spatial errors commonly observed in yield datasets, including GPS errors, yield surges, and voids. Section 3 outlines the approach to map data quality dimensions, specifically accuracy and completeness, to the spatial errors discussed in the preceding section. Section 4 offers a detailed discussion of the novel void detection and correction algorithm. It also elucidates the mathematical implementation of the evaluation and data quality scoring strategies. Section 5 presents the results obtained and extensively discusses their practical implications and potential applications. Finally, in Section 6, a comprehensive summary and conclusion are presented.

2. Spatial Error Processing

Yield map datasets are a vital tool for site-specific and paddock management-based decision-making systems [5,6,12]. These datasets, however, usually contain many errors arising from different sources. Previous research [8] has identified several of these errors, including unknown crop width, the time lag of the grain, inappropriate GPS recording, yield surges, and voids. For the purposes of this research, only GPS errors, yield surges, and voids were considered. These are highlighted in the next section.

2.1. GPS Errors

This paper discusses two types of GPS errors: those occurring while the combine is stationary but still recording data because the header has not been lifted, and those arising from recording data outside the field boundary. To identify the first type of GPS error, the research employs an approach proposed by [8], which uses Pythagoras' theorem to calculate the distance between consecutive points. Any points with zero travel distance are deemed erroneous. For the second type of GPS error, field boundary data is used. Any points on the yield map that lie outside the field boundary are considered erroneous.

2.2. Yield Surges

Yield surges refers to the difference between the actual yield measurement and the measurements obtained from the combine. According to Beck et al. [4], yield surges are rapid changes in indicated yield over a short distance, typically resulting from operator actions such as a sudden decrease in forward speed during a period of high grain flow [4]. In contrast, Robinson et al. [8] suggest a statistical method that utilizes a moving average mean and standard deviation to detect erroneous yield surges. However, this paper employs a distinct approach that uses absolute median deviation, as significant outliers can adversely affect the mean.

2.3. Voids Errors

Fields typically have areas or sections that cannot be planted due to their topographical features, such as waterlogging or hills, making them unsuitable for tilling. Therefore, farmers usually plant around such areas, resulting in the combine harvester not producing any data for those sections during harvesting, resulting in voids or holes in the yield map. Without GPS or yield data, these voids can be challenging to identify.

To generate contour maps and high-resolution yield maps, interpolation techniques are employed to fill the voids with nearby data points. However, the accuracy of the interpolation is impacted as the size of the void increases. Therefore, voids must be identified and treated as unique cases. Figure 1 shows a yield map with a void, with the white portion in the centre representing the void. This paper introduces a novel approach for identifying and addressing voids, which is highlighted in Section 4.1.

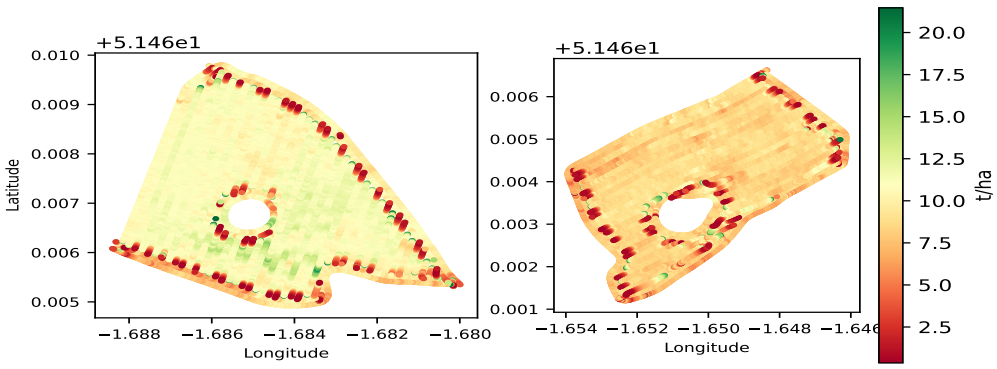


Figure 1. Example yield maps with a void/hole.

3. Data Quality Mapping

Data quality control for yield data is currently based on spatial errors. After spatial interpolation, the RMSE score is calculated as a quality indicator. If the score is small, this indicates good quality yield data. If the score is big, however, this means that the yield data is of poor quality. To calculate the RMSE score, interpolated values have to be compared to values from the weighbridge.

The data from the weighbridge, however, can suffer inaccuracies. This can be caused by several factors, including noise as the machine vibrates and errors from other foreign bodies that enter the system [9]. Weighbridge data is not available for the vast majority of the fields. This leaves most fields without a reference point from which to evaluate yield quality and mapping, and, therefore, it is difficult to determine the trustworthiness of data used to make decisions. Establishing a quality score which can be determined independently from weighbridge data is imperative.

Moreover, as precision agriculture expands, different data sources and data types are being integrated and used simultaneously. Data are from weather stations, soil sensors, and many other IoT-based data collection methods. Data quality control for this kind of data is based on data quality dimensions, and indeed this is the standard for data quality assurance in IoT [10]. Integrating quality assurance for yield datasets with other existing IoT sources to support organizational level decision-making that is based on quality data requires streamlining all quality assurance processes into a single pipeline that uses the same standard metrics.

Data quality dimensions offer a way to assess data quality using associated metrics likes accuracy and completeness. Previous research [13] has built and tested an end-to-end data quality assessment framework that uses DQDs to assess data quality in real-time with no need for a gold standard (weighbridge data). This is ideal for cases where weighbridge data is not present, or has inaccuracies. This quality assurance process can also integrate with other IoT data sources for a holistic end-to-end data quality assessment.

To achieve this, this paper creates a mapping between spatial errors and DQDs. The definitions of spatial errors and relationships between them are informed by previous research [14]. The presence, or lack, of spatial errors leads to a deterministic change in quality evaluation metrics and RMSE. This relationship can be used to determine a quality score for the data, which can be used to assess the trustworthiness of the yield datasets and how much credence the data should be used for decision-making. Figure 2 shows the mapping flow between spatial errors and DQDs.

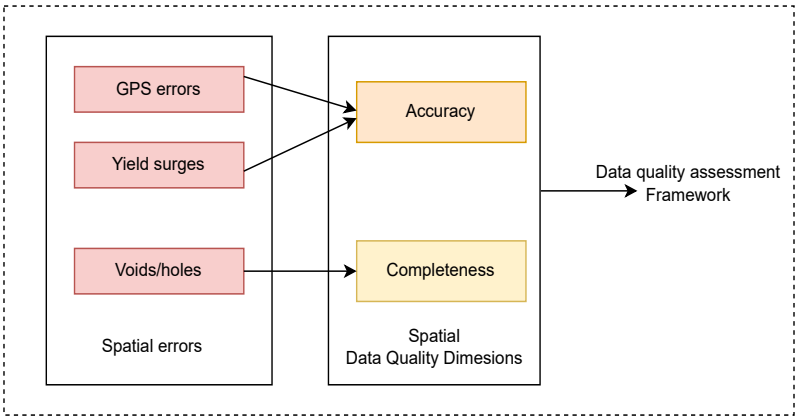


Figure 2. Mapping spatial errors to data quality dimensions.

3.1. Geospatial Data Quality Dimensions

The measure of quality is subjective and largely depends on the context in which it is applied. In manufacturing, for example, quality is often evaluated based on the product's physical attributes [14]. However, when it comes to data quality, it can be challenging to define as data lacks physical attributes. Instead, data quality is determined by intangible properties such as accuracy and completeness, which are collectively referred to as data quality dimensions (DQDs). Using DQDs provides an effective way to measure data quality, and several authors have proposed different DQDs and associated metrics to assess it [10].

Unlike other datasets, geospatial data describe phenomena in multiple dimensions, including spatial, temporal, and thematic components [14]. Therefore, DQDs for geospatial datasets have to be defined similarly. The paper defines spatial and thematic components for accuracy, while only the thematic component was used for completeness. This is due to the unavailability of time data for many farms. The mathematical definitions used here are informed by previous research in [5,8,14].

3.1.1. Accuracy

1. Spatial accuracy (positional accuracy) is applied to the spatial component of a geospatial dataset. Metrics are well defined for point entities, but widely accepted metrics for lines and area are yet to be developed [14]. We define area errors as the points (spatial coordinates that are outside a defined field of interest. These are inappropriate GPS recordings. There are two distinctions, which are points outside the field boundary and outlier points recorded while the machine is stationary.

$$Spatial\ accuracy(A_s) = 1 - \sqrt{\frac{\sum_i^n (spatial\ errors)^2}{n}} \quad (1)$$

   where n is the number of points outside the defined boundary, and spatial error is the distance a given point is from the defined boundary. To define spatial errors, this paper used the same approach defined in [8];
2. Thematic accuracy (or attribute accuracy) varies with a measurement scale. The attribute, in this case, is the yield. Beck et al. [4] define this as yield surges. There have been other techniques that have been used to eliminate this kind of error. For example, ref. [8] used a statistical identifier based on moving average mean and standard deviation. This paper uses median absolute deviation to filter out yield surges as the mean can be affected by outliers, Therefore:

$$Thematic\ accuarcy(A_t) = 1 - \frac{G_e}{n} \quad (2)$$

   where G_e is the number of outlier points in a given field, and n is the total number of points in the field.

To calculate G_e, a statistical method that employs median absolute deviations (MAD) is utilized. Absolute deviation from the median has long been utilized to filter outliers [15]. The median is a measure of central tendency, and is preferable to the mean as it is less susceptible to the presence of outliers, which can have a disproportionate impact. MAD was calculated using the formula defined by Huber et al. [16].

$$MAD = \alpha M_i(|x_i - M_j(x_j)|) \quad (3)$$

where x_j is the original observations, M_j is the median of the series, and α is data normalization constant defined by [17]. It is defined as $\alpha = \frac{1}{Q(0.75)}$, where $Q(0.75)$ is the 0.75 quantile of that underlying distribution. The normalisation step is important because otherwise MAD would estimate the scale up to a multiplicative constant [16] only.

Therefore

$$Resultant\ Accuracy = w_1 * A_s + w_2 * A_t \tag{4}$$

3.1.2. Completeness

1. Thematic completeness is based on voids/holes in the yield maps. Voids are a well-known problem that affects the quality of yield maps and, subsequently, the accuracy of any interpolation technique [18,19]. Currently, there is no defined method to identify and mitigate the effects of voids. Current approaches aim to fill voids. This affects not only the accuracy of the interpolation methods, but also downstream processes like yield prediction that might be based on such erroneous data. This paper implements a novel approach to identifying voids. This is discussed in Section 4.1. Thematic completeness is therefore given by

$$Thematic\ completeness(A_t) = 1 - \frac{GridCount}{n} \tag{5}$$

where *GridCount* is the number of grids that form the void and n is total number of grids for a given yield map.

4. Implementation

The system implementation was divided into two stages. The initial stage involves identifying and classifying spatial errors. Three spatial errors were used in this implementation. The other stage involves mapping the defined spatial error to DQDs and assessing the impact of DQDs on the spatial interpolation. Spatial interpolation was used as an assessment example because its one of the most common pre-processing techniques used to construct usable yield maps [20,21]. Each of these stages is highlighted in detail in the following sections. Figure 3 shows the end-to-end system flow of the implementation, showing void identification and correction in stage one and data quality mapping in stage two.

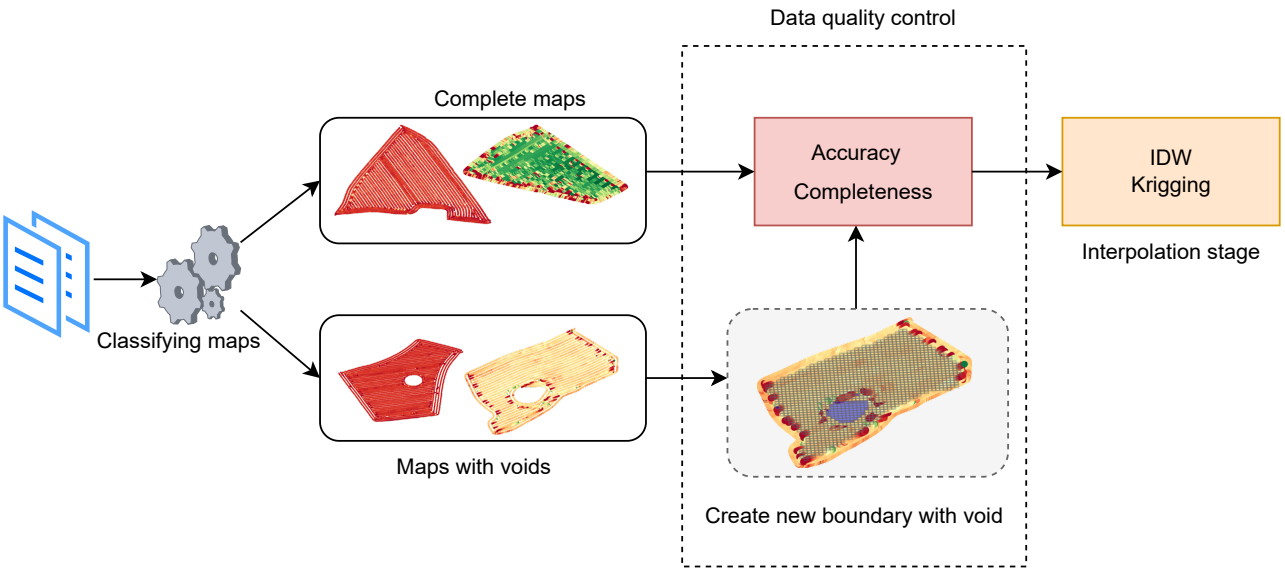


Figure 3. End-to-end process flow of the void identification, correction, and data quality mapping framework.

4.1. Void Error Correction

Yield map errors can be attributed to several causes. Prior research has identified some of these errors [8]. This paper focuses on three types of errors, namely GPS errors, yield surges, and voids. Unlike GPS errors and yield surges, no established methods exist to identify yield maps with voids. Thus, this section presents a novel approach to detecting voids in yield maps.

Two inputs are required to interpolate yield maps spatially: yield map data consisting of GPS points with corresponding yield measurements, and field boundaries defined by a set of GPS points delineating the field's limits. The latter is used to restrict the interpolation process to within the field.

Figure 4 illustrates the importance of restricting yield map interpolation to the limits of the field, as shown by the green boundary line. Without such restriction, interpolation would continue indefinitely beyond the field boundary. When dealing with yield maps that contain voids, however, an additional inner boundary exists (artefact A) that is not accounted for in the original boundary file. As a result, interpolation will continue until the void is filled. This can significantly impact the accuracy of interpolation methods and downstream processes, such as yield prediction, that rely on such data. Therefore, it is crucial to identify yield maps with voids and reconstruct the boundary files to reflect these physical features.

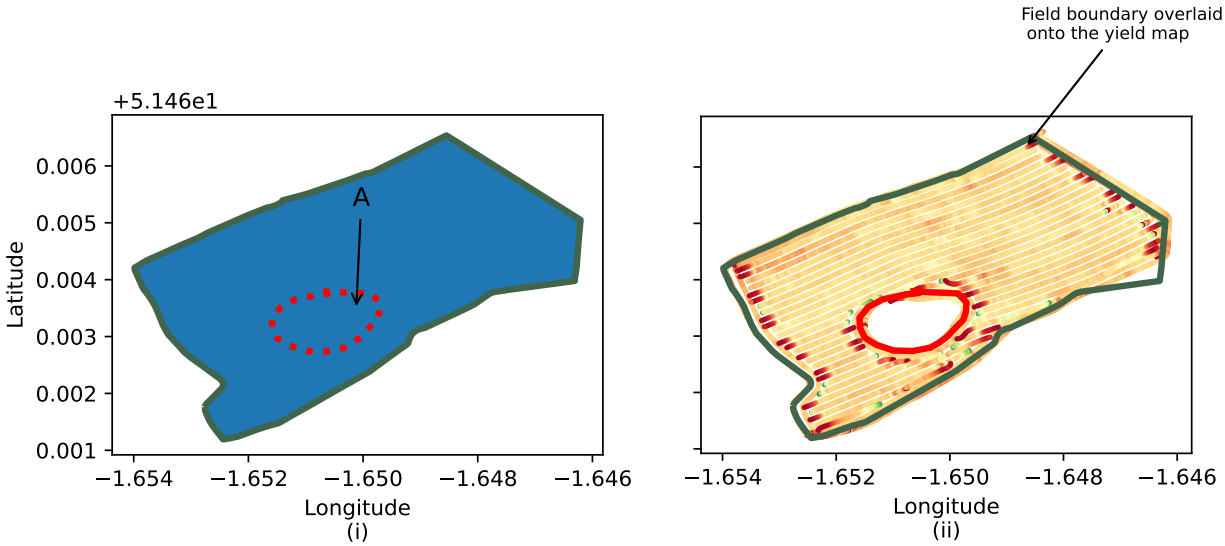


Figure 4. (**i**) Field boundary map; artefact A represents the outline of the void that is missing in the boundary map. (**ii**) Corresponding yield map with field boundary overlaid in green and the void highlighted in red.

Various approaches can help identify voids in yield maps, such as computer vision, image processing, and artificial-intelligence-based solutions, while these methods can achieve positive results, they face challenges such as the need for a considerable amount of data to train and test and high computational costs.

To identify and correct voids, the proposed approach performs the following steps. (1) The boundary map is used to generate a fixed-size grid structure encompassing the entire area of the boundary map. This is illustrated in Figures 5i and 6i. (2) For each set of grid coordinates, search the yield map data to determine if such coordinates overlap with the yield map data. (3) Determine any grids whose coordinates do not overlap with the corresponding yield map to constitute a void. The concept is that if a yield map is complete, each small grid in the field boundary map should be contained in the yield map. Otherwise, the grids within the field boundary map, but not within the yield map, constitute part of the void. (4) Finally, a new boundary map is constructed that includes the void.

An example of the process is shown in Figures 5 and 6. When a yield map has no void, the coordinates of each grid in the boundary map have a one-to-one mapping to the yield map. For example, for each grid highlighted in green in the boundary map (artefact A in Figure 5i), there is a corresponding one in the yield map (artefact A in Figure 5ii). For yield maps with a void, however, grids exist in the boundary map without mapping to the yield map. For example, in Figure 6, the green highlights (artefact A in Figure 6i) have corresponding ones to the yield map (artefact A in Figure 6ii). Grids in (artefact B in Figure 6i), however, have no mapping to the yield map because the void has no corresponding GPS data.

The initial size of the each grid was set to 10 m. This was chosen as the lower limit because the harvest line in the field have the same size. Lowering this value could increase false positives. Different size were tested to determine the optimal grid size that maximises true positive rate. The proposed method is highly computationally efficient and ideal for real-time applications. Unlike computer vision-based techniques that require converting spatial data to images (pixels), which can lead to loss of information, the proposed approach works directly with GPS coordinates, ensuring no information is lost.

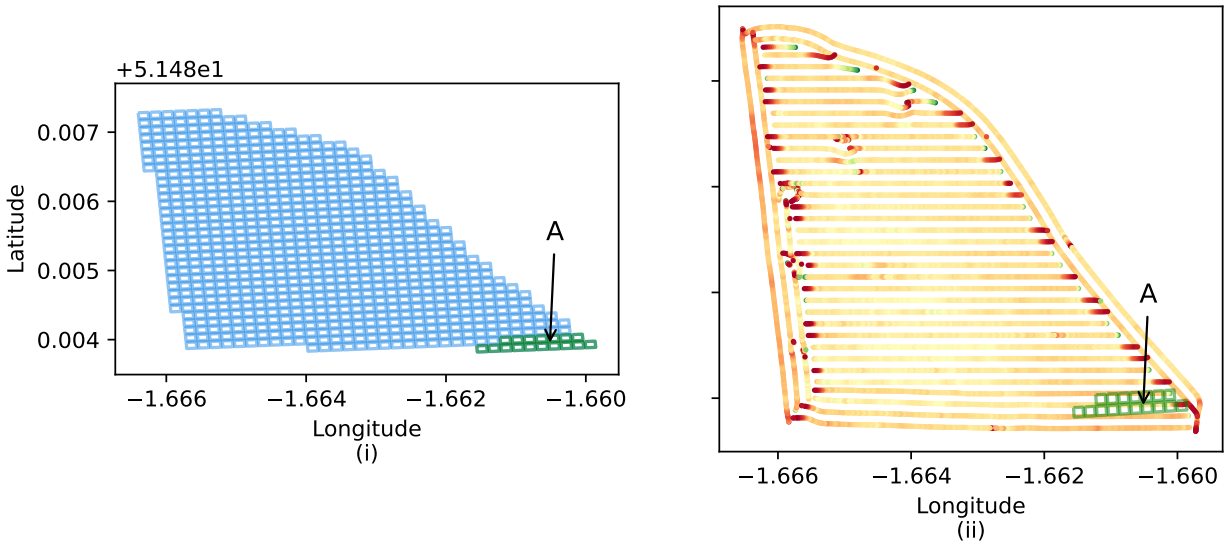


Figure 5. Field boundary map converted into grid with the corresponding yield map for a case with no void.

4.2. Data Quality Scoring

The data quality scoring technique used in this paper is based on previous research [13] that uses trust and DQDs to evaluate the quality of heterogeneous IoT data streams in real-time. Trust is a well-established metric that has been used to validate the reliability of unknown sources [22]. To this end, therefore, given a yield map dataset, the quality score will be given by

$$Quality_{score} = w_1 \cdot Accuracy + w_2 \cdot Completeness + e \quad (6)$$

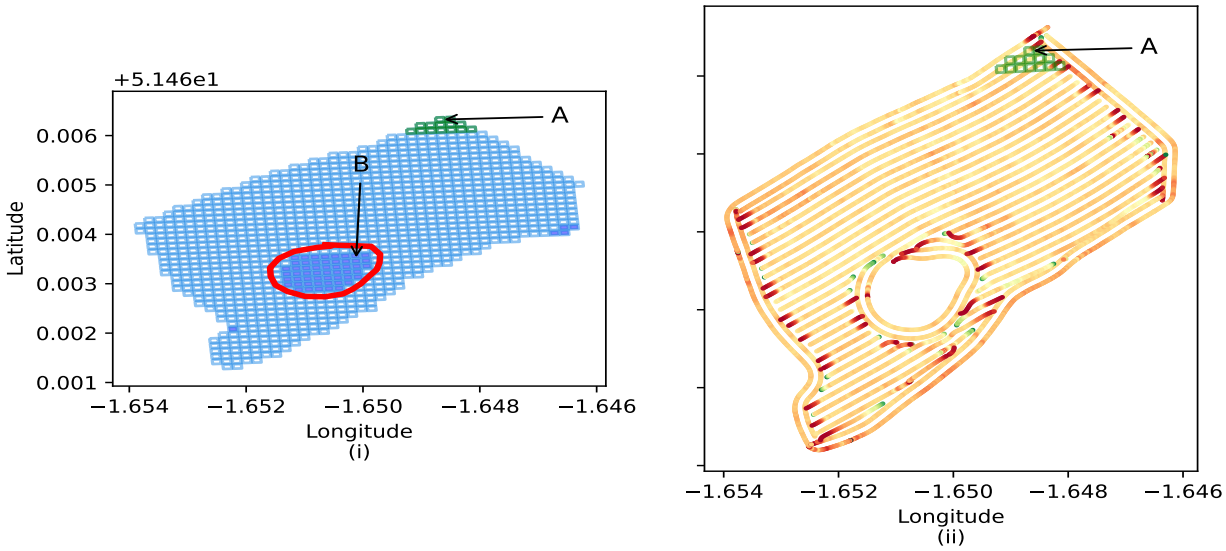


Figure 6. Field boundary map converted into grid with the corresponding yield map for a case with a void.

The weights w_1 and w_2 are determined by each use case. The goal is for each use case to be able to customise its own quality score. The metric e is the experience metric. It ensures that a past quality score and current quality score of the same dataset contribute to the overall score of the dataset. Detailed description and implementation of these can be found in our previous research [11,13,22].

4.3. Evaluation Strategy Using Spatial Interpolation

Using yield maps for decision-making requires high-resolution maps [23]. To this end, several spatial interpolation techniques exist, for example, linear interpolation, inverse distance weighting (IDW), and Kriging [24]. These work by taking known values (yield) and predicting unknown values in the neighbourhood. This process results in improved maps with clear boundaries showing the variation in yield output of the different field sections.

To evaluate the effectiveness of using DQDs for real-time quality assessment of yield maps, this paper uses IDW as an application example. Any interpolation technique, however, can be used without any changes to the downstream processes.

Inverse Distance Weighting (IDW)

IDW interpolation is a deterministic spatial interpolation method that uses known values with corresponding weights to estimate an unknown value at a particular location [25]. One IDW method, sometimes referred to as Shepard's method [26], is given by the following equation

$$\hat{z}(x) = \frac{\sum_i^n w_i z_i}{\sum_{i=1}^n w_i} \tag{7}$$

where z is the estimate at point x, n is the number of surrounding points, and w is given by:

$$w = |x - x_i|^{-\beta} \tag{8}$$

where $\beta \geq 0$ and $|\cdot|$ correspond to the Euclidean distance. Typically values $\beta = 1$ are $\beta = 2$ usually used; however, in this paper, grid search was used to obtain the optimal values.

To evaluate the performance of IDW, the root mean squared error (RMSE) was used. This is given by the following equation:

$$RMSE = \sqrt{\frac{\sum_i^n \parallel \alpha_i - \beta_i \parallel^2}{n}} \tag{9}$$

where n is the number of data points and α_i, β_i are actual and predicated values, respectively.

5. Evaluation

The system implementation is divided into two stages. In the first stage, spatial errors are classified, and in the second stage, the impact of data quality issues (DQDs) on spatial interpolation is assessed. To evaluate the system's performance, two experiments were conducted.

The first experiment aims to evaluate the benefits of adding void detection to the spatial error processing data pipeline using a grid approach. The effectiveness of the proposed algorithm is evaluated using a dataset containing fields with and without voids. The algorithm creates a grid of cells from the field boundary map and compares each cell to the yield map. As the grid size affects the algorithm's performance, a range of grid sizes are compared using accuracy, sensitivity, and specificity as performance metrics.

The second experiment aims to evaluate the efficacy of using a quality score calculated using DQDs that are mapped from spatial errors as a means to assess the quality of yield map datasets. The dataset described in the next section was used. The advantage of using the proposed quality score is that it is not based on a gold standard (weighbridge data). The RMSE of the yield dataset and quality score are calculated before and after filtering spatial errors. The mean of scores of RMSE and quality score for each year are compared.

5.1. Dataset Description

The dataset consists of 524 yield maps collected from 267 fields across 20 farms in the United Kingdom. The biggest field spans over 56.0 hectares; the smallest is approximately 0.1 hectares. The data was collected from 2013 to 2019 and consists of three crops: winter barley, spring barley, and winter wheat. The dataset also includes weighbridge data corresponding to each field. This is used as the true measure of yield per year. A total of 21% of all the yield maps used had voids, and of these, 50% had a high (greater than two tones/per hectare) discrepancy between the measured yield output from the combine and the actual output from the weighbridge. Figure 7 summarises the dataset with the distribution of crops across fields for each year and the void distribution across fields.

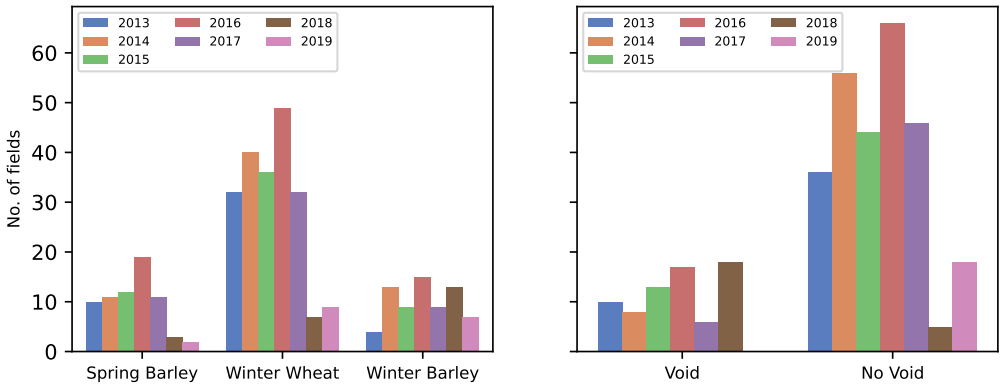


Figure 7. Crop distribution across fields for each year and void distribution across fields per year.

5.2. Results and Discussion

This section is structured into three main parts. The first two sections detail results related to a novel void detection and correction algorithm. Furthermore, they examine the influence of these enhancements on downstream processes, particularly interpolation. The third part evaluates the effectiveness of the data quality dimensions (DQDs) in filtering spatial errors, utilizing spatial interpolation as the evaluation method.

5.2.1. Void Identification

Figure 8 shows the results of the classification algorithm. This is a binary classification problem with yield maps with voids given as positive examples and those without voids considered as negative examples. Yield maps usually contain other kinds of gaps, for example harvest lines (gap between two harvest line) and other anomalies which are not voids, and should not be misclassified as such. Choosing an optimal grid size is important to avoid false positives. Typically the gap between two harvest lines is usually about 10 m. For this reason grid size values below 10 m were not used to avoid this miss identification. Grid sizes above 25 m had the highest score of 100%, 91%, and 82% for sensitivity, specificity, and accuracy, respectively. The main objective of the experiment was to minimise false positives, as this would affect downstream processes. These are defined as below:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{10}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{11}$$

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{12}$$

where TP is true positive, FP is false positive, TN is true negative, and FN is false negative.

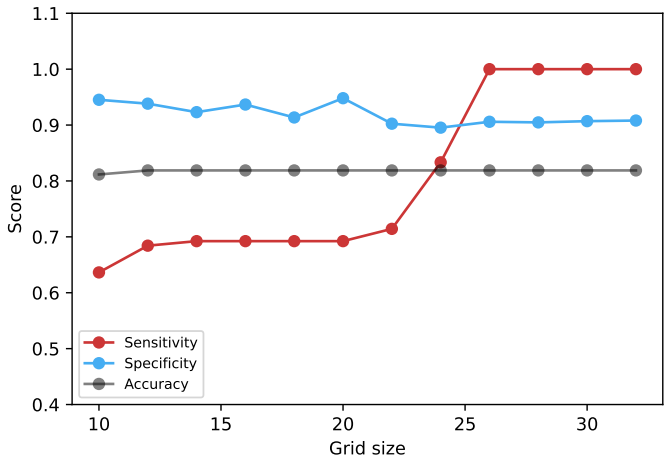


Figure 8. Performance (accuracy, specificity, and sensitivity) of the classification algorithm for different grid size values.

5.2.2. Effects of Void Correction and Other Spatial Errors on RMSE

Figure 9 shows the effects of the presence of voids and other spatial errors on the interpolation process. To assess this effect, the RMSE score after interpolation is used. The RMSE was computed by comparing the interpolated values with the actual yield values from the weighbridge. The lower the RMSE value, the better the interpolation performance, and vice versa. The study compared different search radii values, which also affect interpolation performance. Obtaining an optimal value is critical. Values ranging from 5 to 10 m were used, although there was no significant difference. As shown in Figure 9, after filtering spatial errors, there was a very significant improvement in the overall RMSE score.

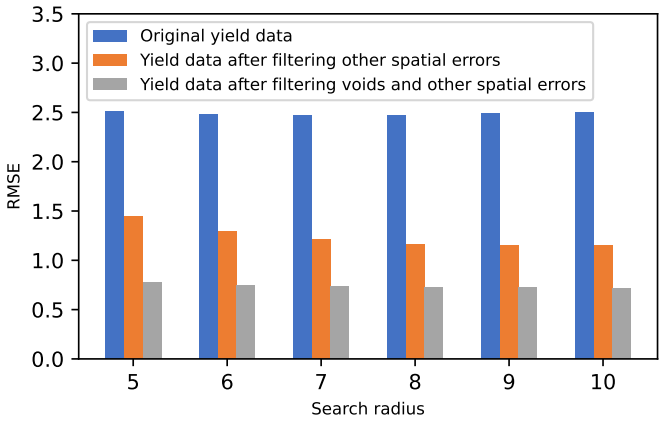


Figure 9. Performance of IDW and the effects of correction of void and other spatial errors under different values of search radius.

Figure 10i displays the original yield map before interpolation. Figure 10ii,iii explores the effects of not considering and considering spatial errors on the interpolation process, respectively. Figure 10ii (artefact A) demonstrates that conventional approaches that do not account for voids and other spatial errors can result in inaccuracies. This can compromise downstream processes such as yield prediction that rely on this data. Additionally, the interpolation process can also impact regions outside of the void, leading to incorrect yield representations and potentially erroneous decision-making, such as in the case of automated fertilizer applications. On the other hand, Figure 10iii takes into account the presence of voids and other spatial errors, and, therefore, effectively addresses these shortcomings.

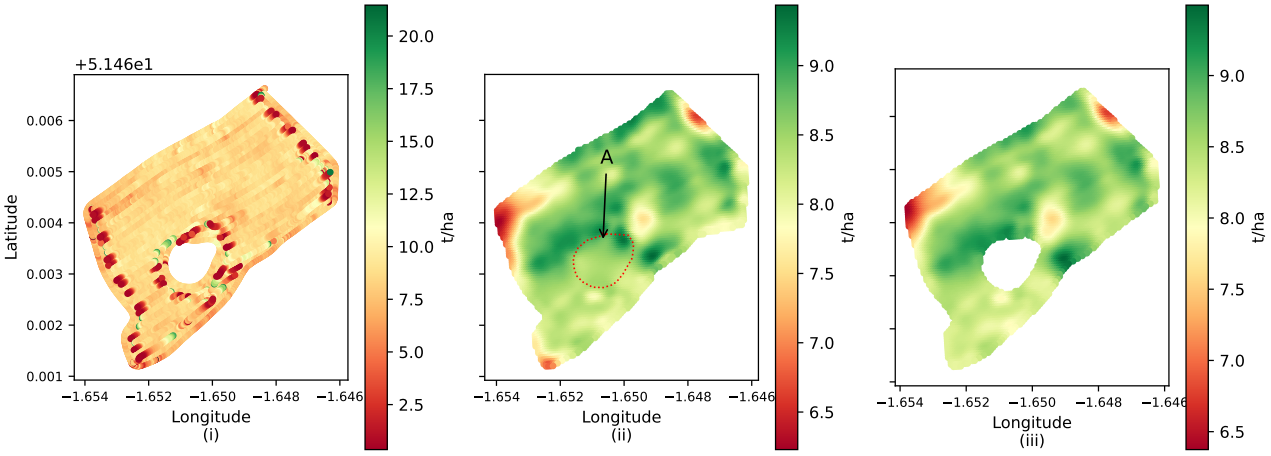
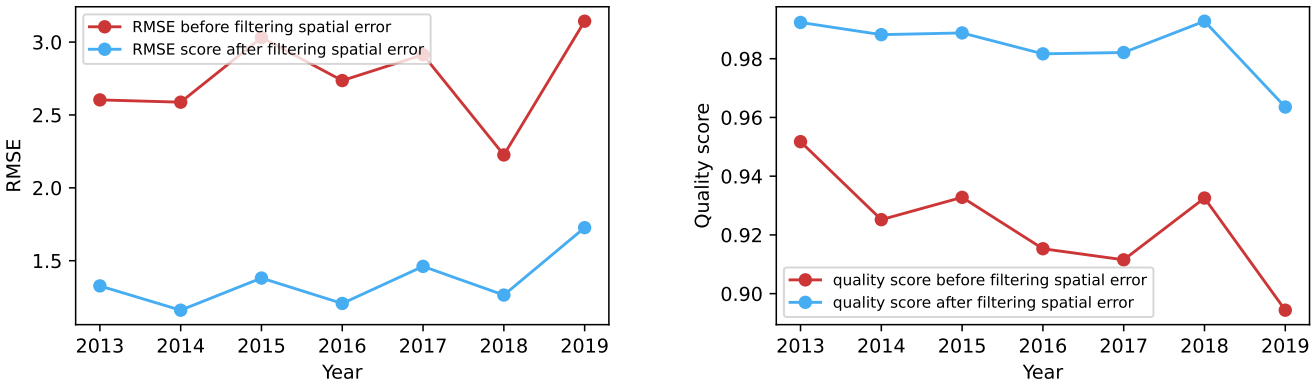


Figure 10. (**i**) Original yield map before interpolation, (**ii**) yield map interpolated without considering spatial DQDs, and (**iii**) yield map interpolated after considering the proposed mapping of spatial DQDs.

5.2.3. Using DQDs as a Score for Yield Data Quality

Figure 11a,b illustrates the yield data's average RMSE and quality score for different years, both before and after filtering out spatial errors. Pearson's correlation coefficient was used to analyse the relationship between the RMSE and the quality score. The comparison was made between the RMSE values before and after filtering and their respective quality scores. An inverse relationship was observed between the RMSE and the proposed quality score. This behaviour is expected since, as the yield data quality improves, the RMSE is anticipated to decrease while the quality score would increase, and vice versa.

The correlation coefficient before filtering out spatial errors was lower than the correlation coefficient after the filtering process. This discrepancy can be attributed to outliers, affecting most interpolation methods and consequently impacting the resulting RMSE score. The stronger correlation observed after filtering out spatial errors suggests that the proposed quality score can be employed as an effective means of assessing yield data quality without requiring a gold standard. The correlation results are summarized in Tables 1 and 2. These are based on the RMSE score presented in Figure 11.



(**a**) (**b**)

Figure 11. Comparing the RMSE and quality score before and after interpolation of yield data for various years. (**a**) RMSE before and after filter spatial errors. (**b**) Quality score before and after filter spatial errors.

Table 1. Pearson correlation between quality score and RMSE before filtering spatial errors

	QualityScoreBefore	**RMSEBefore**
QualityScoreBefore	1	−0.577
RMSEBefore	−0.577	1

Table 2. Pearson correlation between quality score and RMSE after filtering spatial errors

	QualityScoreAfter	**RMSEAfter**
QualityScoreAfter	1	−0.793
RMSEAfter	−0.793	1

6. Conclusions

Technological advancements in agriculture have revolutionized farming practices by implementing precision agriculture. This approach leverages data from various processes to make informed decisions and optimize agricultural practices. However, it is essential to acknowledge that data collected from combines can often contain errors resulting from systematic and operator actions, significantly impacting the decision-making process. To address this challenge, several approaches, such as filtering and interpolation methods, have been proposed to mitigate these errors. Nevertheless, finding a comprehensive solution to identify and filter voids or holes in yield maps remains a persistent challenge.

This paper introduces a novel algorithm specifically designed to identify voids in yield maps. The algorithm effectively maps three types of spatial errors, namely GPS errors, yield surges, and voids, to two commonly used data quality dimensions (DQDs): accuracy and completeness. To the best of our knowledge, no existing solution currently employs this approach. The effectiveness of DQDs in filtering spatial errors was evaluated using spatial interpolation techniques. This work has the potential to significantly enhance the performance of downstream processes for yield map data and establish a unified data quality assessment framework based on consistent metrics across multiple data sources.

While this study focused on addressing three specific spatial errors, namely GPS errors, yield surges, and voids, and two common DQDs of accuracy and completeness, our future work will involve the integration of additional spatial errors and DQDs. We will explore the overall impact on the root mean square error (RMSE). Additionally, future research efforts will aim to incorporate larger zones within farms, different crops, and various harvest combines, as these factors can influence the quality of yield maps. Furthermore, most yield maps contain a single void, but there are instances where this number can increase to two or even three and, as the number of voids increases, the algorithm's accuracy is impacted. Future work will also aim to address this challenge.

Author Contributions: J.B.: Implementation, Writing—original draft, Validation and editing. G.M.P.O.: Supervision and Investigation. R.C.: Supervision and Investigation. D.D.: Supervision, Investigation, Validation, Writing and editing. C.K.: System implementation. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded under the SFI Strategic Partnership Programme (16/S/3296) and is co-funded by Origin Enterprises plc.

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

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

References

1. Shafi, U.; Mumtaz, R.; García-Nieto, J.; Hassan, S.A.; Zaidi, S.A.R.; Iqbal, N. Precision Agriculture Techniques and Practices: From Considerations to Applications. *Sensors* **2019**, *19*, 3796. [CrossRef] [PubMed]
2. Ramachandran, V.; Ramalakshmi, R.; Srinivasan, S. An Automated Irrigation System for Smart Agriculture Using the Internet of Things. In Proceedings of the 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore, 18–21 November 2018; Volume 11, pp. 210–215. [CrossRef]
3. Kootstra, G.; Wang, X.; Blok, P.M.; Hemming, J.; van Henten, E. Selective Harvesting Robotics: Current Research, Trends, and Future Directions. *Curr. Robot. Rep.* **2021**, *2*, 95–104. [CrossRef]
4. Beck, A.; Searcy, S.; Roades, J. Yield data filtering techniques for improved map accuracy. *Appl. Eng. Agric.* **2001**, *17*, 423.
5. Blackmore, S. Remedial correction of yield map data. *Precis. Agric.* **1999**, *1*, 53–66. [CrossRef]
6. Blackmore, S.; Godwin, R.J.; Fountas, S. The Analysis of Spatial and Temporal Trends in Yield Map Data over Six Years. *Biosyst. Eng.* **2003**, *84*, 455–466. [CrossRef]
7. Trevisan, R.G.; Shiratsuchi, L.; Bullock, D.; Martin, N. Improving yield mapping accuracy using remote sensing. In Proceedings of the Precision Agriculture '19, Wageningen, The Netherlands, 8–11 July 2019; Volume 7, pp. 901–908. [CrossRef]
8. Robinson, T.; Metternicht, G. Comparing the performance of techniques to improve the quality of yield maps. *Agric. Syst.* **2005**, *85*, 19–41. [CrossRef]
9. Longchamps, L.; Tisseyre, B.; Taylor, J.; Sagoo, L.; Momin, A.; Fountas, S.; Manfrini, L.; Ampatzidis, Y.; Schueller, J.K.; Khosla, R. Yield sensing technologies for perennial and annual horticultural crops: A review. *Precis. Agric.* **2022**, *23*, 2407–2448. [CrossRef]
10. Chen, M.; Song, M.; Han, J.; Haihong, E. Survey on data quality. In Proceedings of the 2012 World Congress on Information and Communication Technologies, WICT 2012, Trivandrum, India, 30 October–2 November 2012.
11. Byabazaire, J.; O'Hare, G.; Delaney, D. Using Trust as a Measure to Derive Data Quality in Data Shared IoT Deployments. In Proceedings of the Proceedings—International Conference on Computer Communications and Networks, ICCCN, Honolulu, HI, USA, 3–6 August 2020. [CrossRef]
12. Sudduth, K.A.; Drummond, S.T.; Myers, D.B. Yield Editor 2.0: Software for Automated Removal of Yield Map Errors. In *Proceedings of the 2012 Dallas, Texas, 29 July–1 August 2012*; American Society of Agricultural and Biological Engineers: St. Joseph, MI, USA, 2012. [CrossRef]
13. Byabazaire, J.; O'Hare, G.M.; Delaney, D.T. End-to-End Data Quality Assessment Using Trust for Data Shared IoT Deployments. *IEEE Sens. J.* **2022**, *22*, 19995–20009. [CrossRef]
14. Veregin, H. Data quality parameters. *Geogr. Inf. Syst.* **1999**, *1*, 177–189.
15. Leys, C.; Ley, C.; Klein, O.; Bernard, P.; Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* **2013**, *49*, 764–766. [CrossRef]
16. Huber, P.J. *Robust Statistical Procedures*; SIAM: Philadelphia, PA, USA, 1996.
17. Rousseeuw, P.J.; Croux, C. Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.* **1993**, *88*, 1273–1283. [CrossRef]
18. Qiu, Z.; Yue, L.; Liu, X. Void Filling of Digital Elevation Models with a Terrain Texture Learning Model Based on Generative Adversarial Networks. *Remote Sens.* **2019**, *11*, 2829. [CrossRef]

19. Reuter, H.I.; Nelson, A.; Jarvis, A. An evaluation of void-filling interpolation methods for SRTM data. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 983–1008. [CrossRef]
20. Lyle, G.; Clarke, K.; Kilpatrick, A.; Summers, D.M.; Ostendorf, B. A Spatial and Temporal Evaluation of Broad-Scale Yield Predictions Created from Yield Mapping Technology and Landsat Satellite Imagery in the Australian Mediterranean Dryland Cropping Region. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 50. [CrossRef]
21. Mariano, C.; Mónica, B. A random forest-based algorithm for data-intensive spatial interpolation in crop yield mapping. *Comput. Electron. Agric.* **2021**, *184*, 106094. [CrossRef]
22. Byabazaire, J.; O'Hare, G.; Delaney, D. Data quality and trust: A perception from shared data in IoT. In Proceedings of the 2020 IEEE International Conference on Communications Workshops, ICC Workshops 2020—Proceedings, Dublin, Ireland, 7–11 June 2020. [CrossRef]
23. Vega, A.; Córdoba, M.; Castro-Franco, M.; Balzarini, M. Protocol for automating error removal from yield maps. *Precis. Agric.* **2019**, *20*, 1030–1044. [CrossRef]
24. Mueller, T.G.; Pusuluri, N.B.; Mathias, K.K.; Cornelius, P.L.; Barnhisel, R.I.; Shearer, S.A. Map Quality for Ordinary Kriging and Inverse Distance Weighted Interpolation. *Soil Sci. Soc. Am. J.* **2004**, *68*, 2042–2047. . [CrossRef]
25. Lu, G.Y.; Wong, D.W. An adaptive inverse-distance weighting spatial interpolation technique. *Comput. Geosci.* **2008**, *34*, 1044–1055. [CrossRef]
26. Gordon, W.J.; Wixom, J.A. Shepard's method of "metric interpolation" to bivariate and multivariate interpolation. *Math. Comput.* **1978**, *32*, 253–264.

agronomy

Article

Recognition of Edible Fungi Fruit Body Diseases Based on Improved ShuffleNetV2

Xingmei Xu, Yuqi Zhang, Hongcheng Cao, Dawei Yang, Lei Zhou and Helong Yu *

College of Information Technology, Jilin Agricultural University, Changchun 130118, China; xingmeix@jlau.edu.cn (X.X.); zhangyuqi@mails.jlau.edu.cn (Y.Z.); caohongcheng@mails.jlau.edu.cn (H.C.); yangdawei@mails.jlau.edu.cn (D.Y.); zhoulei@mails.jlau.edu.cn (L.Z.)
* Correspondence: yuhelong@jlau.edu.cn

Abstract: Early recognition of fruit body diseases in edible fungi can effectively improve the quality and yield of edible fungi. This study proposes a method based on improved ShuffleNetV2 for edible fungi fruit body disease recognition. First, the ShuffleNetV2+SE model is constructed by deeply integrating the SE module with the ShuffleNetV2 network to make the network pay more attention to the target area and improve the model's disease classification performance. Second, the network model is optimized and improved. To simplify the convolution operation, the 1×1 convolution layer after the 3×3 depth convolution layer is removed, and the ShuffleNetV2-Lite+SE model is established. The experimental results indicate that the accuracy, precision, recall, and Macro-F1 value of the ShuffleNetV2-Lite+SE model on the test set are, respectively, 96.19%, 96.43%, 96.07%, and 96.25%, which are 4.85, 4.89, 3.86, and 5.37 percent higher than those before improvement. Meanwhile, the number of model parameters and the average iteration time are 1.6 MB and 41 s, which is 0.2 MB higher and 4 s lower than that before the improvement, respectively. Compared with the common lightweight convolutional neural networks MobileNetV2, MobileNetV3, DenseNet, and EfficientNet, the proposed model achieves higher recognition accuracy, and its number of model parameters is significantly reduced. In addition, the average iteration time is reduced by 37.88%, 31.67%, 33.87%, and 42.25%, respectively. The ShuffleNetV2-Lite+SE model proposed in this paper has a good balance among performance, number of parameters, and real-time performance. It is suitable for deploying on resource-limited devices such as mobile terminals and helps in realization of real-time and accurate recognition of fruit body diseases of edible fungi.

Keywords: edible fungi fruit body; disease recognition; ShuffleNetV2; attention mechanism

Citation: Xu, X.; Zhang, Y.; Cao, H.; Yang, D.; Zhou, L.; Yu, H. Recognition of Edible Fungi Fruit Body Diseases Based on Improved ShuffleNetV2. *Agronomy* **2023**, *13*, 1530. https://doi.org/10.3390/agronomy13061530

Academic Editor: Paul Kwan

Received: 27 April 2023
Revised: 26 May 2023
Accepted: 30 May 2023
Published: 31 May 2023

1. Introduction

China's edible fungi industry is the fifth largest industry after grain, oil, vegetables, and fruits [1,2], which contributes much to the country's economic construction. The development of edible fungi industry provides an important future food source for humans [3,4]. Edible fungi have high protein content, high nutritional value and high medicinal value, and they have significant value as functional foods and for medicinal purposes [5]. With the increasing demand for edible fungi, it is urgent to ensure the production and quality of edible fungi and maintain the healthy development of the edible fungi industry [6–8]. In recent years, the planting scale and types of edible fungi have gradually increased, followed by an increasing number of disease types [9]. There are many edible fungi diseases in the frustum stage, such as brown rot, soft rot, pilum spot disease, brown spot disease, straw mushroom pellet sclerotium disease, etc., which cause serious economic losses [10]. At present, the problems in the prevention and control of edible fungi diseases are the following: there are many types of edible fungi diseases, and mushroom farmers cannot accurately diagnose edible fungi diseases, failing to achieve effective prevention and control in the initial stage of the disease. As a result, disease control is seriously affected, resulting

in a decline in the quality and yield of edible fungi. The fruit body period of edible fungi is a key period for disease prevention and control. Realizing accurate diagnosis of disease types in the fruit body period of edible fungi can provide theoretical guidance for farmers to accurately spray drugs, which is significant for improving the yield of edible fungi [11,12] and ensuring the healthy and sustainable development of the edible fungi industry [13–15].

Traditional machine learning methods, such as naïve Bayes [16–18], logistic regression [19], and support vector machine [20–22], are not suitable for recognizing edible fungi diseases in the fruit body period due to their shortcomings in high computational complexity, slow convergence rate, and difficulty in processing a large number of complex samples [23,24]. In recent years, deep learning methods have been widely studied in crop disease recognition [25–30]. For instance, Nurul Nabilah et al. [31] took 974 pepper disease images collected by themselves and used traditional methods and deep learning methods for experimental comparison. It was found that the recognition accuracy of deep learning was 92.10%, and the deep learning method was significantly better than the traditional feature extraction method. Chen J. et al. [32] proposed an INC-VGGNet model for rice disease recognition based on the VGGNet framework by introducing the Inception module and using the transfer learning method. The accuracy of the model on the public dataset was 91.83%, and the experimental results indicated that the method can be used for disease classification. With the rapid development of deep learning technology, more and more researchers apply lightweight convolutional neural networks to classify and identify crop diseases [33–37]. Chen Junde et al. [38] proposed a Mobile-DANet model that retains the transition layer structure based on DenseNet, replaces the traditional convolutional layer with a depth-wise separable convolutional layer, and introduces an attention mechanism to learn the relationship between channels and the importance of spatial points to input features. Mobile-DANet achieved an average recognition accuracy of 95.86% on locally collected data. Wang et al. [39] proposed an attention-based deep separable Bayesian optimization neural network for recognizing and classifying rice diseases. The model was improved from MobileNet and trained and tested on four categories of rice disease datasets. Atila et al. [40] used the EfficientNet model to identify plant diseases, and after adjusting the hyperparameters, the recognition accuracy was 99.91%. Kang et al. [41] developed an automatic mushroom recognition system using a convolutional neural network. To better investigate the characteristics of mushroom image data, AlexNet, VGGNet, and GoogLeNet were used for comparative experiments, and the class number expansion and fine-adjustment technology were exploited to realize transfer learning. The final top-five accuracy (the probability that the top five categories contain the actual results) was 96.84%. To sum up, although the research on crop disease recognition based on deep learning methods has achieved good results, it still requires much manual intervention in the recognition process, the number of model parameters is large, and the training time is long, which is not conducive to realizing rapid real-time detection of diseases. Therefore, it is necessary to use lightweight convolutional neural networks to design an efficient and non-destructive method for recognizing edible fungi fruit body diseases.

Aiming at the problems mentioned above, this study proposes a lightweight neural-network-based method for recognizing edible fungi diseases in the fruit body period. This paper improves the ShuffleNetV2 network based on the fusion of an attention mechanism and obtains the ShufflenetV2-Lite+SE model [42–47], which is used to recognize edible fungi diseases in the fruit body period. This method provides certain technical guidance for the recognition of edible fungi diseases.

2. Materials and Methods

2.1. Construction of the Dataset

This study used an image dataset of edible fungi fruit bodies. Part of the data comes from Fusong County, Tonghua County, and other places in Baishan City, Jilin Province. It was filmed by researchers on site, and some of the data comes from the internet and reference books. In order to avoid the problem of misplacing disease categories, we

specifically invited domain experts to screen the data and obtained a dataset containing 649 images of four types (three types of diseases and one type of health), including 224 images of physical disease, 139 images of bacterial disease, 131 images of fungal disease, and 155 images of healthy condition. The images were saved in the jpg format. Some edible fungi fruit body images are shown in Figure 1.

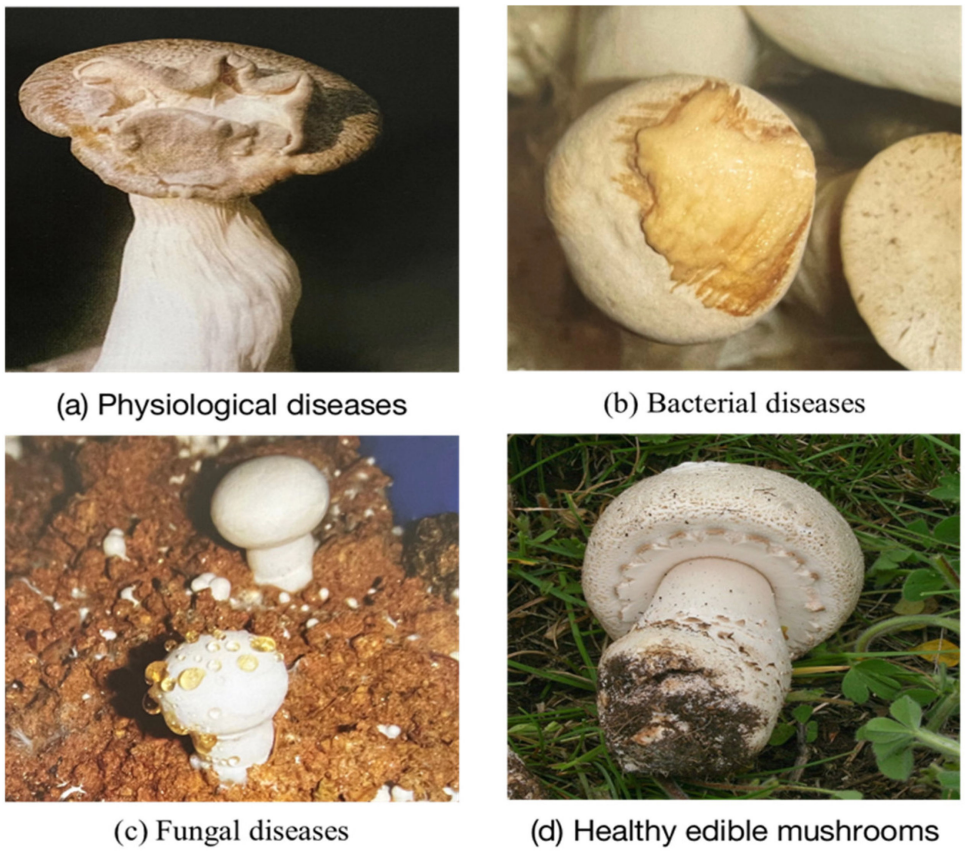


Figure 1. The image types of fruit bodies of edible fungi.

2.2. Data Preprocessing

2.2.1. Data Augmentation

The edible fungi fruit body images of three types of diseases and one type of healthy condition in the dataset were enhanced in the same way to increase the difference of training data and improve the generalization ability of the model. The processing operations include adding Gaussian noise, increasing light intensity, decreasing light intensity, vertical flip, and horizontal flip [48,49]. These operations adjust the angle, brightness, and blur of the original image, and the total number of images in the dataset is 3439 after data augmentation. The balance of the test data is ensured by data augmentation [50], the quality and quantity of data samples are improved, and sufficient data samples are provided for training the convolutional neural network [51]. Some examples of data augmentation results are shown in Figure 2.

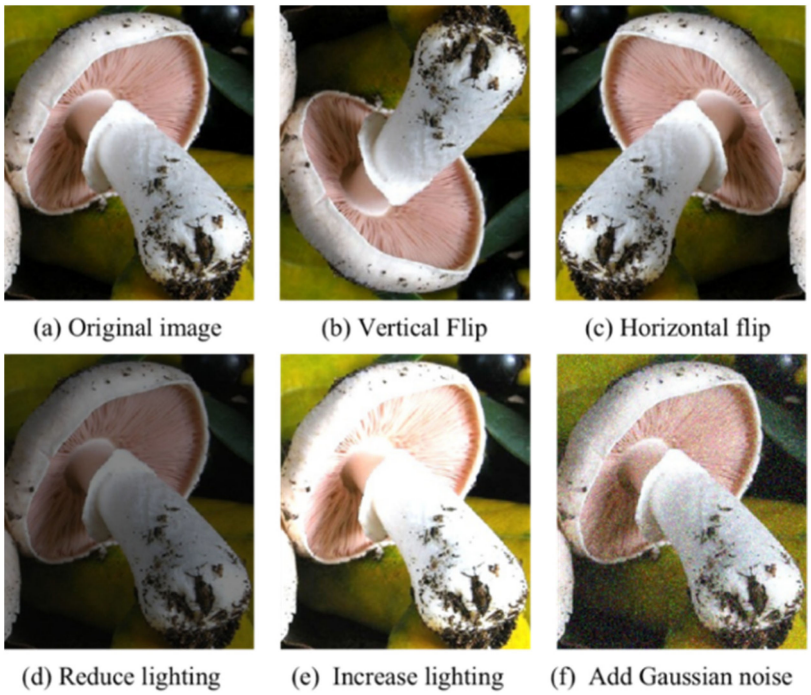


Figure 2. Data enhancement results.

2.2.2. Dataset Partitioning

The label of the edible fungi fruit body disease dataset was set with the Python code [52], including four categories: physiological disease, bacterial disease, fungal disease, and healthy condition after data enhancement, which are represented as Physiological, Bacterial, Fungoid, and Health, respectively. Figure 3 shows the statistical chart of the quantity of each type of condition.

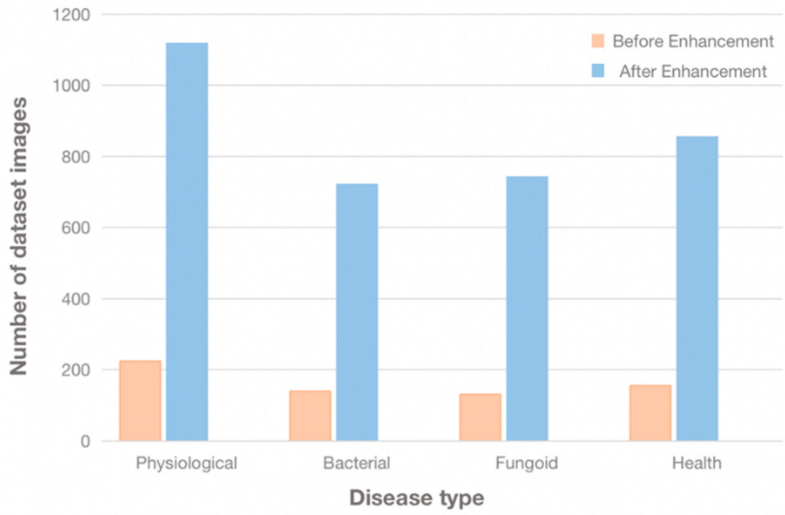


Figure 3. The quantity distribution of each type of image.

In this test, the dataset was divided into a training set and a test set at a ratio of 8:2, with 2752 and 687 samples, respectively. The dataset division is shown in Table 1.

Table 1. Dataset division.

Condition	Training Set	Test Set
Physiological	895	224
Bacterial	576	145
Fungoid	594	149
Health	684	172

3. Recognition Model of Fruit Body Diseases of Edible Fungi

3.1. ShuffleNetV2 Model

ShuffleNetV2 is a lightweight network model proposed by Zhang et al. [53] which aims to greatly reduce the model size and speed up the operation of the model without sacrificing performance efficiency. The biggest innovation of this model is that it fully utilizes the two operations of channel shuffle and group convolution to reduce the calculation amount and the number of parameters of the model. Specifically, channel shuffle is an operation that disrupts the channels of feature maps in order and reconstructs a new feature map to solve the problem of poor information flow caused by group convolution. Excessive group convolution can lead to a large Memory Access Cost (MAC) overhead. In the structure of the ShuffleNetV2 network, the number of output channels of Stage2, Stage3, Stage4, and Conv5 increases successively, while the number of output channels of the stage structure increases successively. As the network depth increases, the feature extraction ability of the model is gradually enhanced, and the detection accuracy is continuously improved. The structure of the ShuffleNetV2 network is shown In Figure 4.

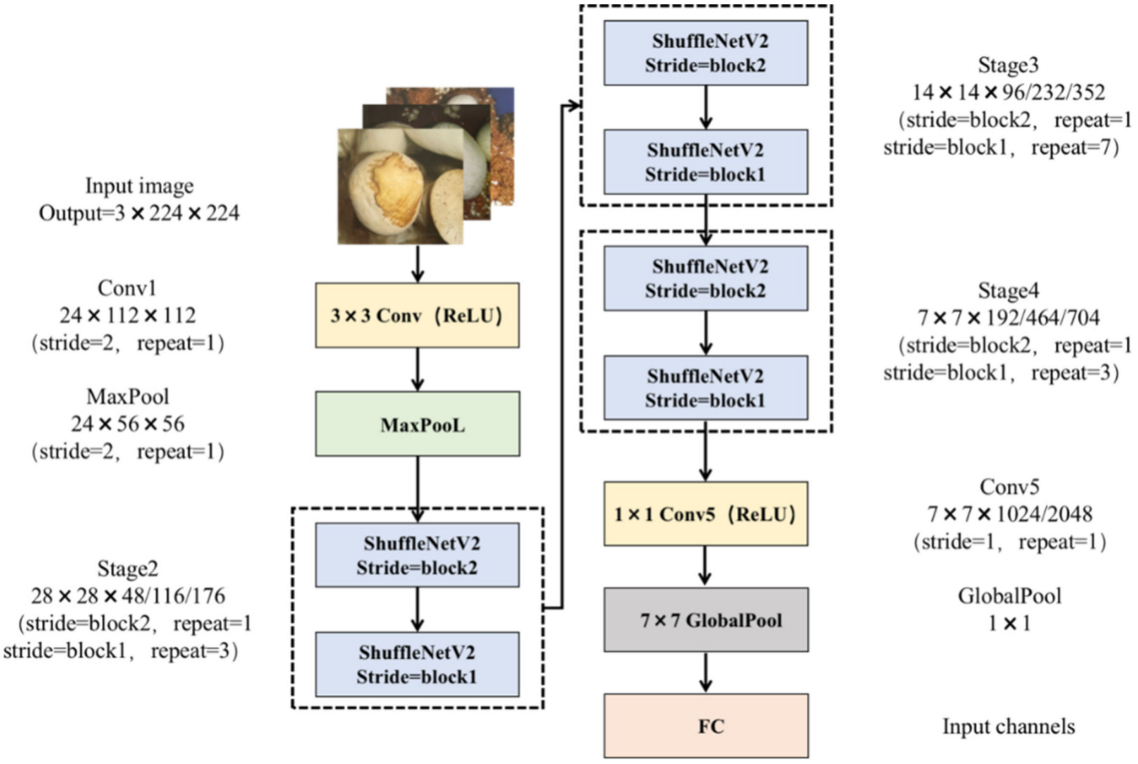


Figure 4. The structure of the ShuffleNetV2 network.

The basic structure of the ShuffleNetV2 network consists of two types of blocks. Block1 randomly divides the input channel into two parts: one part retains its mapping and directly transmits downward; the other part performs separable convolution to extract image features. The output channel of the two parts is combined at the bottom of the module. Then, a random mixing operation is performed on the final output feature graph

channel, and the structure is shown in Figure 5a. Block2 sends all feature diagrams into two network branches, and the output feature diagrams of the two branches are combined at the bottom of the module to double the number of final output channels.

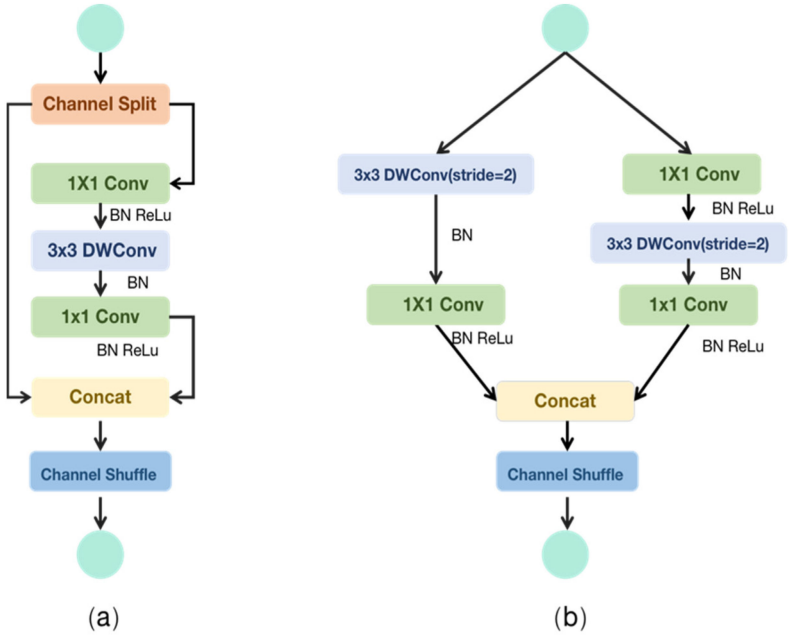


Figure 5. (**a**) Basic unit. (**b**) Spatial down sampling unit (2×).

The structure is shown in Figure 5b.

In ShuffleNetV2, all feature channels have the same weight, and the number of channels doubles every time Block2 passes through. With the doubling of the number of channels, much attention is paid to the feature channels that have a great impact on the classification results. Meanwhile, the depth-separable convolution used in Block2 is sensitive to the location of sensitive features, and too much background information is retained, which easily affects the classification effect.

3.2. Model Improvement

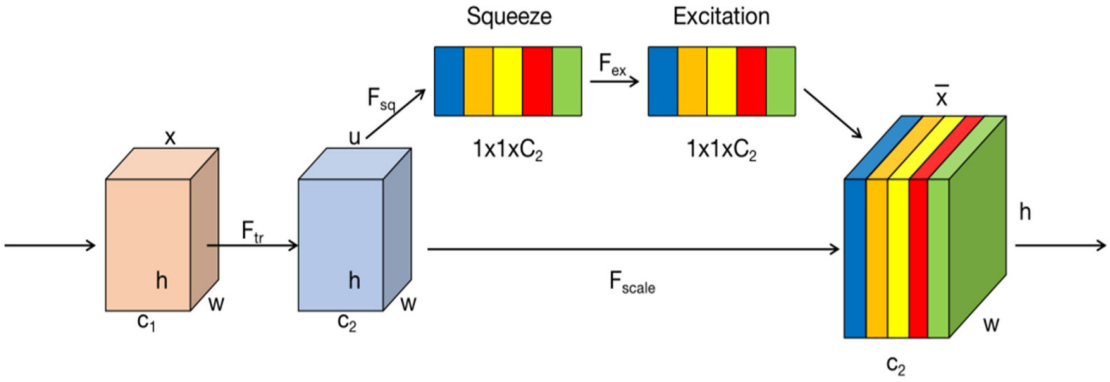
There are some problems with the disease dataset of edible fungi fruit body constructed in this study, such as complex disease image background, high similarity between some diseases and background, large difference in disease size, etc. As a result, the overall recognition is difficult, and the existing model has poor recognition performance. Considering the recognition accuracy and speed of the model, the ShuffleNetV2 model is improved in this study.

3.2.1. Attention Mechanism

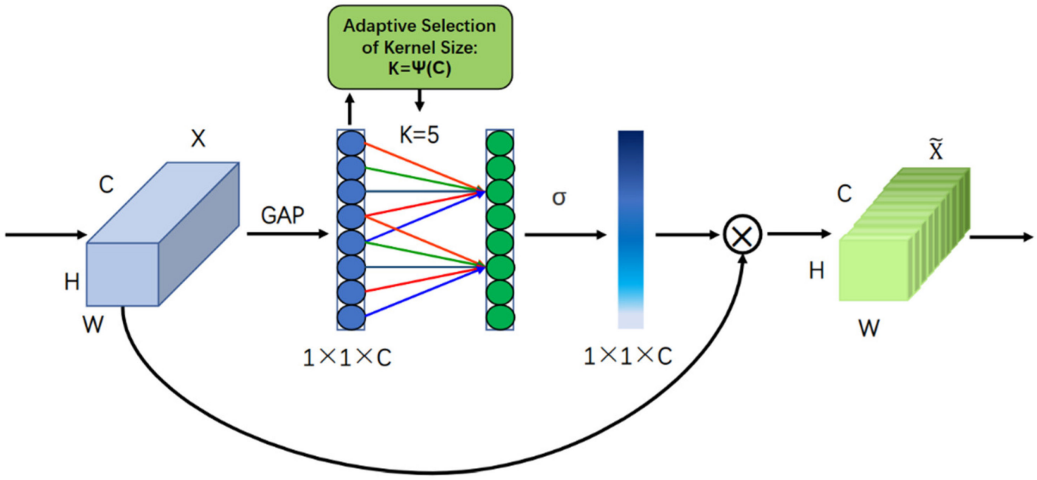
The attention mechanism aims to focus on areas of interest and try to suppress the role of areas of interest in image segmentation as much as possible. In deep learning CNN, attention mechanisms can be divided into two types: channel attention and spatial attention. The channel attention refers to determining the weight relationship between different channels, enhancing the weight of key channels, and suppressing channels with little inhibitory effect. The spatial attention is the determination of the weight relationship between different pixels in the spatial neighborhood, enhancing the weight of key area pixels, allowing the algorithm to pay more attention to the research area we need, and reducing the weight of unnecessary areas.

The SE (Squeeze-and-Excitation) attention mechanism is a method of determining weights in channel attention mode, which achieves priority by assigning weights between different channels. The SE attention module can adjust the weight according to different

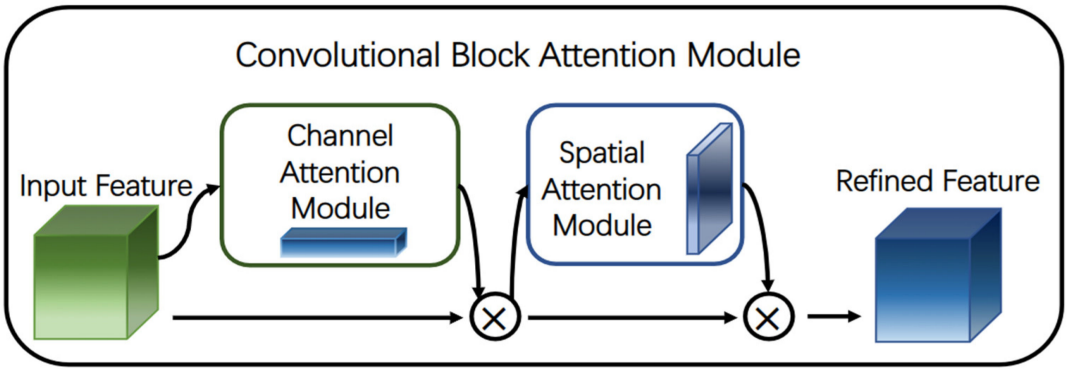
feature channels, automatically enhance the feature channels with rich contrast information in the image, and effectively suppress the feature channels unrelated to the target. The model structure is shown in Figure 6a.



(**a**)



(**b**)



(**c**)

Figure 6. (**a**) The structure of the SE module. (**b**) The structure of the ECA module. (**c**) The structure of the CBAM module.

The ECA (Efficient Channel Attention) attention mechanism is a method of channel attention. To avoid damaging the direct correspondence between channels and their weights, a one-dimensional convolution method is proposed to avoid the impact of dimensionality reduction on data. The model structure is shown in Figure 6b.

The CBAM (Convolutional Block Attention Module) attention mechanism is an algorithm model that combines channel and spatial attention mechanisms. The input feature map is first subjected to channel attention mechanism, followed by spatial attention mechanism operation, and, finally, output, achieving the goal of strengthening the region of interest from both channel and spatial aspects. The model structure is shown in Figure 6c.

3.2.2. Simplify Model Structure

Among the basic components of ShuffleNetV2, three convolution layer operations go through on the right branch, which are the normal 1 × 1 convolution layer (performing Batch Normalization, BN, and ReLU), the 3 × 3 deep convolution layer (performing BN), and the normal 1 × 1 convolution layer (performing BN ReLU). Two 1 × 1 convolution layer operations are used here, but there are actually some superfluous layers. Because the dimensionality-up and dimensionality-down operation is not necessary here, only the fusion of the inter-channel information of the DW convolution using a 1 × 1 convolution layer operation is sufficient. Therefore, in this study, deletion of a 1 × 1 convolution layer after a 3 × 3 deep convolution layer is considered to achieve the goal of model lightweight. The improved ShuffleNetV2-Lite network structure is shown in Figure 7.

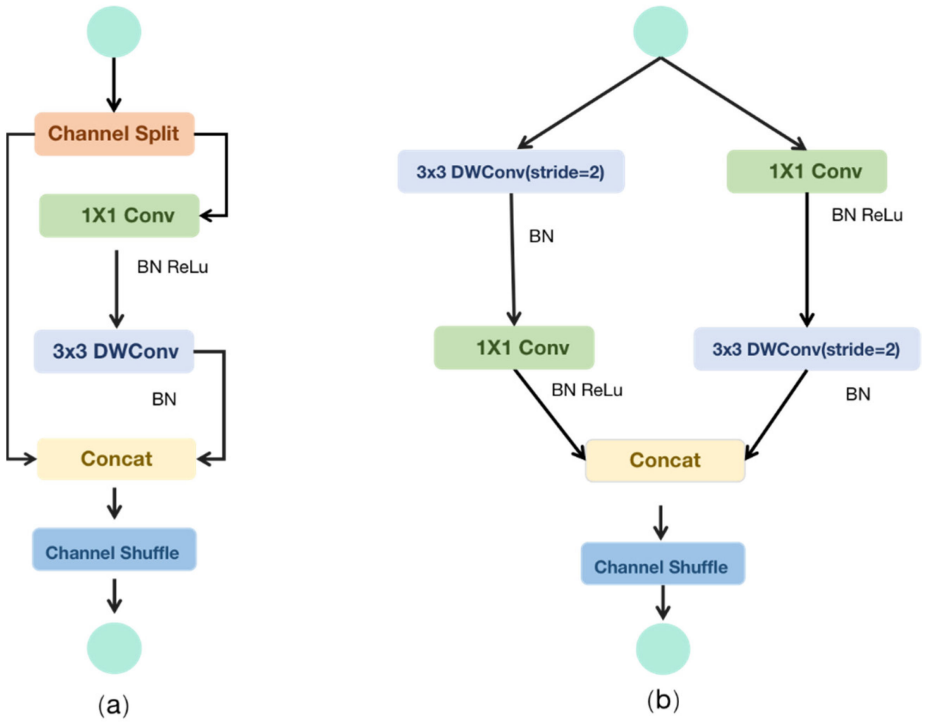


Figure 7. (**a**) The improved basic unit. (**b**) The improved Spatial down from the unit (2×).

In this study, the redesigned ShuffleNetV2-Lite network structure can effectively reduce computational complexity and optimize model performance while ensuring model accuracy.

3.3. Experimental Environment and Hyperparameters

Python3.8 and CUDA11.3 are configured under the Windows 10 operating system to build a convolutional neural network structure with the Pytorch framework as the core. The parameters of the experimental environment are shown in Table 2.

Table 2. The parameters of the experimental environment.

Environment	Parameters
OS	Windows 10
CPU	Intel Xeon Gold 6246R @ 3.40 GHz 32 cores
Memory	128 GB
Deep learning framework	Pytorch-GPU 1.12.1
GPU	NVIDIA Quadro RTX 8000
CUDA version	CUDA Toolkit 11.3
Pytorch version	Pytorch 0.9.1

The parameters for model training are shown in Table 3: the input image size is 224 × 224 pixels, the number of training iterations is 200, the batch size is 64, the learning rate is 0.01, the optimizer is Adam, and the cross-entropy loss function is adopted. To prevent overfitting and under fitting, L2 regularization method and the Class_Weighting loss function weighting method are added to the cross-entropy loss function of the model. We increased the weight of the loss function for the small class sample data of the classifier and reduce the weight of the loss function for the large class sample to solve the problem of data imbalance.

Table 3. Parameters for model training.

Parameters	Values
Epoch	200
Batch size	64
Learning rate	0.01
Optimizer	Adam

3.4. Evaluation Index

This study uses four indicators, accuracy, precision, recall, and Macro-F1, to measure the model performance.

(1) Accuracy represents the proportion of correct results predicted by the model in the total samples. The calculation formula is

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (1)$$

(2) Precision represents the proportion of positive samples predicted correctly by the model in the predicted positive samples. The calculation formula is

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (2)$$

(3) Recall represents the proportion of positive samples predicted by the model to the actual positive samples. The calculation formula is

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (3)$$

(4) Macro-F1 represents the harmonic mean of precision and recall, and it reflects the comprehensive performance of the model. The calculation formula is

$$\text{Macro-F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{precision} + \text{Recall}}, \quad (4)$$

where TP indicates that the predicted positive sample is also the actual positive sample; FP indicates that the predicted positive sample is actually negative; FN indicates that the

predicted negative sample is actually positive; TN indicates that the predicted negative sample is also actually negative.

4. Experimental Results and Analysis

4.1. Model Performance Evaluation with Different Attention Mechanisms

In this experiment, the Convolutional Block Attention Module (CBAM) module, the Efficient Channel Attention (ECA) module, and the Squeeze-and-Excitation (SE) module were embedded into ShuffleNetV2, respectively, and the experiments were conducted on the edible fungi fruit body disease image dataset. The performance comparison of different attention mechanisms is shown in Table 4.

Table 4. Performance comparison of different attention mechanisms.

Model	Accuracy (%)	Params (MB)	Time/Epoch (s)
ShuffleNetV2	91.34	1.4	45
ShuffleNetV2+CBAM	92.91	1.9	57
ShuffleNetV2+ECA	93.15	2.3	61
ShuffleNetV2+SE	93.77	1.8	47

The recognition accuracy of the ShuffleNetV2+SE model was 93.77%, which was 2.43%, 0.14%, and 0.62% higher than that of ShuffleNetV2, ShuffleNetV2+CBAM, and ShuffleNetV2+ECA models, respectively. It can be seen that by adding the attention mechanism, the information interaction between channels can be better realized, and the performance of the model can be improved. Especially after adding the SE module, the model pays more attention to the channel features with the most information, and unimportant channel features are suppressed. In this experiment, more attention was paid to the disease area of the edible fungi fruit body, higher disease recognition accuracy was obtained, and the model achieved the best performance.

The parameter of the ShuffleNetV2+SE model was 1.8 MB, which is 0.4 MB more than that of the original model. The parameters of the ShuffleNetV2+CBAM model and the ShuffleNetV2+ECA model were reduced by 0.1 MB and 0.5 MB, respectively. The average iteration time of the ShuffleNetV2+SE model was 47 s, showing an increase of 2 s compared with the original model. Compared with the ShuffleNetV2+CBAM model and the ShuffleNetV2+ECA model, the average iteration time was reduced by 10 s and 14 s, respectively. In terms of model parameters, the SE module has the best performance among the three attention mechanisms. In terms of training time, adding the attention mechanism can effectively reduce the average iteration time. Considering model recognition accuracy, the number of model parameters, as well as the average iteration time, the SE module was chosen to be embedded in ShuffleNetV2 to construct the model in this study.

4.2. Ablation Experiments for the ShuffleNetV2-Lite+SE Model

The attention SE module is introduced into the ShuffleNetV2 model, and the structure of the model is optimized to obtain the ShuffleNetV2+SE model, the optimized ShuffleNet-Lite model, and the ShufflenetV2-Lite+SE model, respectively. To investigate the performance improvement effect of the ShuffleNetV2 model brought by the attention mechanism and the structural optimization, ablation experiments were conducted. The performance of the model was evaluated in terms of accuracy, precision, recall, and Macro-F1 value on the test set, and the number of model parameters and the average iteration time were used to evaluate the complexity of the model. The performance comparison of different models is shown in Table 5.

Table 5. Performance comparison of different models.

Model	Accuracy (%)	Precision (%)	Recall (%)	Macro-F1 (%)	Params (MB)	Time/Epoch (s)
ShuffleNetV2	91.34	91.54	92.21	90.88	1.4	45
ShuffleNetV2+SE	93.77	93.84	93.59	93.71	1.8	47
ShuffleNetV2-Lite	93.88	93.17	94.70	93.93	1.2	40
ShuffleNetV2-Lite+SE	96.19	96.43	96.07	96.25	1.6	41

As shown in Table 5, the accuracy, precision, recall, and the Macro-F1 value of the ShuffleNetV2-Lite+SE model reach 96.19%, 96.43%, 96.07%, and 96.24%, respectively. The accuracy, precision, recall, and the Macro-F1 value are improved by 2.42, 2.59, 2.48, and 2.54 percent, respectively. Compared with other models, the proposed model achieves higher accuracy and better performance.

The parameters and the average iteration time of the ShufflenetV2-Lite+SE model are 1.6 MB and 41 s, respectively. Compared with the ShuffleNetV2+SE model with the highest accuracy, the model parameters are reduced by 11.11 percent, and the iteration time is reduced by 12.77 percent. Thus, the proposed model has lower model complexity than other models.

The improved model proposed in this study shows a better balance between the performance, complexity, and real-time performance of the model. It is suitable for deploying on embedded resource-constrained devices such as mobile terminals and helps to realize real-time and accurate recognition of edible fungi fruit body diseases.

4.3. Performance Comparison of ShuffleNetV2-Lite+SE with Other Models

To further evaluate the recognition effect of the improved network model on edible fungi disease images, the ShuffleNetV2-Lite+SE model was compared with the representative lightweight convolutional neural networks including MobileNetV2, MobileNetV3, DenseNet, and EfficientNet on the same dataset under the same experimental environment and network parameter configuration. The accuracy, precision, recall, and Macro-F1 value of these models on the test set were used to evaluate the model performance, and the number of model parameters and the average iteration time were used to evaluate the complexity of the model. The performance comparison of ShuffleNetV2-Lite+SE with other models is shown in Table 6, and the accuracy of the model on the test set, the change in loss value, and the confusion matrix during iterations are shown in Figures 8–10.

Table 6. The performance comparison of ShuffleNetV2-Lite+SE with other models.

Model	Accuracy (%)	Precision (%)	Recall (%)	Macro-F1 (%)	Params (MB)	Time/Epoch (s)
MobileNetV2	85.72	85.33	85.93	85.63	3.5	66
MobileNetV3	91.72	91.38	91.40	91.39	5.1	60
DenseNet	88.50	88.58	87.04	87.80	7.8	62
EfficientNet	89.29	89.46	89.98	89.84	5.3	71
ShuffleNetV2-Lite+SE	96.19	96.43	96.07	96.25	1.6	41

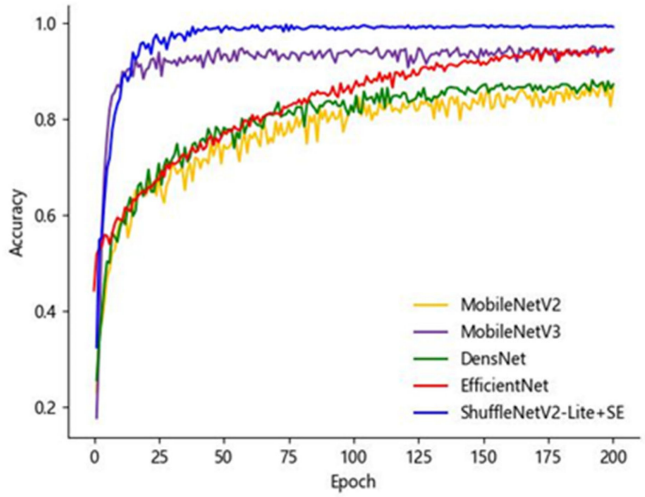


Figure 8. The accuracy of different models.

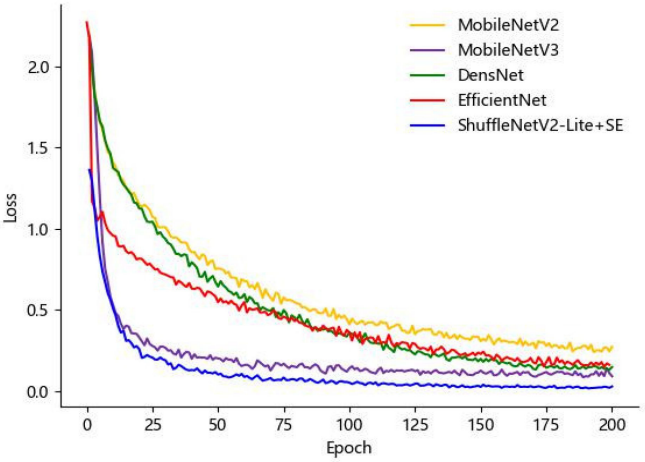


Figure 9. The loss values of different models.

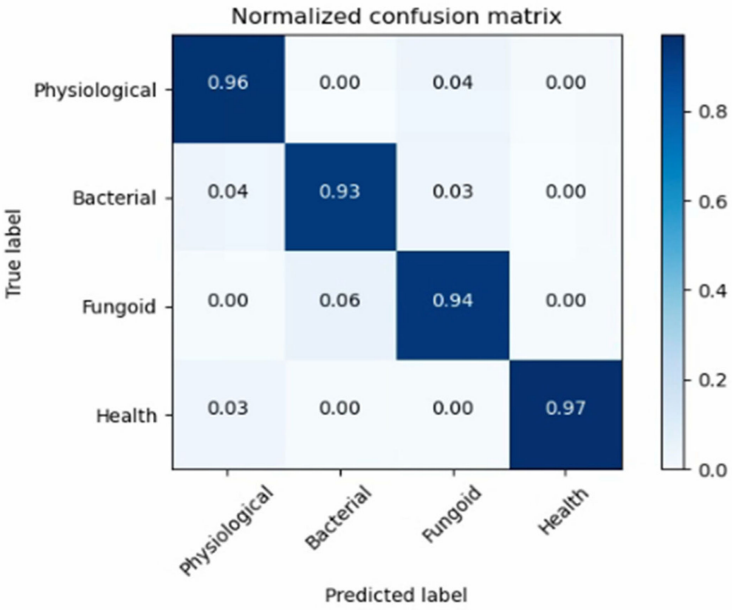


Figure 10. The confusion matrix.

It can be seen from Table 6 that the ShuffleNetV2-Lite+SE model achieves an accuracy of 96.19%, a precision of 96.43%, a recall of 96.07%, and a Macro-F1 value of 96.24%, respectively. Compared with the best-performing MobileNetV3 model, the accuracy, precision, recall, and Macro-F1 value are improved by 4.47, 5.05, 4.67, and 4.86 percent, respectively. Compared with other models, the proposed model has higher accuracy and better performance.

The parameters and the average iteration time of the ShuffleNetV2-Lite+SE model are 1.6 MB and 41 s, respectively. Compared with the best-performing MobileNetV3 model, the parameters of the ShufflenetV2-Lite+SE model are reduced by 68.63%, and the average iteration time is reduced by 31.67%, indicating that the ShufflenetV2-Lite+SE model has lower model complexity than other models.

It can be seen from Figure 8 that the accuracy of the five models gradually flattens out as the number of iteration rounds increases. After the 20th iteration round, the accuracy curve of the ShuffleNetV2-Lite+SE model is relatively smooth, there is no over-fitting phenomenon, and the accuracy becomes stable at more than 90%, while the accuracy of the other models only reaches about 80%. After the 70th iteration round, the accuracy of all models tends to be stable, and the accuracy of the ShuffleNetV2-Lite+SE model reaches 96.19%, which is the highest among all models. The results indicate that the network has good recognition accuracy.

Figure 9 shows that in the initial training stage, the loss values of all networks continuously decrease and eventually tend to be stable without large fluctuations. The ShuffleNetV2-Lite+SE model has a significantly better convergence speed than other models. After the 29th iteration round, the loss value of the ShuffleNetV2-Lite+SE model remains below 0.2, and after the 55th iteration round, the overall loss value curve change becomes stable, and the loss value of the five models is lower than 0.5. In addition, the loss value of the ShuffleNetV2-Lite+SE model is significantly lower than that of the other models, showing a higher classification ability of this model.

Figure 10 shows the classification confusion matrix of the ShuffleNetV2-Lite+SE model for images of the three types of diseases and one type of health. It can be seen that the prediction accuracy of the images of the physiological disease is 96%, the prediction accuracy of the images of the bacterial disease is 93%, and the prediction accuracy of the images of the fungoid disease is 94%. The prediction accuracy of the images of the healthy condition is 97%. Among them, the accuracy of disease image prediction is relatively low for the bactrial and fungoid groups. The reason is that there are problems in the dataset of edible fungi fruit body diseases, such as complex disease image backgrounds, high similarity between some diseases and the background, and significant differences in disease sizes, which cause the model to misjudge. Overall, the ShuffleNetV2-Lite+SE model has a good recognition effect on the four types of images.

5. Conclusions

In this study, the recognition of fruit body diseases of edible fungi was taken as the research object, and the problems of low recognition accuracy, complex model structure, and slow recognition speed of the existing model were addressed based on the improvements of ShuffleNetV2.

(1) The constructed edible fungi fruit body disease dataset contains images of three types of diseases and one type of health condition, with a total of 649 images. After data augmentation, the total number of images in the dataset is 3439.

(2) The CBAM module, ECA module, and SE module are embedded into ShuffleNetV2, respectively, to enhance the information interaction between channels and improve the model's performance. Experimental results indicate that the recognition accuracy of the ShuffleNetV2+SE model on the test set is 2.43 percent higher than that of the original model. Compared with the ShuffleNetV2+CBAM model and the ShuffleNetV2+ECA model, the recognition accuracy of the Shufflenetv2+SE model is improved by 2.43 percent; the number of model parameters of the ShuffleNetV2+SE

model is reduced by 0.1 MB and 0.5 MB, and the average iteration time is reduced by 10 s and 14 s, respectively.

(3) The attention SE module is introduced into the ShuffleNetV2 model, and the structure of the model is optimized to obtain the ShuffleNetV2+SE model, the optimized ShuffleNet-Lite model, and the ShufflenetV2-Lite+SE model, respectively. Experimental results indicate that the accuracy, precision, recall, and the Macro-F1 value of the ShuffleNetV2-Lite+SE model reach 96.19%, 96.43%, 96.07%, and 96.24%, respectively, which is higher than those of other models. Compared with the ShuffleNetV2+SE model with the highest accuracy, our model reduces the number of model parameters by 11.11%, and the average iteration time by 12.77%, so it has lower model complexity than other models.

(4) The ShuffleNetV2-Lite+SE model is compared with representative lightweight convolutional neural networks, including MobileNetV2, MobileNetV3, DenseNet, and EfficientNet. Compared with the best-performing MobileNetV3 model, the parameters of the ShufflenetV2-Lite+SE model are reduced by 68.63%, and the average iteration time is reduced by 31.67%. The experimental results show that the ShuffleNetV2 Lite+SE model has higher accuracy and lower model complexity, and has certain advantages compared to the existing relevant research. It can be deployed on mobile terminal devices to promote real-time and accurate recognition of diseases in edible fungi fruit bodies.

The ShuffleNetV2-Lite+SE model proposed in this study reduces the model complexity and considers the recognition accuracy and speed of edible fungi fruit body diseases, which lays the foundation for the recognition of edible fungi fruit body diseases. The next step will be to integrate the outdoor environment and factory planting image data and attempt to transplant the model to mobile platforms to test its real-time recognition effect on edible fungi fruit body diseases, providing assistance for the prevention and control of edible fungi fruit body diseases.

Author Contributions: The contributors are X.X. and Y.Z. for conceptualization; H.C. for methodology; investigation, D.Y. and L.Z.; H.Y. for formal analysis; X.X. and Y.Z. for investigation/writing—original draft/supervision; X.X. and H.Y. for visualization; H.Y. for writing—review/editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of Jilin Province (YDZJ202201ZYTS544) and the Technology Development Plan Project of Jilin Province (20200403176SF).

Data Availability Statement: Data supporting the findings of this study are available from the corresponding author.

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

References

1. Arman, H.; Shreya, G.; Devarai, S.K. Current trends in health-promoting potential and biomaterial applications of edible mushrooms for human wellness. *Food Biosci.* **2023**, *51*, 102290.
2. Rachele, D.C.; Liam, P.; Teresina, M. A systematic review on drivers influencing consumption of edible mushrooms and innovative mushroom-containing products. *Appetite* **2023**, *182*, 106454.
3. Zhao, J.; Wang, T.; Zhang, C.; Han, X.; Yan, J.; Gan, B. A comparative analysis of the umami taste of five fresh edible mushrooms by simulating the chemical environment of oral digestion in vitro. *LWT* **2023**, *176*, 114522. [CrossRef]
4. Xu, J.; Xu, D.; Hu, Q.; Ma, N.; Pei, F.; Su, A.; Ma, G. Immune regulatory functions of biologically active proteins from edible fungi. *Front. Immunol.* **2023**, *13*, 1034545. [CrossRef] [PubMed]
5. Bhushan, A.; Kulshreshtha, M. The Medicinal Mushroom Agaricus bisporus: Review of Phytopharmacology and Potential Role in the Treatment of Various Diseases. *J. Nat. Sci. Med.* **2018**, *1*, 4–9.
6. Bok, Y.Y.; Ji, O.M.; Lee, O.Y.; Gyun, S.P.; Yeul, J.K.; Sik, K.W. Development trend of the mushroom industry. *J. Mushroom* **2016**, *14*, 142–154.
7. Yin, H.; Yi, W.; Hu, D. Computer vision and machine learning applied in the mushroom industry: A critical review. *Comput. Electron. Agric.* **2022**, *198*, 107015. [CrossRef]
8. Zhou, L.; Guo, S.; Shu, M.; Liang, L.; Li, Y. Study on three major effects and guarantee system for quality management of edible mushroom products. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *185*, 012019. [CrossRef]

9. Changtian, L.; Shuai, X. Edible mushroom industry in China: Current state and perspectives. *Appl. Microbiol. Biotechnol.* **2022**, *106*, 3949–3955.
10. Bellettini, M.B.; Bellettini, S.; Fiorda, F.A.; Pedor, A.C.; Bach, F.; Fabela-Moron, M.F.; Hoffmann-Ribani, R. Diseases and pests noxious to *Pleurotus* spp. mushroom crops. *Rev. Argent. Microbiol.* **2018**, *50*, 216–226. [CrossRef]
11. In, L.D.; Hwan, L.J.; Ho, J.S.; Jong, O.S.; Chui, D.I. Crop Disease Diagnosis with Deep Learning-Based Image Captioning and Object Detection. *Appl. Sci.* **2023**, *13*, 3148.
12. Yasuhito, O. Sustainability perspectives for future continuity of mushroom production: The bright and dark sides. *Front. Sustain. Food Syst.* **2022**, *6*, 447.
13. Li, E.; Wang, L.; Xie, Q.; Gao, R.; Su, Z.; Li, Y. A novel deep learning method for maize disease identification based on small sample-size and complex background datasets. *Ecol. Inform.* **2023**, *75*, 102011. [CrossRef]
14. Xu, L.; Cao, B.; Ning, S.; Zhang, W.; Zhao, F. Peanut leaf disease identification with deep learning algorithms. *Mol. Breed.* **2023**, *43*, 25. [CrossRef]
15. Xu, P.; Fu, L.; Xu, K.; Sun, W.; Tan, Q.; Zhang, Y.; Zha, X.; Yang, R. Investigation into maize seed disease identification based on deep learning and multi-source spectral information fusion techniques. *J. Food Compos. Anal.* **2023**, *119*, 105254. [CrossRef]
16. Edoardo, R.; Cinzia, V.; Alessio, F. Quantile-distribution functions and their use for classification, with application to naïve Bayes classifiers. *Stat. Comput.* **2023**, *33*, 55.
17. Moch, L.; Syaiful, R.H.; Mochammad, H.; Faishol, A.M.; Nikmatus, S.Z. Feature Extraction and Naïve Bayes Algorithm for Defect Classification of Manalagi Apples. *J. Phys. Conf. Ser.* **2022**, *2394*, 012014.
18. Ye, Z.; Zuo, T.; Chen, W.; Li, Y.; Lu, Z. Textual emotion recognition method based on ALBERT-BiLSTM model and SVM-NB classification. *Soft Comput.* **2023**, *27*, 5063–5075. [CrossRef]
19. Lei, X.; Zhang, W. Logistic regression algorithm to identify candidate disease genes based on reliable protein-protein interaction network. *Sci. China Inf. Sci.* **2021**, *64*, 179101. [CrossRef]
20. Archana, S.K.; Srinivasan, R.B.; Prabakar, N.T.; Francis, A.S. A novel method to improve computational and classification performance of rice plant disease identification. *J. Supercomput.* **2022**, *78*, 8925–8945. [CrossRef]
21. Yogeswararao, G.; Malmathanraj, R.; Palanisamy, P. Fractional weighted nuclear norm based two dimensional linear discriminant features for cucumber leaf disease recognition. *Multimed. Tools Appl.* **2022**, *81*, 38735–38755. [CrossRef]
22. Prabu, M.; Chelliah, B.J. An intelligent approach using boosted support vector machine based arithmetic optimization algorithm for accurate detection of plant leaf disease. *Pattern Anal. Appl.* **2022**, *26*, 367–379. [CrossRef]
23. Jitendra, T.V.; Tausif, D.; Tarun, S. Identification of Plant Diseases Using Multi-Level Classification Deep Model. *Int. J. Ambient. Comput. Intell. (IJACI)* **2022**, *13*, 21.
24. Singh, T.P.; Pritee, K.; Tanuja, S.; Apaiajita, O. Trends in vision-based machine learning techniques for plant disease identification: A systematic review. *Expert Syst. Appl.* **2022**, *208*, 118117.
25. Amudha, M.; Brindha, K. Multi Techniques for Agricultural Image Disease Classification and Detection: A Review. *Nat. Environ. Pollut. Technol.* **2022**, *21*, 2165–2175. [CrossRef]
26. Ganbayar, B.; Hyun, N.S.; Ryoung, P.K. Deep learning-based plant classification and crop disease classification by thermal camera. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 10474–10486.
27. Islam, M.A.; Sikder, M.H. A Deep Learning Approach to Classify the Potato Leaf Disease. *J. Adv. Math. Comput. Sci.* **2022**, *37*, 143–155. [CrossRef]
28. Karthik, R.; Dinesh, T.R.; Shivam, B.; Sreejan, C. A novel deep learning architecture for disease classification in Arabica coffee plants. *Concurr. Comput. Pract. Exp.* **2023**, *35*, e7625.
29. Rehman, S.; Khan, M.A.; Alhaisoni, M.; Armghan, A.; Alenezi, F.; Alqahtani, A.; Vesal, K.; Nam, Y. Fruit Leaf Diseases Classification: A Hierarchical Deep Learning Framework. *Comput. Mater. Contin.* **2023**, *75*, 1179–1194. [CrossRef]
30. Wang, Y.; Chen, Y.; Wang, D. Convolution Network Enlightened Transformer for Regional Crop Disease Classification. *Electronics* **2022**, *11*, 3174. [CrossRef]
31. Loti, N.N.A.; Noor, M.R.M.; Chang, S.-W. Integrated analysis of machine learning and deep learning in chili pest and disease identification. *J. Sci. Food Agric.* **2021**, *101*, 3582–3594. [CrossRef]
32. Chen, J.; Chen, J.; Zhang, D.; Sun, Y.; Nanehkaran, Y.A. Using deep transfer learning for image-based plant disease identification. *Comput. Electron. Agric.* **2020**, *173*, 105393. [CrossRef]
33. Bhuiyan, M.A.B.; Abdullah, H.M.; Arman, S.E.; Rahman, S.S.; Al Mahmud, K. BananaSqueezeNet: A very fast, lightweight convolutional neural network for the diagnosis of three prominent banana leaf diseases. *Smart Agric. Technol.* **2023**, *4*, 100214. [CrossRef]
34. Lin, J.; Chen, X.; Pan, R.; Cao, T.; Cai, J.; Chen, Y.; Peng, X.; Cernava, T.; Zhang, X. GrapeNet: A Lightweight Convolutional Neural Network Model for Identification of Grape Leaf Diseases. *Agriculture* **2022**, *12*, 887. [CrossRef]
35. Wang, T.; Xu, H.; Hai, Y.; Cui, Y.; Chen, Z. An Improved Crop Disease Identification Method Based on Lightweight Convolutional Neural Network. *J. Electr. Comput. Eng.* **2022**, *2022*, 1–16. [CrossRef]
36. Bao, W.; Yang, X.; Liang, D.; Hu, G.; Yang, X. Lightweight convolutional neural network model for field wheat ear disease identification. *Comput. Electron. Agric.* **2021**, *189*, 106367. [CrossRef]
37. Chen, Y.; Chen, X.; Lin, J.; Pan, R.; Cao, T.; Cai, J.; Yu, D.; Cernava, T.; Zhang, X. DFCANet: A Novel Lightweight Convolutional Neural Network Model for Corn Disease Identification. *Agriculture* **2022**, *12*, 2047. [CrossRef]

38. Chen, J.; Wang, W.; Zhang, D.; Zeb, A.; Nanehkaran, Y.A. Attention embedded lightweight network for maize disease recognition. *Plant Pathol.* **2020**, *70*, 630–642. [CrossRef]
39. Zeng, W.; Li, M. Crop leaf disease recognition based on Self-Attention convolutional neural network. *Comput. Electron. Agric.* **2020**, *172*, 105341. [CrossRef]
40. Atila, Ü.; Uçar, M.; Akyol, K. Plant leaf disease classification using EfficientNet deep learning model. *Ecol. Inform.* **2021**, *61*, 101182. [CrossRef]
41. Euncheol, K.; Yeongtae, H.; Seok, O.I. Mushroom Image Recognition using Convolutional Neural Network and Transfer Learning. *KIISE Trans. Comput. Pract.* **2018**, *24*, 53–57.
42. Zhang, S.; Yang, H.; Yang, C.; Yuan, W.; Li, X.; Wang, X.; Zhang, Y.; Cai, X.; Sheng, Y.; Deng, X.; et al. Edge Device Detection of Tea Leaves with One Bud and Two Leaves Based on ShuffleNetv2-YOLOv5-Lite-E. *Agronomy* **2023**, *13*, 577. [CrossRef]
43. Türkmen, S.; Heikkilä, J. An efficient solution for semantic segmentation: ShuffleNet V2 with atrous separable convolutions. In Proceedings of the Image Analysis: 21st Scandinavian Conference, SCIA 2019, Norrköping, Sweden, 11–13 June 2019.
44. Wei, J.; Yu, P.; Bo, X.; Juncheng, W. A Real-Time Apple Targets Detection Method for Picking Robot Based on ShufflenetV2-YOLOX. *Agriculture* **2022**, *12*, 856.
45. Li, X.; Wen, Z.; Hua, Q. Vehicle License Plate Recognition Using Shufflenetv2 Dilated Convolution for Intelligent Transportation Applications in Urban Internet of Things. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 3627246. [CrossRef]
46. Chen, Z.; Yang, J.; Chen, L.; Jiao, H. Garbage classification system based on improved ShuffleNet v2. *Resour. Conserv. Recycl.* **2022**, *178*, 106090. [CrossRef]
47. Liu, Z.; Jiang, J.; Du, Y.; Xu, Z. A Band Influence Algorithm for Hyperspectral Band Selection to Classify Moldy Peanuts. *IEEE Access* **2021**, *9*, 147527–147536. [CrossRef]
48. Min, B.; Kim, T.; Shin, D.; Shin, D. Data Augmentation Method for Plant Leaf Disease Recognition. *Appl. Sci.* **2023**, *13*, 1465. [CrossRef]
49. Wan, X.; Zhang, X.; Liu, L. An Improved VGG19 Transfer Learning Strip Steel Surface Defect Recognition Deep Neural Network Based on Few Samples and Imbalanced Datasets. *Appl. Sci.* **2021**, *11*, 2606. [CrossRef]
50. Mateusz, B.; Atsuto, M.; Mazurowski, M.A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw.* **2018**, *106*, 249–259.
51. Parab, M.; Bhanushali, A.; Ingle, P.; Kumar, B.N.P. Image Enhancement and Exposure Correction Using Convolutional Neural Network. *SN Comput. Sci.* **2023**, *4*, 204. [CrossRef]
52. Valentina, M.A.; Mircea, T.R.; Serban, M.; Mihaeia, C. On Spectral-Spatial Classification of Hyperspectral Images Using Image Denoising and Enhancement Techniques, Wavelet Transforms and Controlled Data Set Partitioning. *Remote Sens.* **2022**, *14*, 1475.
53. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. *arXiv* **2017**, arXiv:1707.01083v2.

agronomy

MDPI

Article

Real-Time Detection of Crops with Dense Planting Using Deep Learning at Seedling Stage

Shuolin Kong, Jian Li, Yuting Zhai, Zhiyuan Gao, Yang Zhou and Yanlei Xu *

College of Information and Technology, Jilin Agricultural University, Changchun 130118, China; 20201277@mails.jlau.edu.cn (S.K.); lijian@jlau.edu.cn (J.L.); 20221850@mails.jlau.edu.cn (Y.Z.); gzy@mails.jlau.edu.cn (Z.G.); zhouyang@jlau.edu.cn (Y.Z.)
* Correspondence: yanleixu@jlau.edu.cn

Abstract: Crop seedlings are similar in appearance to weeds, making crop detection extremely difficult. To solve the problem of detecting crop seedlings in complex field environments, a seedling dataset with four crops was constructed in this study. The single leaf labeling method was proposed as an alternative to conventional labeling approaches to improve the detection accuracy for dense planting crops. Second, a seedling detection network based on YOLOv5 and a transformer mechanism was proposed, and the effects of three features (query, key and value) in the transformer mechanism on the detection accuracy were explored in detail. Finally, the seedling detection network was optimized into a lightweight network. The experimental results show that application of the single leaf labeling method could improve the mAP0.5 of the model by 1.2% and effectively solve the problem of missed detection. By adding the transformer mechanism module, the mAP0.5 was improved by 1.5%, enhancing the detection capability of the model for dense and obscured targets. In the end, this study found that query features had the least impact on the transformer mechanism, and the optimized model improved the computation speed by 23 ms·frame^{-1} on the intelligent computing platform Jetson TX2, providing a theoretical basis and technical support for real-time seedling management.

Keywords: crop seedling detection; dense target detection; lightweight transformer; YOLOv5

Citation: Kong, S.; Li, J.; Zhai, Y.; Gao, Z.; Zhou, Y.; Xu, Y. Real-Time Detection of Crops with Dense Planting Using Deep Learning at Seedling Stage. *Agronomy* **2023**, *13*, 1503. https://doi.org/10.3390/agronomy13061503

Academic Editor: Roberto Marani

Received: 28 April 2023
Revised: 18 May 2023
Accepted: 23 May 2023
Published: 30 May 2023

1. Introduction

Precision agriculture, which is aimed at reducing the cost of agricultural production, environmental pollution and the automation of crop management [1], is gaining widespread attention and is being investigated by various agricultural researchers. In precision agriculture, the accurate acquisition of crop status and position is crucial for facilitating precise fertilization, weed control and full automation of crop management [2]. It serves as a key factor in reducing fertilizer waste, excessive herbicide use and controlling costs [3,4]. Crop detection becomes more challenging during the seedling stage, as crop seedlings bear a resemblance to weeds and are more susceptible to death from possible environmental factors. Hence, the design of a model capable of accurately detecting crop seedlings in complex environments becomes increasingly crucial.

In the past, crop detection often focused on individual crop species as targets and utilized conventional visual algorithms for detection [5,6]. Gai et al. [7] used the 2D connected components method, two-dimensional multiscale wavelet transformation and marker-controlled watershed segmentation algorithm to segment broccoli and lettuce. The accuracy for segmenting broccoli and lettuce achieved 92.4% and 96.6%, respectively. Chen et al. [8] extracted Gabor features from corn images and built a support vector machine (SVM) model to learn the Gabor features of corn images for corn detection. In addition, Hamuda et al. [9] used cauliflower HSV spatial images as input for cauliflower detection based on the dilation algorithm and moment method, which obtained a detection accuracy of 99.04%.

With the development of deep learning, convolutional neural networks (CNN) were developed, a vision algorithm with a stronger generalization ability and higher accuracy compared to traditional vision algorithms. CNN neural networks are widely used in precision agriculture and smart agriculture, such as automatic species identification [10], disease identification [11,12] and fruit ripeness analysis [13–15]. Most of the research on crop and weed localization is based on object detection networks, such as You Only Look Once (YOLO) [16–19] series models and Region-CNN (RCNN) series models [20–22]. Zou et al. [23] combined images with and without weeds to generate new weed images, and trained a semantic segmentation network called UNet, obtaining an accuracy of 92.21%. Punithavathi et al. [24] proposed a detection model based on Faster RCNN for crop and weed detection and used the extreme learning machine algorithm to optimize the hyperparameters of the deep learning model to obtain a higher detection accuracy. Chen et al. [25] detected weeds in sesame fields based on the YOLOv4 detection network and used local attention pooling to replace maximum pooling in spatial pyramid pooling and SEnet modules to replace logical modules in local attention pooling. The model obtained a 96.16% detection accuracy. Although these studies have provided a solid theoretical and experimental foundation for crop detection, the development of comprehensive automated crop management is still challenging. One of these challenges is that current studies mainly focus on crops that have grown to a degree where they can be easily distinguished from their surroundings. Another challenge is the precise localization of crops with dense planting (placing multiple seeds in a single planting hole), particularly during the seedling stage.

Additionally, we have identified several issues regarding the labeling process. In previous detection studies, researchers typically labeled the entire crop as the target, as shown in Figure 1a,b. However, this method presented three potential issues.

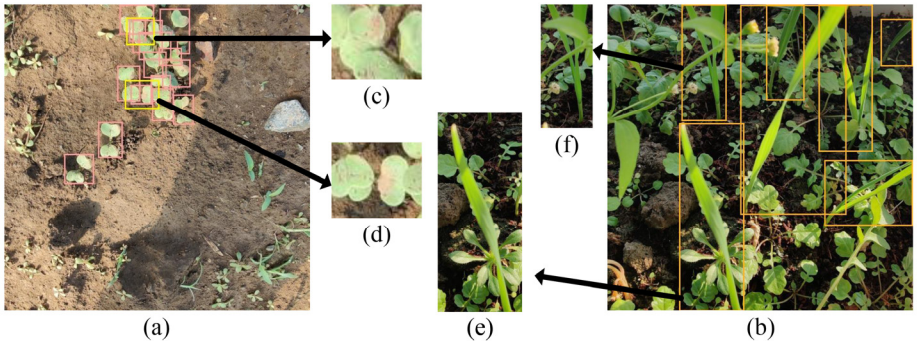


Figure 1. Schematic diagram of the whole crop labeling method. (**a**) Schematic of radish; (**b**) schematic diagram of wheat; (**c**) radish with only one leaf; (**d**) covered radish leaves; (**e**) wheat with only one leaf; (**f**) wheat with two leaves.

The first issue was that because of the dense nature of the crops, as shown in Figure 1d, the worker needs to be meticulous to identify multiple leaves belonging to the same crop. This process will produce errors in labeling. The second issue was that some crops can only be labeled to a single leaf due to occlusion, as shown in Figure 1c. As a result, there was a significant gap in the information about the same target features, making it difficult for the deep learning model to learn the feature patterns and reducing identification accuracy. Lastly, the third issue is that crops exhibit different growth rates, resulting in some crops having only one leaf, while others may have multiple leaves, as shown in Figure 1e,f. This inconsistency also reduced the detection accuracy. Although the whole crop labeling method can aid in identifying each crop, these three scenarios can lead to labeling errors. Additionally, in some cases, there may be a significant difference in the shape of the target, which can result in reduced accuracy for detecting the crop. Therefore, it is important to develop more effective labeling strategies to improve the accuracy of crop detection.

To address the issue of low accuracy in detecting crops from dense crop seedlings in complex environments, we constructed a crop seedling dataset that included crops such as

soybean, wheat, radish and cucumber, grown with different planting methods, sizes and growing environments. In addition, two labeling strategies were proposed in a scientific manner for dense planting crops. Furthermore, we proposed a deep learning structure transformer mechanism applied to computer vision that has been investigated recently. Since the transformer mechanism has been shown to be more resistant to interference and to extract features more rationally [26,27], it improves both the detection and recognition accuracy when combined with convolutional neural networks [28]. Therefore, we took inspiration from these studies that incorporated the transformer mechanism into their CNN-based model and expected to improve the detection accuracy of the model for various crop seedlings in complex environments.

In summary, we carried out the following work:

1. An image dataset was constructed, comprising seedlings from four different crops, all grown in environments with a substantial weed presence. Additionally, a dense planting method was used for some crops.
2. In order to improve the accuracy of the model for the detection of densely grown crops, two labeling strategies were proposed from the perspective of crop type.
3. A detection model for dense targets based on YOLOv5 and the transformer mechanism was proposed from the perspective of model structure.
4. Finally, in order to improve the detection efficiency, the model was lightened and improved based on the impact of three different features in the transformer mechanism on the accuracy.

The workflow diagram of the study is shown in Figure 2.

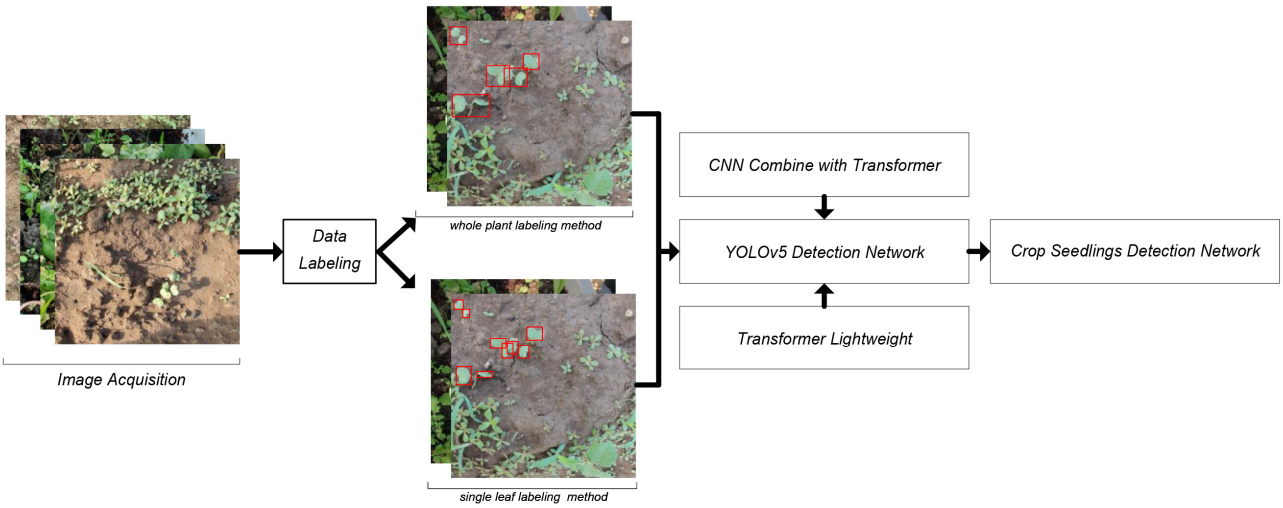


Figure 2. The research flowchart used in this study to detect crops at seedling stage.

2. Materials and Methods

2.1. Image Acquisition

The image capture device was a smartphone Oneplus8P (the manufacturer was BBK Electronics, Shenzhen, China), with a main camera lens of 48 million pixels, and the captured picture pixels were 3000 × 3000. The camera's ISO was set to 400, the color temperature was set to 5000 and the shutter speed was set to 1/50 s. The photographs were taken at a height ranging from 20 to 40 cm. In total, we collected 2140 digital RGB images in JPG format for use in this study.

In order to improve the model's generalization capacity and validate its efficacy in detecting a wide range of crops in complex environmental conditions, we meticulously built a dataset of crop seedlings with our own shooting. These data would be used to train and validate the model.

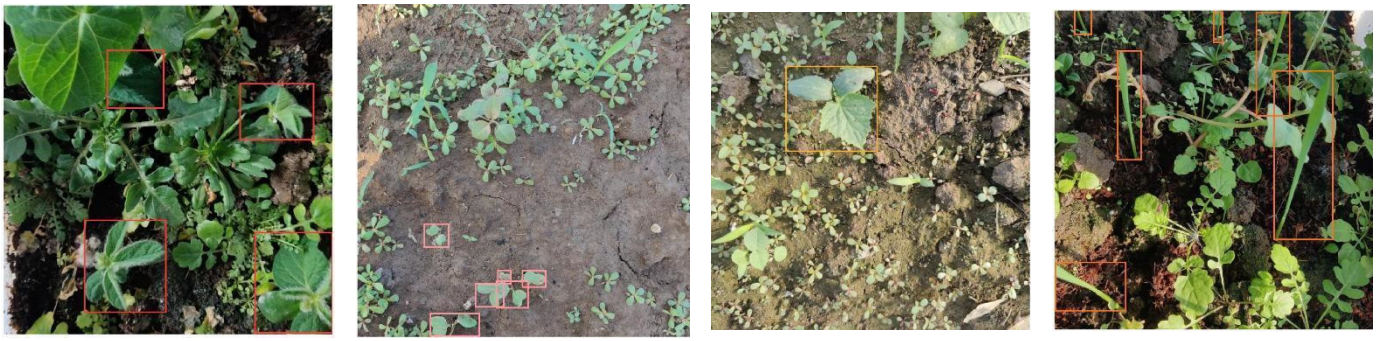
Four crops were targeted, namely soybeans, radishes, cucumbers and wheat. These crops were specifically chosen for two reasons. Firstly, soybean, radish, cucumber and

wheat are representative cash crops. Secondly, the seedling stages of various weed species, such as setaria viridis, eleusine indica, wild pea and petunia, bear resemblance to the seedling stages of the aforementioned crops.

The cultivation of these crops took place within a greenhouse located at Jilin Agricultural University in China. The greenhouse measures 5 m in width and 40 m in length. To create an environment rich in weed species, no weeding was conducted prior to crop planting and throughout the crop cultivation process. Specifically, soybeans and wheat were directly and randomly sown in the field, while cucumbers were planted at 20 cm intervals with 1–3 seeds placed in each planting hole. Radishes were also planted at 20 cm intervals, with more than 5 radish seeds placed in each planting hole. A total of 50 soybean and wheat seeds were planted, 27 cucumber seeds were planted and 30 planting holes were utilized for radishes.

During the seedling stage of the crops, the entire crop was photographed. Specifically, for soybeans, the seedling stage refers to the period from the emergence of cotyledons to the growth of the third true leaf. The same applies to cucumbers. For radishes, the seedling stage specifically refers to the period from the emergence of cotyledons to the growth of the first true leaf. As for wheat seedlings, the seedling stage refers to the period when the entire crop is less than 15 cm in height. Thirty specimens of each crop were planted and photographs were taken every three days starting from when the first pair of true leaves fully unfolded.

The crop seedling data comprised soybean, which was of medium size and planted sparsely, resulting in partial obscurity by weeds (Figure 3a), making detection slightly difficult (the size represents the proportion of the target crop in the entire image). Similarly, radish was small and densely planted, with seedlings not obscured by weeds but covered by other radish seedlings (Figure 3b), making it more challenging to detect. In addition, the cucumber was of medium size, planted sparsely and not shaded by other weeds, resulting in the least difficult detection (Figure 3c). Although wheat was planted sparsely, the seedlings were still be covered by other seedlings, which was the same for that of densely planted radishes (Figure 3d), which greatly increased the difficulty of detection. Detailed information on the data of each crop seedling is shown in Table 1. The dataset was split into training and test sets with the ratio of 0.8:0.2.



(**a**) (**b**) (**c**) (d)

Figure 3. Sample image of four crop seedling detection data. (**a**) Soybean; (**b**) radish; (**c**) cucumber; (**d**) wheat. Note: the color (HEX) of the detection box for soybean, radish, cucumber and wheat is FF3838, FF9D97, FF701F, FFB21D.

Table 1. Crop information and respective image collected in this study.

Type	Number	Size	Cover	Plant Method
Soybean	544	Medium	Partially	Sparse
Radish	515	Small	Mostly	Dense
Cucumber	503	Medium	None	Sparse
Wheat	578	Large	Mostly	Sparse

2.2. Labeling Strategy and Data Enhancement

To match the model input size and address the GPU memory limitations, we first reduced the image resolution from 3000 × 3000 to 640 × 640 using bilinear interpolation. Then, Labelimg 1.8.1 (an open source, efficient image labeling tool) was used to label the crop images and the tag data format was set to VOC format.

On the basis of the whole crop labeling method, we proposed the second labeling strategy as another possible option, referred to it as Strategy B, which utilized the single leaf labeling method, as shown in Figure 4. This labeling method has the following advantages. This method involved labeling a single leaf, thereby effectively reducing the labeling difficulty. Workers only needed to label each individual leaf without considering the logical relationship between the leaves. In addition, this method resulted in a more uniform target morphology and clearer features within the labeling box, particularly in cases of dense planting. By adopting this new labeling strategy, the study was able to overcome the limitations of the whole crop labeling method as stated in the introduction and improve the accuracy in crop detection.

(**a**) (**b**)

Figure 4. Schematic diagram of the leaf labeling method. (**a**) Schematic diagram of radish; (**b**) schematic diagram of wheat.

To avoid model overfitting and to enhance the model performance, the amount of training data was expanded by employing various techniques. Two popular methods are image enhancement and mosaic enhancement. The image enhancement method involves applying random rotation and color dithering to the images. The color dithering operation adjusts various image properties such as image saturation, sharpness, brightness and contrast. The values of saturation and sharpness range from 0 to 3.1 and the values of luminance and contrast range from 1 to 2.1. The mosaic enhancement method takes a different approach by creating a single image from multiple randomly selected images. This process involved flipping and scaling the images, dithering the colors and finally stitching them together into a single image. The mosaic method allows the model to learn multiple images simultaneously, which improves the generalization ability of the model. A schematic diagram of the above two expansion methods is shown in Figure 5. The image enhancement was run online by the program, and it did not affect the original data. The enhancement probability we set was 50%, that is, there was a 50% possibility of enhancing the input image in one round of training.

(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 5. Example of augmentations. (**a**) Original image of cucumber; (**b**) rotating image of cucumber; (**c**) original image of cucumber radish; (**d**) color dithering image of cucumber radish; (**e**) augmentative image with mosaic method.

2.3. Crop Seedlings Detection Network

The YOLOv5 is a single-stage target detection model, which is the most mature model in the YOLO series of models [29]. Compared to YOLOv3 [30] and YOLOv4 [31], YOLOv5 achieves a higher computational accuracy and minimal arithmetic power consumption. The model is continuously updated, making it more mature compared to YOLOv7 [32] and it outperforms its predecessors in various fields [33,34]. Therefore, YOLOv5 was used as the base detection network in this study.

The YOLOv5 structure consists of three main parts. Through these three structures, the model could effectively extract image features and locate key areas. The schematic diagram of the model structure is shown in Figure 6. The first part employed an improved cross-stage partial network (CSPNet) as the backbone network to extract the image-based feature information [35]. To improve the feature extraction, the faster feature extraction C3 module and the spatial pyramid pooling (SPP) that unifies multiple scale feature maps into the same size were added to CSPNet [20]. The second part was the neck network that extracted and processed the base image feature information output from the backbone network. To accomplish this, it used the feature pyramid network (FPN) structure [36], which consists of several neck blocks that perform computation and feature fusion. The neck blocks were responsible for aggregating the feature map information in the previous layer of neck blocks or the backbone network. The third component of the model was the detection head. Its primary purpose is for calculating relevant information of the detected targets in the detection box, such as the confidence level and the length and width adjustment value of the detection box. The detection head comprises three sub-components, each with a different feature map size, 76 × 76, 38 × 38, and 19 × 19, respectively. These sub-components detect small-sized, medium-sized and large-sized objects, respectively.

Transformer Neck Block

The transformer mechanism is a cutting-edge deep learning model with strong feature generalization and global feature extraction capabilities. The structure of the transformer mechanism is shown in Figure 7 and is comprised of several key components. Firstly, the Layer Norm represents the normalization process by layer, ensuring that the data are consistent and free from unwanted variations. Then, multi-head attention is the core of the transformer mechanism and is responsible for extracting global feature information. Lastly, the multilayer perceptron layer (MLP) module consists of two fully connected layers and an SiLU activation function, facilitating the transformation of the extracted features into a usable form.

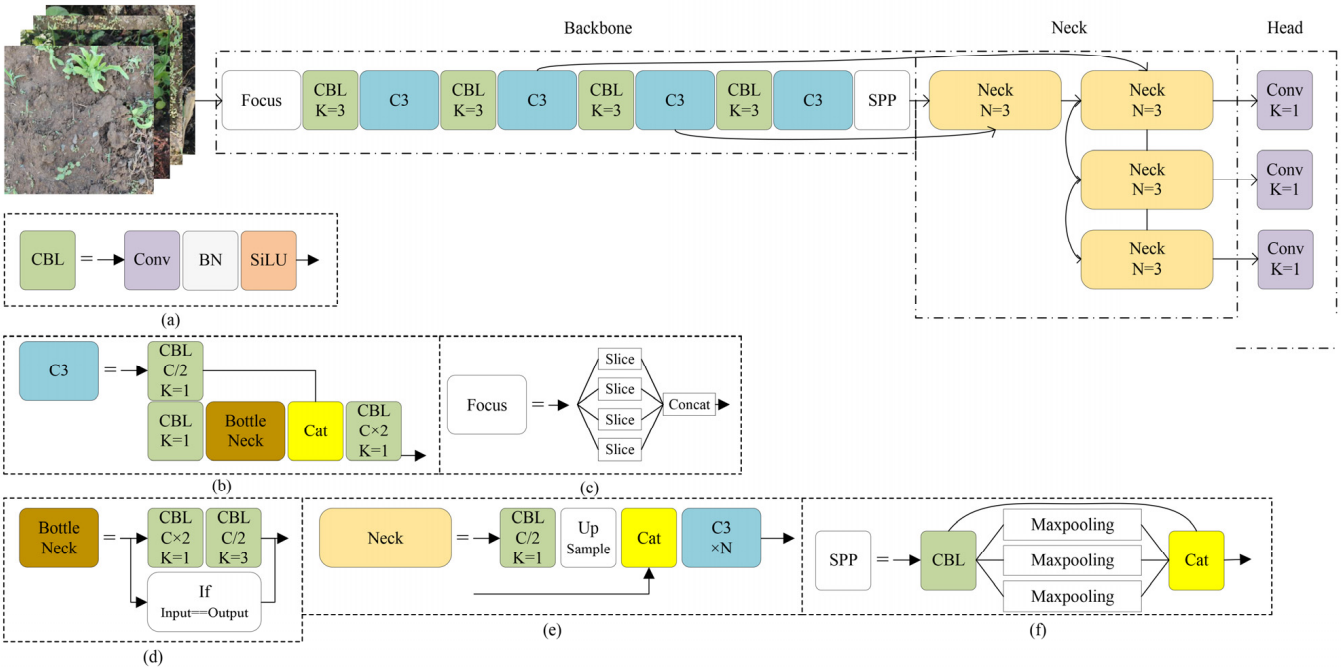


Figure 6. Diagram of the YOLOv5 detection network structure. Where, Conv is the convolutional layer, BN is the batch normalization layer and Cat is the concatenation operation. K is the convolutional kernel size, C is the number of channels and the step size of the convolutional kernel in the CBL module is 2. The second C3 block, the fourth C3 block and the SPP block are repeated 7, 1 and 3 times, respectively. (**a**) CBL module for downsampling operation; (**b**) C3 module is mainly responsible for extracting image features; (**c**) the Focus module is a downsampling module with no parameters; (**d**) BottleNeck is the module responsible for extracting features in the C3 module; (**e**) the Neck module is used to fuse features of different sizes; (**f**) SPP module can standardize the dimensions of feature drawings.

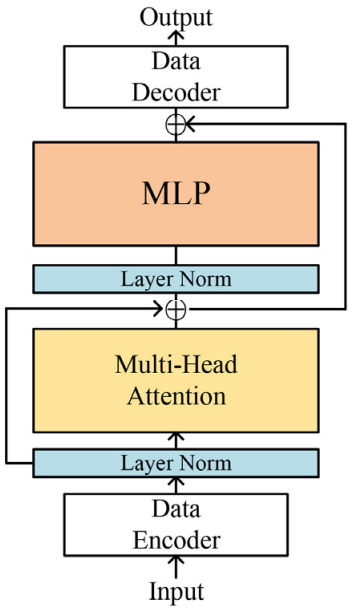


Figure 7. Diagram of the transformer mechanism structure in vision transformer model.

The transformer mechanism's abilities to extract global feature extraction relies on multi-head attention to compute the similarity between each image patch. The computation process of multi-head attention involves several steps. First, the input data pass through

the MLP to obtain three distinct feature sets, as shown in Equation (1), which are named Q, K and V, representing the query, key and value of the input image, respectively. Next, the similarity between Q and K is calculated as shown in Equation (2), where dk is the length of feature K and the Softmax function is used to obtain the weight value α of the similarity. Finally, the weight α is multiplied by V to output the features with global feature information, as shown in Equation (3).

$$Q, K, V = MLP(Input) \tag{1}$$

$$\alpha = Softmax(\frac{Q \cdot K}{\sqrt{d_K}}) \tag{2}$$

$$Output = \alpha \cdot V \tag{3}$$

This study aims to improve the information extraction capability of the model for dense targets. To achieve this, we incorporated the transformer mechanism into the YOLOv5 detection network because of its ability to extract global information. However, we found that adding the transformer mechanism directly to the backbone network resulted in high computational consumption and a weak inductive bias capability. To address this, we drew inspiration from previous studies that combined convolutional neural networks and the transformer mechanism [37,38]. Specifically, we added the transformer mechanism at the end of the neck of YOLOv5 to improve the detection capability of YOLOv5 for dense crop seedlings by using convolutional neural networks to bias the induction of features and reduce the number of feature channels. This approach helped to overcome the problem of weak bias induction and high computation of the transformer mechanism, while improving the detection efficiency. The improved neck block structure is shown in Figure 8.

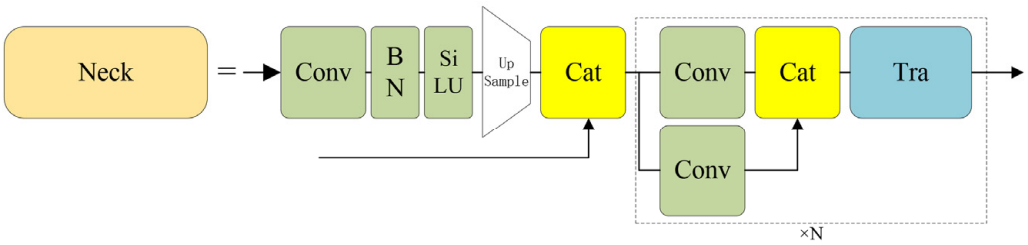


Figure 8. Diagram of the improved neck structure.

While the transformer mechanism has the potential to extract global feature information, which can enhance the model's ability to detect dense targets, it is prone to computational overload and has a tendency to overfit on small datasets [39]. In addition, it is worth noting that the transformer mechanism originated from the transformer model in the domain of natural language processing. In the case of text data, each piece of text has a distinct query, key and value due to the dense feature of the information contained in the text. Due to the sparse nature of the image information, it lacks the explicit query, key and value information that is necessary for the transformer mechanism to operate efficiently. As a result, Q, K and V for image data are simply features with varying information used by the transformer mechanism. However, the extraction of such information requires significant computational power, which is the primary reason for the slower processing speed of the transformer mechanism. In this study, we explored the impact of the three key features, i.e., Q, K and V, on the accuracy of the transformer mechanism. Due to the lack of clarity regarding the information carried by these features, we replaced them with the original input data. This approach allowed us to reduce the computational effort of the model while still examining the effect of these features on the accuracy. We tested two models that use a transformer mechanism, ViT [40] and Swim Transformer [41]. We investigated three types of replacement: replacement Q, replacement V and the simultaneous replacement of Q and V. Since the length of K is identical to that of Q, it is possible to substitute Q and K for each

other when calculating α. As a result, no experiments were conducted in this study for replacement K.

2.4. Training Settings and Hyperparameters

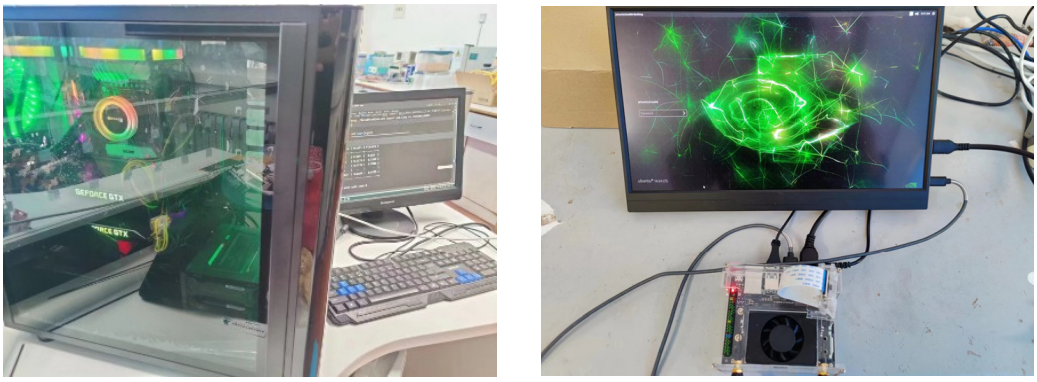
The optimization function used to train the model in this paper was the SGD function. The classification loss used the focal loss function to address the issue of imbalanced data caused by the higher number of crops in the densely grown crop images compared to the sparsely grown crops [42]. The detection loss was calculated using the IOU function [43], and the NMS (Non-Maximum Suppression) method [44] was used to filter redundant detection boxes. To effectively accelerate the model learning speed, we applied the OneCycleLR method for the learning rate. To obtain the optimal detection results of the model on the crop seedling dataset of this paper, we fine-tuned the hyperparameters of YOLOv5, as shown in Table 2. Among them, depth and width, which control the depth and width of the model, were kept constant. To obtain the optimal hyperparameters, the model underwent 500 training iterations, each consisting of 30 epochs. After each training iteration, we adjusted the hyperparameters and finally obtained the optimal set of hyperparameters with the highest accuracy during testing.

Table 2. Hyperparameters used in the training process.

Training Parameters	Values
depth	0.33
width	0.50
initial learning rate	0.009
final learning rate	0.071
momentum	0.95
optimizer weight decay	0.00045
focal loss alpha	0.5
focal loss γ	2

2.5. Experimental Equipment

The experiments were conducted on a Windows 10 operating system, using an NVIDIA Titan X GPU and Intel Xeon E5-2696 v3 CPU to train the deep learning models in PyTorch 1.7.0 with Python 3.6. The experimental device utilized for training the model is shown in Figure 9a.



(**a**) (**b**)

Figure 9. Experimental device. (**a**) Experimental device utilized for training the model; (**b**) Jetson TX2.

The Jetson TX2 (as shown in Figure 9b) operating system was Ubuntu 18.04 using the deep learning framework Pytorch 1.8. The CPU was a CPU cluster consisting of a dual-core Denver2 processor and a quad-core ARM Cortex-A57, with 4 GB of LDDR4 memory and a 256-core Pascal GPU.

2.6. Performance Evaluation

To measure the performance of each trained model on the testing set, evaluation indicators were used in this study. Specifically, we evaluated the model performance by single species average precision (AP), mean detection precision (mAP), floating-point operations per second (FLOPs), computational speed of the training platform (server speed) and the computational speed of the carrying platform (TX2 speed). The unit of speed was ms·frame^{-1}, which was how long it took to calculate a frame. The AP was determined by precision and recall, the formula of precision is shown in Equation (4) and the formula of recall is shown in Equation (5), where *TP* is true positive, *FP* is false positive and *FN* is false negative. *TP*, *FP* and *FN* are determined by the IOU threshold. The IOU value is the overlap area between the detection box calculated by the model and the manually labeled detection box.

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

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

The formula for calculating AP is shown in Equation (6) and the value of the IOU was taken as 0.5 when evaluating the AP for a single crop. The mAP is the mean value of AP for the four species. The mAP had different IOU values. The mAP was named mAP0.5 when the value of the IOU was 0.5. The mAP0.5–0.95 was the mean value of the AP for the four species between IOU thresholds from 0.5 to 0.95 at intervals of 0.05 to calculate the AP. This index represented a more rigorous evaluation of the detection accuracy.

$$AP_i = \int_0^1 p_i(r_i)dr \tag{6}$$

Floating-point operations (FLOPs) measure the computational memory consumed by the model for each operation in the convolutional and fully connected layers during forward propagation of the model. The formula for calculating the FLOPs is displayed in Equation (7), where *Ci* is the convolutional layer input channel and *K* is the convolution kernel size. *H* and *W* are the height and weight of the convolutional layer output feature map, respectively, and *Co* is the output channel. *I* and O are the input and output numbers in the fully connected layers, respectively. The unit of *FLOPs* is *G* (Giga)

$$FLOPs = [\sum_{conv=1}^{n} (2C_I K^2 - 1)HWC_O] + [\sum_{full=1}^{n} (2I - 1)O] \tag{7}$$

3. Results

3.1. Results of Two Strategies and Transformer

The results obtained from the models on the testing set are shown in Table 3. In strategy A, the AP of the model with the transformer was improved for soybean (results of YOLOv5, YOLOv5ViT and YOLOv5ST: 90.0–91.4–91.9%), radish (77.3–80.7–81.1%) and cucumber (91.9–94.3–95.3%), demonstrating that the transformer mechanism could enhance the accuracy of the model for both dense and sparse sowing densities. However, we found that the AP for wheat was not significantly improved or even reduced. This is because wheat is planted densely and the wheat leaves are slender in shape, resulting in each detection box containing too many background features. As a result, the model was unable to focus on the region of interest of wheat.

This result also confirms our hypothesis that a large discrepancy in the feature information within the detection boxes of objects of the same category can lead to a reduction in precision. In strategy B, the same high AP was maintained for both sparsely planted cucumbers and soybeans, and an even higher AP was achieved for the model with the transformer mechanism. With strategy B, the detection box size of wheat was reduced, making each detection box contain only one leaf and reducing the feature gap in the detection box of the wheat. However, the characteristic of wheat leaves still caused too many

background features to be included in the box, resulting in more dense detection boxes and a lower wheat AP for YOLOv5. Nonetheless, we observed that the model with transformer mechanism significantly improved the detection accuracy of wheat relative to strategy A (results of strategy A: 73.2–73.7–72.9%; results of strategy B: 72.7–74.0–73.1%), indicating that the transformer mechanism is well-suited for detection in complex environments and medium-sized targets.

Table 3. The detection results of the whole crop labeling (strategy A) method and the single leaf labeling method (Strategy B) on the testing dataset.

Strategy	Model	mAP0.5 (%)	mAP0.5–0.95 (%)	AP (%)				FLOPs (G)	Server Speed (ms·frame^{-1})	TX2 Speed (ms·frame^{-1})
				Soybean	Radish	Cucumber	Wheat			
A	YOLOv5	83.1	52.8	90.0	77.3	91.9	73.2	4.7	13	106
	YOLOv5 ViT	85.0	53.8	91.4	80.7	94.3	73.7	5.3	17	145
	YOLOv5 ST	85.3	53.7	91.9	81.1	95.3	72.9	18.0	19	184
B	YOLOv5	84.3	52.6	91.4	81.9	91.3	72.7	4.7	13	106
	YOLOv5 ViT	85.8	53.3	90.9	83.9	94.2	74.0	5.3	17	145
	YOLOv5 ST	85.9	53.9	91.4	84.3	94.9	73.1	18.0	19	184

The training accuracy curves for YOLOv5, YOLOv5 ViT and YOLOv5 ST are shown in Figure 10. The curves showed that all three models achieved convergence after about 100 rounds. Because the underlying networks were based on YOLOv5, the curves displayed similar trends across the three models, as shown in Figure 10a. The model using strategy A had a large oscillation of the accuracy curve, especially before 100 epochs, as shown in Figure 10b. The large oscillation represents the difficulty of the model to learn the feature patterns, thereby highlighting the issue of a large gap in the information of the target individual features that existed in the overall labeling method. The model using strategy B, as shown in Figure 10c, had a lower initial accuracy due to the greater number of targets to be detected in each image. However, the gap in each individual feature's information was relatively small, the oscillation amplitude was lower and the accuracy improved more rapidly.

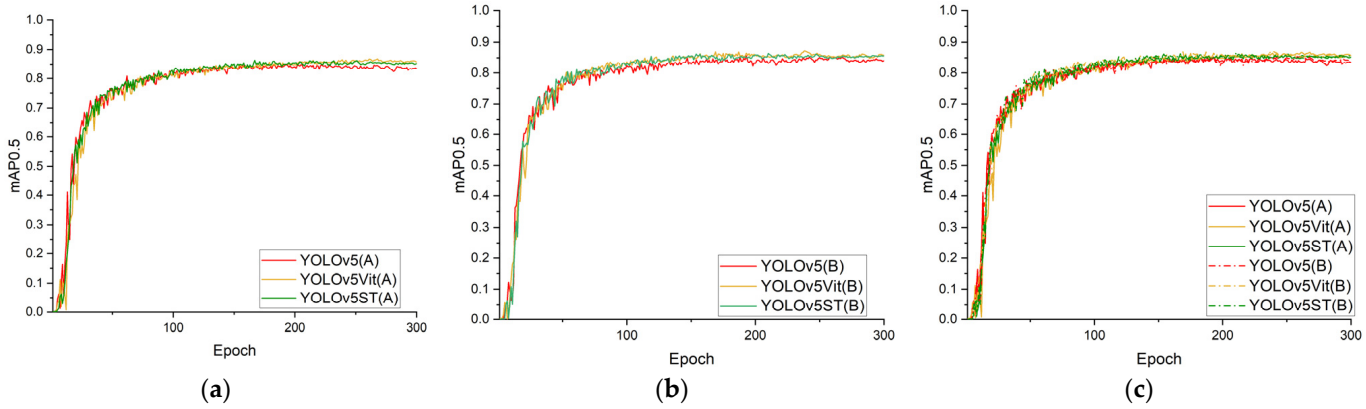


Figure 10. Training curves for the two strategies. (**a**) Comparison of strategy A; (**b**) comparison of strategy B; (**c**) comparison of the two strategies simultaneously.

Although the mAP of strategy B did not show a significant improvement compared to strategy A, it has practical advantages in detecting individual crops. As shown in the detection map of radishes in Figure 11, strategy B provided information on each crop, even if it was only partial, as shown in Figure 11e–g. In contrast, strategy A produced

missed individuals as shown in the detection images b, c and d. Therefore, despite the slight difference in the mAP, strategy B can provide more comprehensive and accurate information on individual crops, resolving the problem of difficult feature learning due to strategy A.

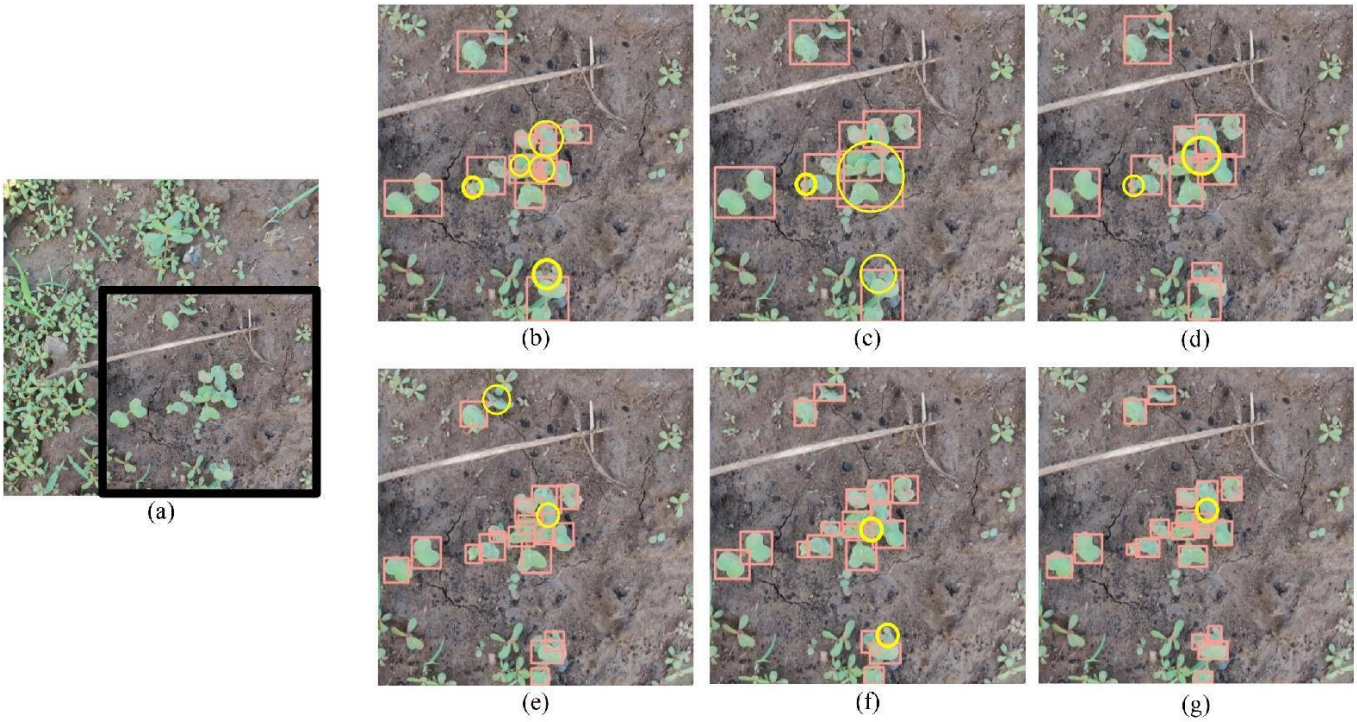


Figure 11. Examples of radish detection. (**a**) Original radish image; (**b**) detection image of YOLOv5 with strategy A; (**c**) detection image of YOLOv5ViT with strategy A; (**d**) detection image of YOLOv5ST with strategy A; (**e**) detection image of YOLOv5 with strategy B; (**f**) detection image of YOLOv5ViT with strategy B; (**g**) detection image of YOLOv5ST with strategy B. Note: yellow circled targets are missed.

The results of the wheat assay are shown in Figure 12. Although the whole crop labeling method was used (strategy A), multiple detection boxes for one crop were produced in the red circles (Figure 12b,c). This phenomenon is more akin to the single leaf labeling method. In addition, the phenomenon observed in Figure 12b highlighted the potential issue that crops with different numbers of leaves can lead to box selection confusion problems. The results in Figure 12c highlighted another potential issue of the presence of crop shading, as a result, the model may interpret two parts of the same crop as two crops. Furthermore, it was also verified that when the background within the detection box was too complex, it could interfere with the model's ability to effectively identify and classify the crop. Therefore, it was verified that the single leaf tagging method is more suitable for dense crop detection. These findings prove that the single leaf marker method may be more suitable for crop detection in high density fields.

In Figure 12e, the object detection results produced by the YOLOv5 model followed strategy B but did not frame the obscured parts (indicated by blue circles). While the detection model with the transformer successfully framed the two parts that led to a separation due to occlusion in one box, verifying that the transformer effectively provides global feature extraction capability.

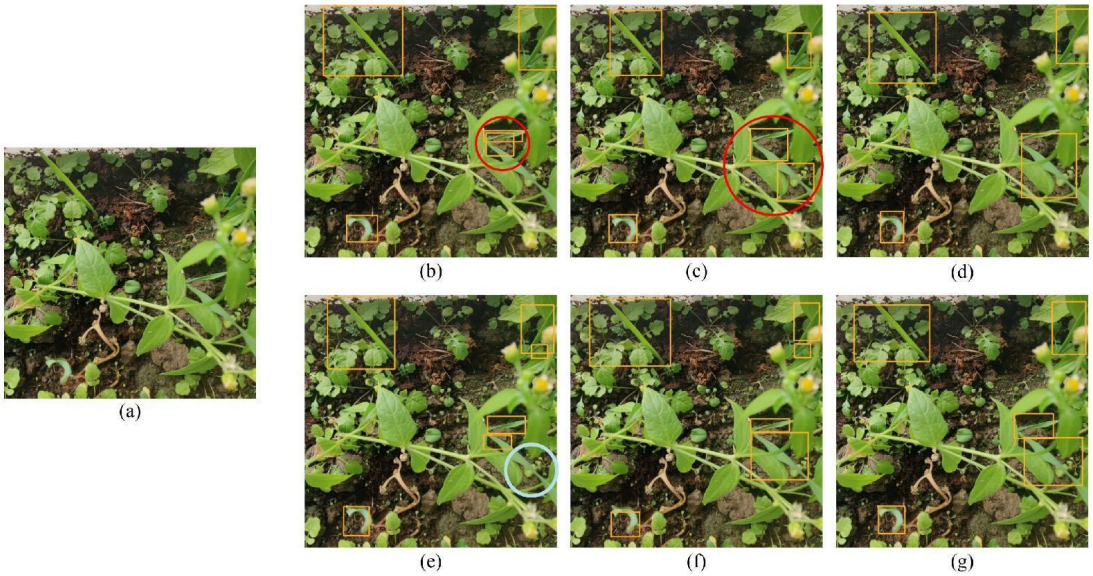


Figure 12. Examples of wheat detection. (**a**) Original wheat image; (**b**) detection image of YOLOv5 with strategy A; (**c**) detection image of YOLOv5ViT with strategy A; (**d**) detection image of YOLOv5ST with strategy A; (**e**) detection image of YOLOv5 with strategy B; (**f**) detection image of YOLOv5ViT with strategy B; (**g**) detection image of YOLOv5ST with strategy B. Note: red circled targets are not in compliance with the strategy. Blue circled targets are the missing parts.

For sparsely grown crops such as cucumbers and soybeans (Figure 13), even though the background was more complex, most of the models achieved a higher detection accuracy and produced satisfactory detection results even without any additional improvements. These findings suggested that there is no need to use strategy B for fields with low plant density during detection.

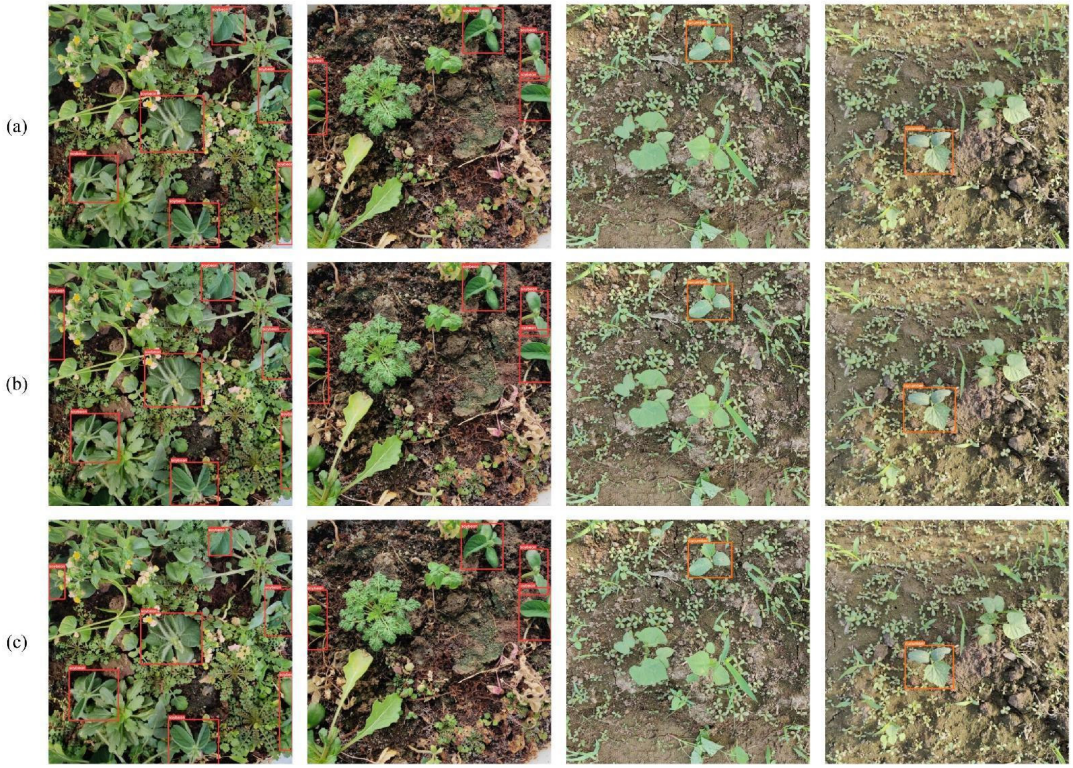


Figure 13. Examples of cucumber and soybean detection. (**a**) Detection image of YOLOv5; (**b**) detection image of YOLOv5ViT; (**c**) detection image of YOLOv5ST.

3.2. Results of the Different Transformer Detection Head

The results of eliminating the feature extraction module in the transformer mechanism are shown in Table 4.

Table 4. Detection results of the lightweight transformer on the testing dataset.

Strategy	Model	Eliminate Component	mAP0.5 (%)	mAP0.5–0.95 (%)	AP/%				FLOPs (G)	Server Speed (ms·frame^{-1})	TX2 Speed (ms·frame^{-1})
					Soybean	Radish	Cucumber	Wheat			
A	YOLOv5 ViT	-	85.0	53.8	91.4	80.7	94.3	73.7	5.3	17	145
		Q	84.2	51.9	90.9	79.9	93.2	72.6	5.0	16	128
		V	84.4	52.1	91.4	80.1	94.0	72.2	5.0	16	128
		QV	83.7	51.5	89.0	79.3	93.7	72.6	4.9	15	114
	YOLOv5 ST	-	85.3	53.7	91.9	81.1	95.3	72.9	10.7	19	184
		Q	84.4	52.3	90.7	81.1	94.7	71.1	9.4	18	164
		V	85.0	53.8	91.4	80.7	94.3	73.7	9.4	18	164
		QV	84.2	51.9	90.9	79.9	93.2	72.6	8.1	17	149
	YOLOv5	-	84.4	52.1	91.4	80.1	94.0	72.2	4.7	13	106
B	YOLOv5 ViT	-	85.8	53.3	90.9	83.9	94.2	74.0	5.3	17	145
		Q	84.6	52.9	91.2	79.9	93.7	73.7	4.9	16	128
		V	85.0	52.1	90.2	82.4	93.2	74.1	4.9	16	128
		QV	83.2	51.9	90.2	78.1	92.6	72.1	5.1	15	114
	YOLOv5 ST	-	85.9	53.9	91.4	84.3	94.9	73.1	10.7	19	184
		Q	83.8	52.1	89.7	79.2	93.5	72.8	9.4	18	164
		V	84.1	52.8	90.3	80.1	93.1	73.0	9.4	18	164
		QV	83.1	51.1	89.4	78.7	92.6	71.8	8.1	17	149
	YOLOv5	-	84.3	52.6	91.4	81.9	91.3	72.7	4.7	13	106

Benefiting from the window multi-head self-attention structure of the swim transformer, the mAP0.5 of YOLOv5ST was the highest among strategies A and B. Meanwhile, the elimination of feature blocks made the model faster. The elimination of V increased the computational speed to 164 ms·frame^{-1}, while the computational speed after eliminating Q and V increased to 149 ms·frame^{-1}. However, with the elimination of feature blocks, the mAP0.5 of YOLOv5ST decreased significantly. In strategy B, the computational accuracy of YOLOv5ST with feature V removed was reduced by 1.8%, and the simultaneous removal of Q and V led to a reduction in the computational accuracy of 2.8%.

The model robustness was higher for ViT with a simple structure. In strategy B, YOLOv5ViT reduced the mAP0.5 by 0.8% after eliminating V, which was smaller than that of YOLOv5ST. The simultaneous elimination of Q and V led to a reduction of 2.6% in the mAP0.5, which was similar to the reduction in YOLOv5ST. This demonstrated that the simultaneous removal of Q and V led to a reduction in the global feature extraction capability of the transformer. Although the improvement of the YOLOv5ViT computation speed was not as obvious as YOLOv5ST, which was only 23 ms·frame^{-1} faster on TX2, the YOLOv5ViT computation speed remained lower due to the relatively simple structure. The impact of removing the feature extraction module on the AP of cucumber was minimal. This was because the cucumber target was more prominent and the environment had fewer similar weeds without occlusions, indicating that YOLOv5 can detect this class of simple targets with a higher AP even without using the transformer mechanism.

In summary, YOLOv5ViT was the least sensitive to the effects of feature removal and had a faster speed, so YOLOv5ViT with V removed was used for further comparisons with other YOLO series models.

3.3. Comparison with Other YOLO Models

The results of each series of YOLO models are shown in Table 5. The YOLOv5 model used in this study had smaller FLOPs than the latest YOLOv7 model, which were slightly faster to compute in the server. Since there was no structural reparameterization in YOLOv5,

the computation speed in TX2 was faster than that in YOLOv7. This was the reason why we chose YOLOv5, a more mature model. The model with the transformer mechanism proposed in this paper had the same accuracy in the mAP0.5–0.95 as YOLOv7 in strategy A. Although the mAP0.5 of soybean and cucumber in strategy B was slightly lower than that of YOLOv7, both remained above 90%. The faster computation speed of our model on TX2 makes it more suitable for future embedding of the algorithm into small intelligent platforms for automated crop management applications.

Table 5. The detection results of different YOLO models on the testing dataset.

Strategy	Model	mAP0.5 (%)	mAP0.5–0.95 (%)	AP (%)				FLOPs (G)	Server Speed (ms·frame^{-1})	TX2 Speed (ms·frame^{-1})
				Soybean	Radish	Cucumber	Wheat			
A	YOLOv3	53.7	36.3	65.9	37.2	79	36.7	24.5	38	207
	YOLOv4	63.6	41.4	75.4	52.1	87.7	39.1	22.3	27	244
	YOLOv5	83.4	50.6	90	77.3	91.9	73.2	4.7	13	106
	YOLOX	83.2	49.9	90.6	79.1	91.8	71.3	15.2	22	121
	YOLOv7	83	52.1	90.4	74.6	92.1	72.9	26.7	15	141
	Ours	84.4	52.1	91.4	80.1	94	72.2	5.0	16	128
B	YOLOv3	53.9	35.5	64.7	38.4	78.9	33.5	24.5	38	207
	YOLOv4	63.3	40.7	75.7	52.6	88.7	36.3	22.3	27	244
	YOLOv5	84.3	52.6	91.4	81.9	91.3	72.7	4.7	13	106
	YOLOX	83.6	51	90.6	80.2	92.8	70.9	15.2	22	121
	YOLOv7	84.9	52.8	91.1	80.5	94.6	73.5	26.7	15	141
	Ours	85	52.1	90.2	82.4	93.2	74.1	4.9	16	128

4. Discussion

Accurate crop detection during the seedling stage enables the reduction in the damage inflicted by agricultural robots on crops, while simultaneously enhancing the efficiency of tasks such as fertilization and weed control. Research on crop detection using deep learning is quite extensive, such as Zou et al. [23] who employed advanced image generation techniques and captured crop images with complex backgrounds. However, their segmentation model required additional algorithmic processing to obtain the precise location of the crops. Furthermore, their computational speed was reported as 51 ms, whereas our proposed model achieved a faster computation speed of only 16 ms. Chen et al. [8] detected weeds in sesame fields based on the YOLOv4 detection network, but only one crop species was targeted. The same problem also occurred in the study by Hamuda [9] et al. The study conducted by Punithavathi et al. [24] utilized a dataset consisting of six crop species and eight weed species. However, the crops in this dataset were grown in an environment with a sparse weed presence and their computational speed was reported as 43 s, which was significantly slower compared to our model.

In our study, we also observed that the use of the single leaf labeling method resulted in a lower mAP0.5–0.95 accuracy compared to the whole crop labeling method. We speculate that this discrepancy may be attributed to the denser distribution of targets within the images when using the single leaf annotation method, which challenged the model's ability to achieve a higher precision detection. However, the model still demonstrated a good approximation of the target position, leading to improved accuracy in mAP0.5.

In the experiments eliminating Q and V, eliminating V had the lowest influence on the accuracy, while eliminating both Q and V had the highest influence on the accuracy, as shown in Table 4. The transformer computed the Q, K and V features of the input data, calculated the similarity weights between Q and K, and multiplied these similarity weights with V to obtain the global characteristics of V. Therefore, the results can show that the reason for the decrease in accuracy due to the simultaneous elimination of Q and V may have been that the transformer lost the ability to acquire global characteristics because the

similarity weights carried the information of V (at this time, both Q and V were equal to the input data) due to our elimination of Q when calculating the similarity weights between Q and K.

In addition, we noticed that some studies used deep learning techniques for the simultaneous detection of crops and weeds. However, this paper detected crops only. This decision was made because including weeds as additional targets in environments with dense weeds would result in imbalanced data for each class, leading to decreased model accuracy. Furthermore, labeling dense weeds is an impractical task. Moreover, weed control tasks do not require the use of deep learning or machine learning for weed detection. Once the position of crops is obtained, simple threshold segmentation methods can be used to detect weeds. Finally, by utilizing the coordinates of crops, avoidance strategies can be implemented to prevent accidental harm to crops during weed control operations.

This study contributes to the detection of dense crops and the lightweight optimization of the transformer, providing accurate positioning of crop seedlings before fertilization and spraying in robotic systems. And part of the crop seedling data and the actual detection video have been made public, which can be found at the following link 'https://github.com/xiaozhi101/crop_detection' (accessed on 27 April 2023). However, this study focused on the rectangular detection boxes, resulting in an insignificant improvement in wheat detection. Therefore, detection algorithms that can frame polygonal detection boxes will be investigated in the future to accurately detect crops with elongated leaves.

5. Conclusions

In this study, a target detection network was proposed for the efficient localization of crop seedlings in complex environments. The target detection network consisted of a YOLOv5 network and transformer module. First, to improve the accuracy and efficiency of the model for dense seedling detection, two labeling strategies, the whole crop labeling method (strategy A) and the single leaf labeling method (strategy B), were proposed. The results showed that the mAP0.5 could be improved from 83.1% to 84.3% using the whole crop labeling method, and from 77.3% to 81.9% for radishes (dense target). Second, the addition of the transformer module improved the mAP0.5 from 83.1% to 85% in strategy A and from 84.3% to 85.8% in strategy B, which effectively improved the detection precision of the model for complex environments and dense targets. Finally, the process of lightweight improvement of the transformer module revealed that feature extraction module V had the least impact on the features. By eliminating V, the computation speed was reduced by 1 ms·frame^{-1} in the server and 17 in the minicomputer TX2, and the mAP0.5 was reduced by only 0.6%, offering the possibility of real-time management of crop seedlings. Such results are particularly beneficial for complex environments and dense targets, highlighting the effectiveness of the transformer module in enhancing the performance of the model.

Author Contributions: Conceptualization, Y.X. and S.K.; Methodology, S.K. and J.L.; Validation, S.K. and Y.X.; Writing—Original Draft Preparation, Y.Z. (Yuting Zhai); Writing—Review & Editing, Y.Z. (Yang Zhou) and Y.X.; Supervision, Z.G. and Y.X.; Funding Acquisition, Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financially supported by the Jilin Provincial Science and Technology Development Plan Project (Grant No. 20230202035NC), and the Science and Technology Development Plan Project of Changchun (Grant No. 21ZGN28).

Data Availability Statement: Not applicable.

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

References

1. Patricio, 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]
2. Jha, K.; Doshi, A.; Patel, P.; Shah, M. A comprehensive review on automation in agriculture using artificial intelligence. *Artif. Intell. Agric.* **2019**, *2*, 1–12. [CrossRef]

3. Fabregas, R.; Kremer, M.; Schilbach, F. Realizing the potential of digital development: The case of agricultural advice. *Science* **2019**, *366*, eaay3038. [CrossRef]
4. Tudi, M.; Ruan, H.D.; Wang, L.; Lyu, J.; Sadler, R.; Connell, D.; Chu, C.; Phung, D.T. Agriculture Development, Pesticide Application and Its Impact on the Environment. *Int. J. Environ. Res. Public Health* **2021**, *18*, 1112. [CrossRef] [PubMed]
5. Cubero, S.; Aleixos, N.; Molto, E.; Gomez-Sanchis, J.; Blasco, J. Advances in Machine Vision Applications for Automatic Inspection and Quality Evaluation of Fruits and Vegetables. *Food Bioprocess Technol.* **2011**, *4*, 487–504. [CrossRef]
6. Burgos-Artizzu, X.P.; Ribeiro, A.; Guijarro, M.; Pajares, G. Real-time image processing for crop/weed discrimination in maize fields. *Comput. Electron. Agric.* **2011**, *75*, 337–346. [CrossRef]
7. Gai, J.Y.; 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.* **2020**, *37*, 35–52. [CrossRef]
8. Chen, Y.J.; Wu, Z.N.; Zhao, B.; Fan, C.X.; Shi, S.W. Weed and Corn Seedling Detection in Field Based on Multi Feature Fusion and Support Vector Machine. *Sensors* **2021**, *21*, 212. [CrossRef]
9. Hamuda, E.; Mc Ginley, B.; Glavin, M.; Jones, E. Automatic crop detection under field conditions using the HSV colour space and morphological operations. *Comput. Electron. Agric.* **2017**, *133*, 97–107. [CrossRef]
10. Garibaldi-Marquez, F.; Flores, G.; Mercado-Ravell, D.A.; Ramirez-Pedraza, A.; Valentin-Coronado, L.M. Weed Classification from Natural Corn Field-Multi-Plant Images Based on Shallow and Deep Learning. *Sensors* **2022**, *22*, 3021. [CrossRef]
11. Thakur, P.S.; Sheorey, T.; Ojha, A. VGG-ICNN: A Lightweight CNN model for crop disease identification. *Multimed. Tools Appl.* **2023**, *82*, 497–520. [CrossRef]
12. Lee, S.H.; Goeau, H.; Bonnet, P.; Joly, A. New perspectives on plant disease characterization based on deep learning. *Comput. Electron. Agric.* **2020**, *170*, 12. [CrossRef]
13. Mu, Y.; Chen, T.S.; Ninomiya, S.; Guo, W. Intact Detection of Highly Occluded Immature Tomatoes on Plants Using Deep Learning Techniques. *Sensors* **2020**, *20*, 2984. [CrossRef]
14. Li, X.C.; Cai, C.L.; Zheng, H.; Zhu, H.F. Recognizing strawberry appearance quality using different combinations of deep feature and classifiers. *J. Food Process Eng.* **2022**, *45*, 11. [CrossRef]
15. Miragaia, R.; Chavez, F.; Diaz, J.; Vivas, A.; Prieto, M.H.; Monino, M.J. Plum Ripeness Analysis in Real Environments Using Deep Learning with Convolutional Neural Networks. *Agronomy* **2021**, *11*, 2353. [CrossRef]
16. 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.
17. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:210708430.
18. Gao, A.; Geng, A.; Zhang, Z.; Zhang, J.; Hu, X.; Li, K. Dynamic detection method for falling ears of maize harvester based on improved YOLO-V4. *Int. J. Agric. Biol. Eng.* **2022**, *15*, 22–32. [CrossRef]
19. Hu, H.; Kaizu, Y.; Zhang, H.; Xu, Y.; Imou, K.; Li, M.; Huang, J.; Dai, S. Recognition and localization of strawberries from 3D binocular cameras for a strawberry picking robot using coupled YOLO/Mask R-CNN. *Int. J. Agric. Biol. Eng.* **2022**, *15*, 175–179. [CrossRef]
20. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef]
21. Cheng, B.; Wei, Y.; Shi, H.; Feris, R.; Xiong, J.; Huang, T. Revisiting rcnn: On awakening the classification power of faster rcnn. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 453–468.
22. Rehman, Z.U.; Khan, M.A.; Ahmed, F.; Damasevicius, R.; Naqvi, S.R.; Nisar, W.; Javed, K. Recognizing apple leaf diseases using a novel parallel real-time processing framework based on MASK RCNN and transfer learning: An application for smart agriculture. *IET Image Process.* **2021**, *15*, 2157–2168. [CrossRef]
23. Zou, K.L.; Chen, X.; Wang, Y.L.; Zhang, C.L.; Zhang, F. A modified U-Net with a specific data argumentation method for semantic segmentation of weed images in the field. *Comput. Electron. Agric.* **2021**, *187*, 9. [CrossRef]
24. Punithavathi, R.; Rani, A.D.C.; Sughashini, K.R.; Kurangi, C.; Nirmala, M.; Ahmed, H.F.T.; Balamurugan, S.P. Computer Vision and Deep Learning-enabled Weed Detection Model for Precision Agriculture. *Comput. Syst. Sci. Eng.* **2023**, *44*, 2759–2774. [CrossRef]
25. Chen, J.Q.; Wang, H.B.; Zhang, H.D.; Luo, T.; Wei, D.P.; Long, T.; Wang, Z.K. Weed detection in sesame fields using a YOLO model with an enhanced attention mechanism and feature fusion. *Comput. Electron. Agric.* **2022**, *202*, 12. [CrossRef]
26. Shao, R.; Shi, Z.; Yi, J.; Chen, P.-Y.; Hsieh, C.-J. On the adversarial robustness of vision transformers. *arXiv* **2021**, arXiv:2103.15670.
27. Tuli, S.; Dasgupta, I.; Grant, E.; Griffiths, T.L. Are convolutional neural networks or transformers more like human vision? *arXiv* **2021**, arXiv:2105.07197.
28. Srinivas, A.; Lin, T.-Y.; Parmar, N.; Shlens, J.; Abbeel, P.; Vaswani, A. Bottleneck transformers for visual recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 16519–16529.
29. You Only Look Once V5. Available online: https://github.com/ultralytics/YOLOv5 (accessed on 11 May 2022).
30. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
31. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.-M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
32. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.

33. Yan, B.; Fan, P.; Lei, X.Y.; Liu, Z.J.; Yang, F.Z. A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. *Remote Sens.* **2021**, *13*, 1619. [CrossRef]
34. Yao, J.; Qi, J.M.; Zhang, J.; Shao, H.M.; Yang, J.; Li, X. A Real-Time Detection Algorithm for Kiwifruit Defects Based on YOLOv5. *Electronics* **2021**, *10*, 1711. [CrossRef]
35. Wang, C.-Y.; Liao, H.-Y.M.; Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W.; Yeh, I.-H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 390–391.
36. 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.
37. Petit, O.; Thome, N.; Rambour, C.; Themyr, L.; Collins, T.; Soler, L. U-net transformer: Self and cross attention for medical image segmentation. In Proceedings of the Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, 27 September 2021; pp. 267–276.
38. Li, Y.; Yao, T.; Pan, Y.; Mei, T. Contextual transformer networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 1489–1500. [CrossRef] [PubMed]
39. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in vision: A survey. *ACM Comput. Surv. (CSUR)* **2022**, *54*, 1–41. [CrossRef]
40. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S. An image is worth 16 $\times$ 16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
41. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
42. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
43. Rezatofighi, H.; Tsoi, N.; Gwak, J.Y.; Sadeghian, A.; Reid, I.; Savarese, S. Computer vision and pattern recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
44. Neubeck, A.; Van Gool, L. Efficient non-maximum suppression. In Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), Hong Kong, China, 20–24 August 2006; pp. 850–855.

agronomy

MDPI

Article

Detection and Counting of Small Target Apples under Complicated Environments by Using Improved YOLOv7-tiny

Li Ma [1], Liya Zhao [1], Zixuan Wang [1], Jian Zhang [2,3,*] and Guifen Chen [4,*]

[1] College of Information Technology, Jilin Agricultural University, Changchun 130118, China; mali@jlau.edu.cn (L.M.); zhaoliya@mails.jlau.edu.cn (L.Z.); 2012100221@mails.jlau.edu.cn (Z.W.)
[2] Faculty of Agronomy, Jilin Agricultural University, Changchun 130118, China
[3] Department of Biology, University of Columbia Okanagan, Kelowna, BC V1V 1V7, Canada
[4] Institute of Technology, Changchun Humanities and Sciences College, Changchun 130118, China
* Correspondence: jian.zhang@ubc.ca (J.Z.); chenguifen@jlau.edu.cn (G.C.)

Abstract: Weather disturbances, difficult backgrounds, the shading of fruit and foliage, and other elements can significantly affect automated yield estimation and picking in small target apple orchards in natural settings. This study uses the MinneApple public dataset, which is processed to construct a dataset of 829 images with complex weather, including 232 images of fog scenarios and 236 images of rain scenarios, and proposes a lightweight detection algorithm based on the upgraded YOLOv7-tiny. In this study, a backbone network was constructed by adding skip connections to shallow features, using P2BiFPN for multi-scale feature fusion and feature reuse at the neck, and incorporating a lightweight ULSAM attention mechanism to reduce the loss of small target features, focusing on the correct target and discard redundant features, thereby improving detection accuracy. The experimental results demonstrate that the model has an mAP of 80.4% and a loss rate of 0.0316. The mAP is 5.5% higher than the original model, and the model size is reduced by 15.81%, reducing the requirement for equipment; In terms of counts, the MAE and RMSE are 2.737 and 4.220, respectively, which are 5.69% and 8.97% lower than the original model. Because of its improved performance and stronger robustness, this experimental model offers fresh perspectives on hardware deployment and orchard yield estimation.

Keywords: YOLOv7-tiny-Apple; small target; fruit detection and counting; digital agriculture

Citation: Ma, L.; Zhao, L.; Wang, Z.; Zhang, J.; Chen, G. Detection and Counting of Small Target Apples under Complicated Environments by Using Improved YOLOv7-tiny. *Agronomy* **2023**, *13*, 1419. https://doi.org/10.3390/agronomy13051419

Academic Editor: Roberto Marani

Received: 30 April 2023
Revised: 17 May 2023
Accepted: 19 May 2023
Published: 20 May 2023

1. Introduction

Apples are a vital agricultural commodity worldwide and a significant contributor to economic development in the agricultural sector. In the United States, Washington State alone accounts for 67% of the entire apple production [1]. Apples are not only delicious but are also packed with a variety of nutrients, including vitamins C and E. These nutritional benefits have made apples a popular choice among consumers. [2,3]. The majority of orchard management today is performed by human labor, which is not only physically taxing but also ineffective, expensive, and prone to error [4].

The integration of intelligence in orchards, driven by the advancement of smart agriculture [5], has emerged as a crucial factor in obtaining precise product information. Nonetheless, detecting fruits accurately in natural environments presents significant challenges. Issues such as fluctuating lighting conditions, overlapping shading, and the resemblance between distant small fruits and the background can lead to inaccuracies in fruit detection [4,6,7]. Accurate fruit detection holds great research value and practical significance for the development of automated harvesting and yield estimation in orchards. Therefore, it is an important area of focus for researchers in this field.

In recent years, there have been some notable advancements in the automated detection and management of orchards, which can be categorized into two main approaches: traditional methods and deep learning algorithms. Traditional image processing methods

have primarily focused on extracting information such as color, shape, and basic features from images [8]. These extracted features are then classified using techniques such as support vector machines or artificial neural networks [9], forming the basis for fruit detection and segmentation [10]. Tian et al. [11] proposed a localization technique based on image depth information to find the circle center, fit the contour, and improve recognition accuracy with a recognition rate of 96.61%. Lin et al. [12] developed a support vector machine-based segmentation algorithm for citrus detection using density clustering to reduce the effect of the environment with excellent performance. Wang et al. [13] used the Retinex image enhancement algorithm, and two-dimensional discrete wavelet transform to apply it to fruit detection based on color features and texture features, with a final accuracy of 85.6%. Wang et al. [14] proposed an algorithm for identifying and locating obscured apples based on K-mean clustering, with a higher localization rate and 89% recognition accuracy than Hough transform and contour curvature methods. Zhang et al. [15] proposed an insulator profile detection method based on an edge detection algorithm. The Canny algorithm was selected as the main algorithm for insulator profile detection, which provides an efficient, accurate, and reliable way for the automated detection of insulator profiles with some practical value.

This paper has selected a deep learning algorithm with strong autonomous learning and feature extraction abilities to address these challenges. These algorithms demonstrate strong robustness and generalization capabilities in fruit detection, making them highly effective for automating orchard processes. Li et al. [16] proposed a target detection algorithm based on YOLOv4-tiny detection of green peppers, combining attention mechanism and multi-scale prediction ideas, with an average accuracy of 95.11%, model size of 30.9 MB, and FPS of 89. Tian et al. [4] presented an enhanced YOLOv3 model for detecting apples at various growth stages. They incorporated DenseNet feature enhancement propagation and enhanced feature reuse techniques to enhance the feature layers at low and medium resolutions. Their model achieved an average detection time that can handle a resolution of 3000 $\times$ 3000 per frame, outperforming the original model and Faster-R-CNN. Zhang et al. [17] proposed a YOLOv4-tiny-based apple detection model with the backbone introducing GhostNet feature extraction network with a CA attention mechanism, introducing depth-separable convolution in the neck and YOLO head for reconstruction, and FPN adding CA attention module to enhance the feature extraction of small targets with an average accuracy of 95.72%. The model has a 37.9 MB model size and a 45.2 FPS. Tu et al. [18] proposed an improved method based on Multi-Scale Fast Region Convolutional Neural Networks (MS-FRCNN) to detect lower-level features by merging feature maps from shallower convolutional feature maps, used in the region of interest pooling, effectively improving the detection of small passionfruit. Uddhav Bhattarai and Manoj Karkee [19] proposed CountNet, a weakly supervised flower/fruit counting network based on deep learning, to learn the number of objects from image-level annotations as input, yielding good MAE and EMSE in an orchard setting. Qian et al. [20] proposed HOG+SVM and an improved YOLOv5-based method for fast recognition of multiple apples in complex environments with parameter reconstruction, the inclusion of an attention mechanism module and fine-tuning of the loss function to better extract the features of different apples and improve the recognition ability of the model. Jan Weyler et al. [21] proposed a method to predict the bounding boxes of crops and weeds automatically, as well as the key points of leaves, with a good ability to estimate leaf numbers, and this method achieved excellent performance in complex scenarios with overlapping plants at different growth stages compared to Mask R-CNN. Chen et al. [22] proposed a Citrus-YOLOv7 model to detect citrus, introducing a small target detection layer, lightweight convolution, and CBAM attention mechanism to achieve multi-scale feature extraction and fusion with an average accuracy of 97.29%, an average detection time of 69.38 ms, and a model size of 24.26 MB. Sun et al. [23] proposed an optimized Retinanet-PVTv2 that introduced a gradient coordination mechanism to detect small green fruits/begonia in a nocturnal environment and showed APs of 85.2% and 61.0% on the NightFruit and Gala datasets, respectively. Yonis Gulzar [24] proposed a

TL-MobileNetV2 model used to classify 40 fruits, removing the classification layers present in the MobileNetV2 architecture, adding five different layers, and also using different pre-processing and model adjustments, effectively improving the efficiency and accuracy of the model with an accuracy of 99%. Normaisharah Mamat et al. [25] proposed an automatic image annotation advancement method using repetitive annotation tasks for the automatic annotation of objects, which was evaluated on different YOLO networks, and the results showed that the proposed method is fast and highly accurate in fruit classification, providing great value in image annotation. Yasir Hamid [26] proposed a deep convolutional neural network model using MobileNetV2 architecture and data augmentation techniques to implement an intelligent seed classification system, and the results show that it has high accuracy and is important for the sustainability of seed analysis. Sonam Aggarwal [27] proposed an artificial intelligence-based stacked integration approach to predict protein subcellular localization, which combines the three powerful pre-trained models, and the results showed that combining different weak convolutional neural networks gave better predictions than individual models with better network performance.

This study proposes an improved YOLOv7-tiny detection and counting network for small target apples to better cope with the inevitable realities of natural environments, complex backgrounds, and complex weather. The dataset was enriched with data enhancement techniques, adding three innovations to the backbone and neck of the model, improving detection accuracy for small target apples, and reducing the number of parameters in the model. The results show that the model has high robustness and generalization performance in detection and counting and can be deployed to mobile applications to facilitate automated orchard management.

The following are the contributions of the authors' work in this study:

(1) Construction of an Apple dataset with complex weather using the MinneApple public dataset, cropping and marking this, and simulating rain and fog scenarios using data augmentation techniques;

(2) Borrowing from DenseNet idea to construct a backbone network by adding to shallow features through skip connections to reduce the loss of small target features;

(3) Multi-scale fusion using P2BiFPN to detect small fruits with low resolution at a distance;

(4) A lightweight ULSAM attention mechanism is used to discard redundant features without increasing the model volume.

This study aims to enhance the model's ability to detect and count small target apples with complex backgrounds in natural environments and to adapt to complex weather conditions.

2. Materials and Methods

2.1. Construction of the Apple Dataset

2.1.1. Image Acquisition and Pre-Processing

The MinneApple dataset was used for this experiment. It was gathered at the University of Minnesota Horticultural Research Center (HRC) between June 2015 and September 2016 [28]. Video clips from various parts of the orchard were collected using a standard Samsung Galaxy S4 phone, and images were then captured to create the desired images "https://conservancy.umn.edu/handle/11299/206575 (accessed on 10 September 2022)". The dataset used in this study was collected from both sunny and shaded areas of tree rows, spanning multiple days to capture diverse lighting conditions. The dataset was challenging, as it included heavily shadowed and crowded scenes with apples that were difficult to locate. Additionally, the dataset comprised various regions with different apple types. The dataset selected in this paper is the training set of Image images in the detection in MinneApple. In this study, we cropped and labeled the images and classified them into three categories: GreenApple, Red1Apple, and Red2Apple. The GreenApple tree has a thick trunk with smooth, green-colored outer skin. The Red1Apple tree is loosely structured, with a yellow-green base color on the fruit and a slightly powdery texture. The Red2Apple

tree is conical in shape, with smooth and evenly colored outer skin, appearing bright red. GreenApple is mostly under bright lighting conditions, and they have an extremely similar background. Red1Apple is in weak lighting conditions with insufficient brightness, while Red2Apple is in backlit and dim lighting conditions. For these three categories of fruit, we pre-processed the data.

The dataset is mainly for the detection of small target apples, as shown in Figure 1. We selected some images to crop them into 841 images of different sizes, reducing the input image size from 1280 × 720 to images of different resolution sizes, including 302 GreenApple of 320 × 320 resolution size, 252 Red1Apple of 320 × 320 resolution size and 410 × 410 resolution size, and 287 Red2Apple of 320×320 resolution size. The cropped images were manually annotated using the online web-based tagging tool makesense.ai "https://www.makesense.ai/ (accessed on 20 September 2022)" and finally exported in the generated .voc format, which was converted to YOLO format to complete the construction of the initial dataset.

Figure 1. Cropping of the data set.

2.1.2. Data Augmentation

A dataset with complex weather was constructed using image enhancement methods to simulate different levels of rain and fog scenarios. This is shown in Figure 2. This is used to test the generalization and robustness of our model. From the overall dataset of three types of apples, 80 images were randomly selected from each fruit class for single-fold data augmentation. The iaa.Fog() function was used to simulate foggy weather conditions, and then the iaa.Rain(speed = (0.1, 0.3)) function was used to simulate rainy weather conditions. This completed the enhancement processing under different weather conditions. Data cleaning was performed in this process, and the final enhanced dataset obtained contained 232 foggy weather images and 236 rainy weather images. The original images and the enhanced images form a dataset with complex weather, containing a total of 829 images, as shown in Table 1. The constructed dataset is randomly partitioned into training, testing, and validation sets according to the ratio of 7:2:1. Finally, 586 images of the training set, 164 images of the test set, and 66 images of the validation set were obtained.

Table 1. The volume of data before and after enhancement and for different weather.

Category	Raw Data Volume	Weather Data Enhancement			Experimental Data Volume
		Normal Scenario	Fog Scenario	Rain Scenario	
GreenApple	302	142	79	76	297
Red1Apple	252	92	73	80	245
Red2Apple	287	127	80	80	287

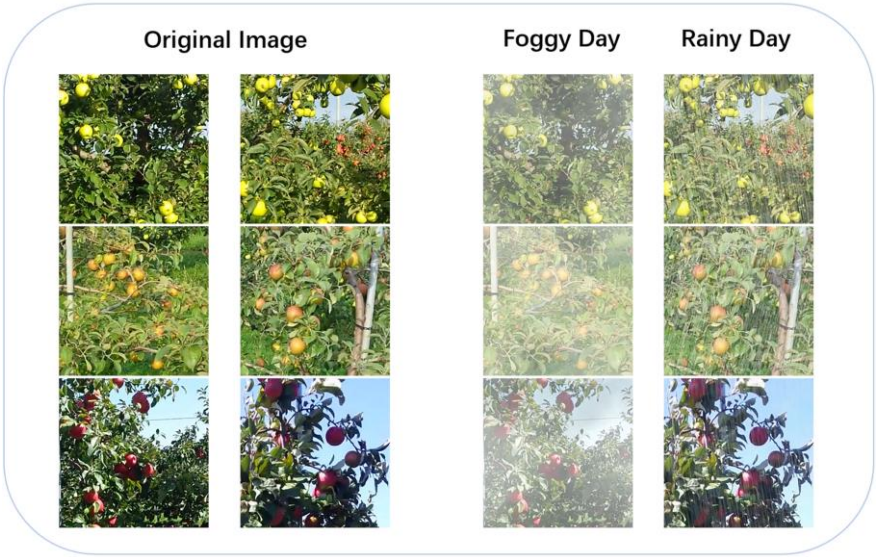


Figure 2. Data enhancement in different weather.

2.2. Selection of Models

CNN-based networks are crucial for fruit detection in terms of target detection. Some of the more established and well-liked detection networks include the Faster Rcnn [29], SSD [30], CenterNet [31], and YOLO series. Object detection is transformed into a regression problem by the single-stage object detection network known as YOLO. By processing the image with a single CNN, YOLO can directly obtain the class and location coordinates of the object. This end-to-end detection network greatly increases the speed of detection. The YOLOv7 network model [32] of the YOLO family has good detection performance. One of them, YOLOv7-tiny, is a lightweight model of the YOLOv7 series. Ease of deployment and detection with accuracy and real-time performance is key to small-target apple detection. In this regard, we compared the lightweight YOLOv7-tiny model with Faster Rcnn-VGG, SSD-MobileNetV2, and CenterNet-ResNet50 for the study. As shown in Table 2, it can be seen that the YOLOv7-tiny detection effect is more pronounced, while the detection speed is fast and model volume is small, and can provide some subsequent mobile deployment technical support.

Table 2. Results of the inter-model selection comparison.

Model	F1 Score/%	Precision/%	Recall/%	mAP0.5/%	Params/MB	FPS
YOLOv7-tiny	72.3	77.2	68.1	74.9	6.01	126.56
Faster Rcnn-VGG	50.33	41.16	66.51	55.95	547.0	24.90
SSD-MobileNetV2	47.3	95.07	31.93	64.16	16.6	81.98
CenterNet-ResNet50	48.3	95.48	40.56	71.22	131.0	69.74

2.3. YOLOv7-tiny-Apple Construction

The YOLOv7-tiny model consists of the backbone, neck, and head. The backbone network is mainly composed of CBL, ELAN-T, and SPPCSPC to perform feature extraction. To enhance the perceptual field of the model and strengthen the features, the neck network PANet merges the characteristics learned at various scales. The head is convolved by 1×1 to transform the features into the final prediction information to obtain the final prediction results. This model has excellent speed and outstanding network performance. For the detection in the complex background of the natural environment, to solve the problem of distant small fruits, missing contours, and weather interference caused by the decline in detection accuracy, this paper has improved YOLOv7-tiny, as shown in Figure 3. The shallow feature fusion helps to reduce the loss of small targets by borrowing the DenseNet

idea in the backbone part and adding a skip connection(a) in the P3 part. The neck is fused at multiple scales by upgrading PANet to P2BiFPN(b), adding a small target detection layer while removing the large target detection layer, and mixing shallow features with deep features for feature fusion reuse and improved detection accuracy. A lightweight ULSAM attention mechanism(c) is added here to focus on the right target discarding redundant features.

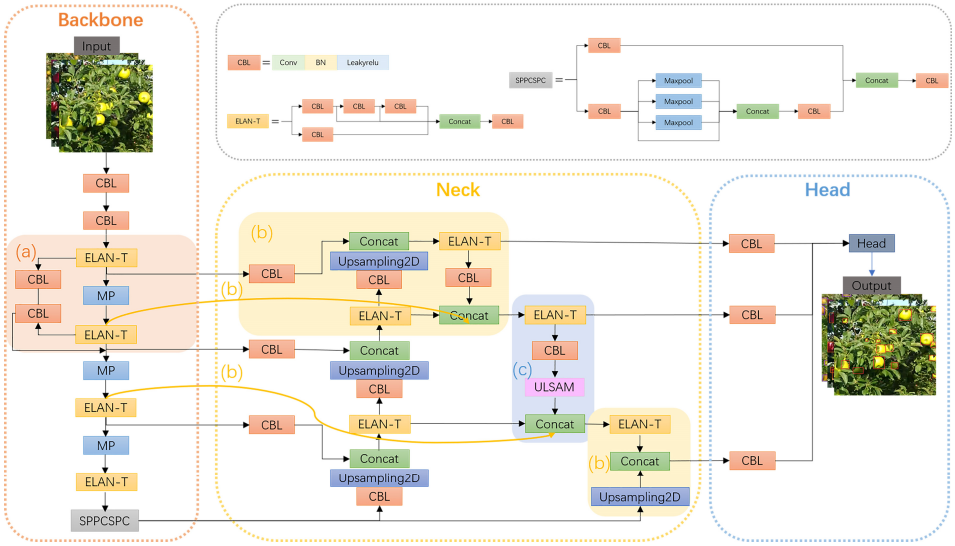


Figure 3. Apple detection and counting model based on YOLOv7-tiny-Apple, (a) skip connection, (b) P2BiFPN, and (c) ULSAM attention mechanism.

2.3.1. Shallow Feature Fusion

Drawing on the DenseNet [33] idea, we make better use of feature information and improve the efficiency of information propagation between layers. DenseNet connects all layers for channel merging and feature reuse, as shown in Figure 4. x_0, x_1,x_2,x_3,x_4 indicate the feature map of each output layer, and h1,h2,h3,h4 indicate the nonlinear transformation; each layer can accept all the previous feature map (l-1) layers [34], and the feature map of each layer is expressed in Equation (1).

$$x_l = H_l[x_0, x_1, \ldots, x_{l-1}] \tag{1}$$

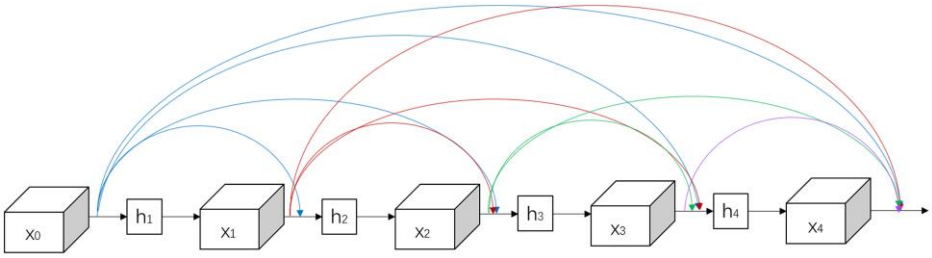


Figure 4. DenseNet feature extraction process.

This experiment proposes to add a skip connection to the p3 position of the backbone network, which can more effectively utilize the feature information to achieve feature reuse and reduce the loss of small target feature information in the transmission process. For this reason, this experiment is set up with two layers; too many connections can cause slow detection speed problems. The input is a 416 × 416 size image that passes through a downsampling convolution layer with a channel number of 32 × 32 and a 3 × 3 convolution kernel and passes to a downsampling convolution layer with a 64 × 64 feature map size and a 3 × 3 size convolution kernel. The convolutional layer of the backbone network

consists of Conv (convolution), BN (batch normalization), and the LeakyRelu activation function. P2 is followed by a convolutional downsampling, and a convolution is connected after p3 to connect it to p2, for which the skip connection is integrated to propagate the connected features forward again, greatly preserving the small target apple features in the process.

2.3.2. Fusion of Extremely Small Target Features in P2BiFPN

The conventional FPN [35], as in Figure 5a, possesses only a single top-down information flow, which causes the loss of feature information. PANet [36] is used in the YOLOv7-tiny network, as shown in Figure 5b, and is a bottom-up information enhancement based on the FPN, and the ability to acquire features still needs to be improved. In this study, the small target detection layer P2 is added to YOLOv7-tiny for stronger feature fusion, preserving the semantic information of small, shallow targets. It is achieved by removing the P5 detection head, fusing the original P5 feature layer with the P4, and finally, outputting the P4 detection head. Drawing on BiFPN [37], which is improved based on PANet, using the idea of bi-directional fusion is a simple and efficient feature fusion mechanism where each input feature has a different resolution, for which efficient feature fusion is performed. As shown in Figure 5, it can be seen that the concat connection of P3P4 in the backbone and P3P4 in the neck, respectively, top-down and bottom-up operations will achieve higher feature fusion. The detection accuracy was improved while the number of parameters was reduced, and this was fused to the whole neck, which we named P2BiFPN, as in Figure 5c. For the detection of small targets, especially small fruits with low resolution at a distance, they possess more reliable accuracy under weather interference.

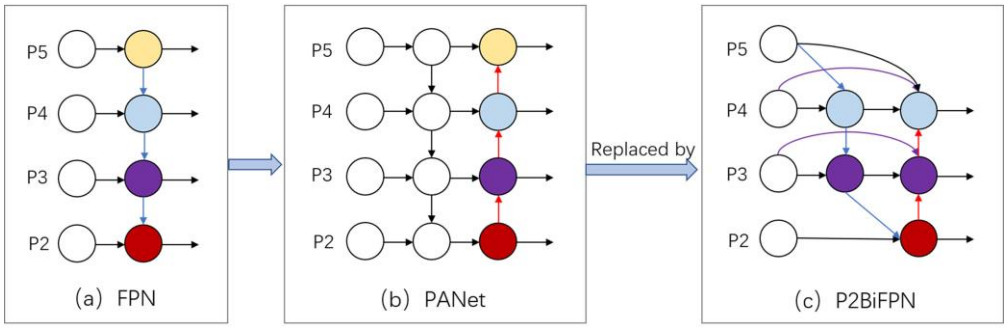


Figure 5. The improvement process of the neck network, with different colored circles representing elements of different sizes, (**a**) FPN, (**b**) PANet, and (**c**) P2BiFPN.

2.3.3. ULSAM Discards Superfluous Features

The detection and counting of small target apples are inaccurate in the case of complex backgrounds and weather changes, especially since GreenApple has great similarity to the background. To solve these problems and discard useless features, we propose adding a ULSAM (Ultra-Lightweight Subspace Attention Mechanism) to the neck [38], which can learn each attention mapping in each feature subspace. As shown in Figure 6, m is the number of input channels, G indicates that each group has G feature maps, h, w is the spatial dimension of the feature maps, forming a set of intermediate feature maps, using the linear relationship between different feature subspaces to integrate the channel information and have a more efficient effect on the learning of features, and do not burden the network. In this experiment, the ULSAM module is added to PANet with the parameters set to a 128 $\times$ 128 feature map size, a 26 $\times$ 26 picture size, and a default g = 4 with connection downsampling. It is well embedded into the network, increasing the detection accuracy while the model size does not change, satisfying the detection model of the lightweight network.

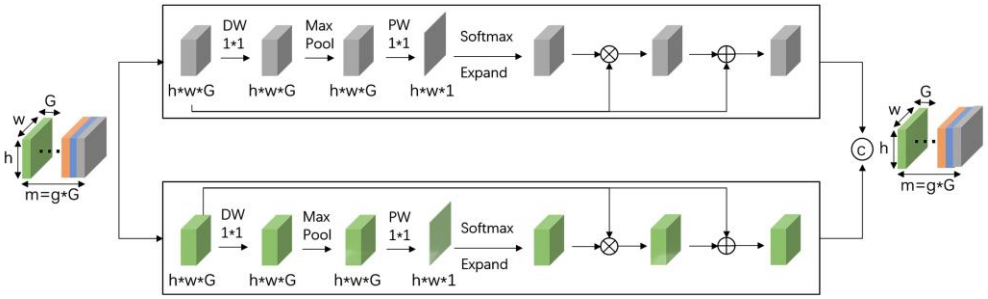


Figure 6. ULSAM structure diagram.

2.4. Experiment Environments

In this experiment, a desktop computer is used, and the running environment is the Pytorch deep learning framework built with the Ubuntu 18.04.6LTS system, configured with 11th Gen Intel(R) Core (TM) i7-11700k@3.60GHz*16, GPU is NVIDIA GeForce GTX 1080Ti/PCIe/SSE2, using CUDA10.2, OpenCV, Cudnn, and other related libraries to implement the detection model of small target applies in the context of complex environments.

2.5. Evaluation Indicators

In this experiment, the average metrics of three types of apples were selected and validated on a randomly divided test set. The mean F1 Score (mF1), Precision, Recall, and mean Average Precision(mAP@0.5) evaluation metrics were used to test the performance of the network, and N represents the number of categories. The evaluation metric is used to measure the performance of the whole model [39], calculated as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \tag{2}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \tag{3}$$

$$\text{F1} = \left(\frac{1}{\text{N}}\sum 2\frac{\text{R} \times \text{P}}{\text{R} + \text{P}}\right)^2 \tag{4}$$

$$\text{AP} = \int_0^1 \text{P}(\text{R})\text{dR} \times 100\% \tag{5}$$

$$\text{mAP} = \frac{\sum_1^{\text{N}} \int_0^1 \text{P}(\text{R})\text{dR}}{\text{N}} \times 100\% \tag{6}$$

where TP represents the number of true positive correctly identified samples, FP represents the number of false positive misidentified negative samples, and FN represents the number of false negative missed positive samples.

In evaluating apple counting, RMSE (Root Mean Square Error) [40] and MAE (Mean Absolute Error) [40] are utilized to measure the effectiveness of counting. K is the number of images, pi is the true number of apple labels, and qi is the number of detected apples, calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{\text{K}}\sum_{\text{i}=1}^{\text{K}}(\text{p}_\text{i} - \text{q}_\text{i})^2} \tag{7}$$

$$\text{MAE} = \frac{1}{\text{K}}\sum_{\text{i}=1}^{\text{K}}|\text{p}_\text{i} - \text{q}_\text{i}| \tag{8}$$

3. Results

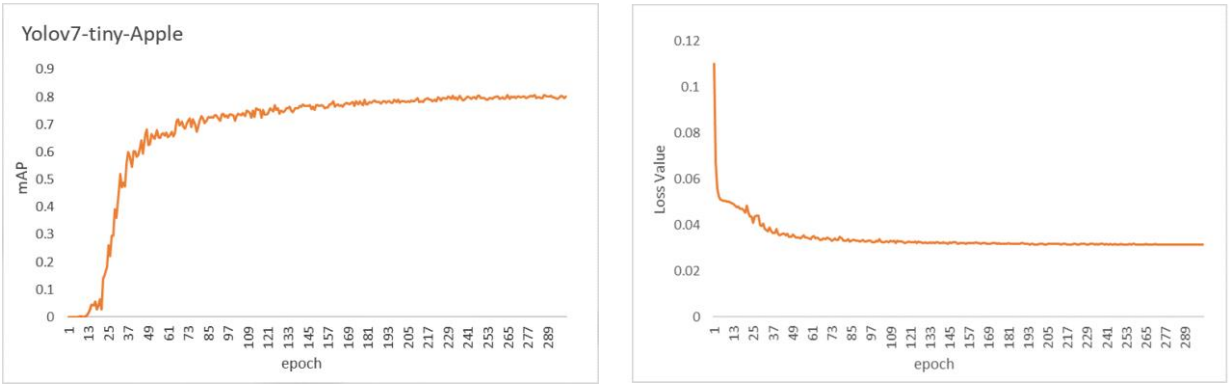
3.1. Training Process

In this study, the same initial training parameters are set for each group of experiments. We input an image size of 416 × 416, an epoch value of 300, a learning rate of 0.01, a momentum of 0.937, and a weight_decay of 0.0005, and the optimizer chosen is Adam, as shown in Table 3. The data is recorded using Tensorboard during the training process, and the training set loss is written for each iteration; the validation set loss is written for each training round, and the model weights are saved.

Table 3. All hyperparameters and values.

Parameter	Value
Input size pixels	416 × 416
Training epochs	300
Learning rate	0.01
Momentum	0.937
Weight decay	0.0005
Optimizer	Adam

As shown in Figure 7, the model's detection accuracy values and training loss values vary with the number of iterations during the training process. Gradually converging from 50 rounds onwards, the detection accuracy gradually increases, and the loss of the model gradually decreases, and finally, the 250 rounds level off, and the accuracy and loss values no longer change. It can be seen that the YOLOv7-tiny-Apple model does not suffer from over- and under-fitting and gradient disappearance [41] and can be applied to the detection and counting of small target fruits.



(**a**) (**b**)

Figure 7. Loss and mAP with iteration. (**a**) mAP curve chart; (**b**) Loss Value curve chart.

3.2. Ablation Experiments

In this experiment, ablation experiments verify each improvement point's effectiveness. We propose a total of three innovation points, and the results of the ablation experiment are shown in Table 4.

After improving the first skip connection j, it is clear that the network adding the skip connection improves mAP by 3.6%, mF1, Precision, and Recall by 3.1%, 2.6%, and 3.5%, respectively, on the original base network, and the number of parameters of the model increases by 0.16 M. The second improvement point, P2BiFPN, improves mAP by 3.8%, mF1, Precision, and Recall by 3.5%, 1.8%, and 4.8%, respectively, compared with the base network, and the amount of parameters of the model is reduced by 1.11 MB, which greatly reduces the complexity of the model. The third improvement point is that the ULSAM attention mechanism increases the detection accuracy while the number of parameters

of the model does not change, satisfying the detection model of the lightweight network. Compared with the base network, mAP improved by 2.9%, mF1, Precision, and Recall improved by 2.3%, 1.4%, and 3.3%, respectively; the number of parameters of the model did not change, while FPS was reduced but could meet the fast detection of the model.

Table 4. Comparison between the improvement points of the ablation experiment.

j	ULSAM	P2BiFPN	F1 Score/%	Precision/%	Recall/%	mAP0.5/%	Params/MB	FPS
			72.3	77.2	68.1	74.9	6.01	158.73
√			75.4	79.8	71.6	78.5	6.17	156.25
	√		74.6	78.6	71.4	77.8	6.01	133.33
		√	75.8	79.0	72.9	78.7	4.90	123.45
√	√		76.0	82.7	70.5	79.1	6.17	125
	√	√	76.6	79.5	74.0	79.8	4.90	104.16
√		√	75.8	81.2	71.4	79.7	5.06	111.11
√	√	√	76.8	80.1	74.1	80.4	5.06	101.01

All three improvement points significantly enhanced the accuracy of the base model in this experiment. We systematically combined them in pairs, as shown in Figure 8, demonstrating the feasibility of the model's detection accuracy after each combination. Finally, by integrating all three improvements, we achieved an mAP (mean Average Precision) value of 80.4%, Precision, Recall, and mF1 values of 80.1%, 74.1%, and 76.8%, respectively. Moreover, the resulting model has a compact size of 5.06 MB. This shows that the improvement of this experiment is more effective in feature extraction in the complex background of the natural environment, largely discarding the redundant interference and having superior detection performance.

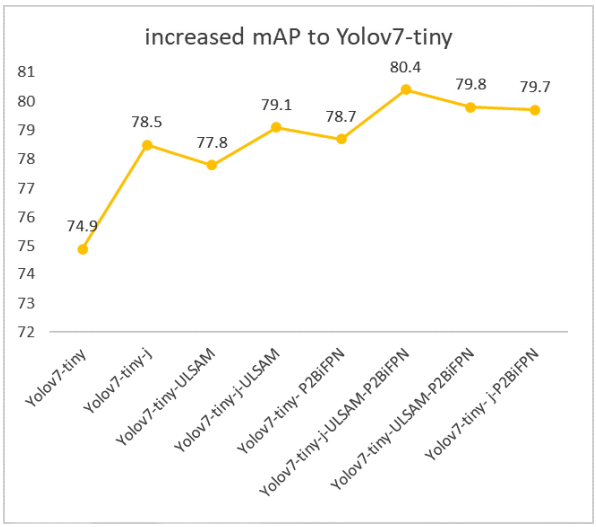


Figure 8. mAP line graph of ablation experiment.

For counting, this study uses the images in the test set for testing, and the counting results are shown in Table 5. From the results, it can be seen that for the first innovation point j skip connection, RMSE is reduced by 0.281, and MAE is slightly higher by 0.006; for the second innovation point, P2BiFPN, RMSE is reduced by 0.285, and MAE is slightly higher by 0.091; for the third innovation point, the ULSAM attention mechanism, RMSE is reduced by 0.32, and MAE is reduced by 0.043. The counting metrics of this experiment, RMSE reduced by 0.416 and MAE reduced by 0.165, are also effective in terms of counting and have performance improvement.

Table 5. Count results of ablation experiments.

Model	MAE	RMSE
YOLOv7-tiny	2.902	4.636
YOLOv7-tiny-j	2.908	4.355
YOLOv7-tiny-ULSAM	2.859	4.316
YOLOv7-tiny-j-ULSAM	2.902	4.534
YOLOv7-tiny-P2BiFPN	2.993	4.351
YOLOv7-tiny-j-ULSAM- P2BiFPN	2.737	4.220
YOLOv7-tiny-ULSAM-P2BiFPN	2.889	4.312
YOLOv7-tiny- j-P2BiFPN	2.902	4.507

3.3. Comparison Experiments

The more classical and popular Faster Rcnn-VGG, SSD-MobileNetV2, CenterNet-ResNet50, the newer and prominent lightweight networks YOLOv5s [42], YOLOx-tiny [43], YOLOv6t [44], and YOLOv7 were selected for comparison with the seven models of YOLOv7-tiny-Apple in this study, and the detection results are shown in Table 6.

Table 6. Accuracy comparison between different models.

Model	F1 Score/%	Precision/%	Recall/%	mAP0.5/%	Params/MB	FPS
YOLOv7-tiny	72.3	77.2	68.1	74.9	6.01	126.56
Faster Rcnn-VGG	50.33	41.16	66.51	55.95	547.0	24.90
SSD-MobileNetV2	47.3	95.07	31.93	64.16	16.6	81.98
CenterNet-ResNet50	48.3	95.48	40.56	71.22	131.0	69.74
YOLOv7	74.2	80.5	69.0	77.8	36.49	55.55
YOLOv5s	75.42	81.3	70.5	78.4	7.02	178.57
YOLOx-tiny	70.6	77.2	65.2	72.14	5.06	84.83
YOLOv6t	75.1	79.7	71.0	76.4	9.67	105.15
YOLOv7-tiny-Apple	76.8	80.1	74.1	80.4	5.06	101.01

By comparing the experimental tests, it was found that the mAP of YOLOv7-tiny-Apple was the highest among the models, reaching 80.4%, which was 2% higher than that of YOLOv5, the best-performing model, ensuring the detection accuracy, while the number of parameters was also the smallest. Compared with Recall, the model sensitivity of YOLOv7-tiny-Apple is 6%, 5.1%, 3.6%, 8.9%, and 3.1% higher than the other models, respectively. When comparing Precision, SSD-MobileNetV2, and CenterNet-ResNet50 stand out more but have no advantage in the remaining areas. The value of mF1 was compared with several models and found that the value of this model was the highest and had better detection accuracy for detecting different classes of fruits. Relative to the FPS, the FPS of YOLOv7-tiny-Apple was 101.01Hz, which was higher than Faster Rcnn-VGG, SSD-MobileNetV2, CenterNet-ResNet50, YOLOv7, and YOLOx-tiny and lower than the other three models, but it is sufficient for fast detection of fruits. This shows the superiority of the present model compared to other models, and compared with other current YOLO small model series, we can see that the present model has higher detection accuracy while the model is also the smallest, which is sufficient to achieve lightweight, small target detection in the complex background of a natural environment, illustrating the superiority of the present experimental model.

The test set photos were also used to detect the counting images from the comparison trials. To examine the efficacy of this experimental approach, Figure 9 displays the MAE and RMSE counting impacts of the evaluation model. It is evident that the YOLOv7-tiny-Apple model's MAE and RMSE, which are 2.737 and 4.220, respectively, better than the counting effects of other models and are the best counting effects. As shown in Table 7.

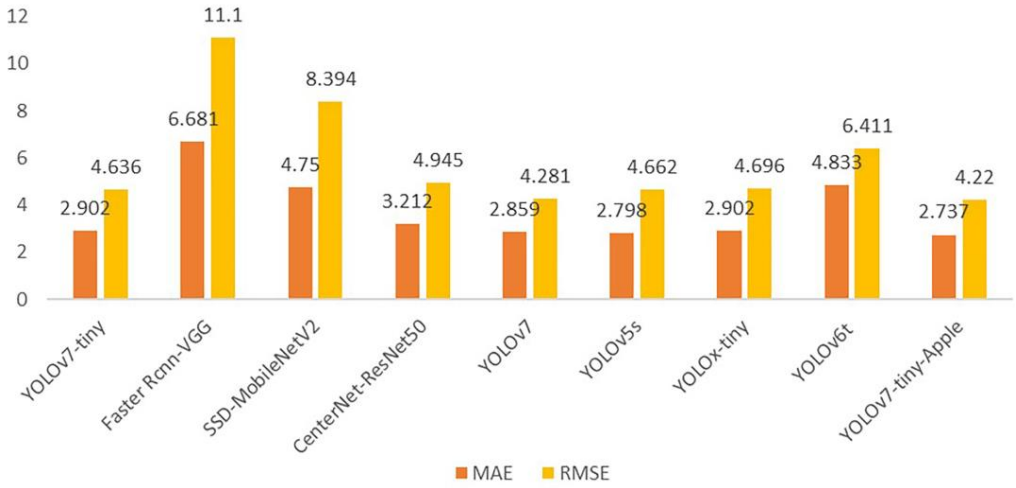


Figure 9. Histogram of counting effect between different models.

Table 7. Comparison of experimental count results.

Model	MAE	RMSE
YOLOv7-tiny	2.902	4.636
Faster Rcnn-VGG	6.681	11.10
SSD-MobileNetV2	4.750	8.394
CenterNet-ResNet50	3.212	4.945
YOLOv7	2.859	4.281
YOLOv5s	2.798	4.662
YOLOx-tiny	2.902	4.696
YOLOv6t	4.833	6.411
YOLOv7-tiny-Apple	2.737	4.220

4. Discussion

Based on the aforementioned discussion, it is evident that the detection model employed in this experiment exhibits high accuracy. To assess the model's robustness in detecting apples within complex environments, we conducted a comparative visual analysis. The analysis distinguished between correctly detected boxes and missed detection cases, represented by oval white boxes. Common challenges observed in both cases include heavy leaf shading and loss of fruit outline. By examining these factors, we were able to evaluate the effectiveness of fruit detection in three distinct scenarios.

4.1. Fruits in Dim or Backlit Conditions

As shown in Figure 10, the comparison of fruit detection results in dim or backlight conditions shows that YOLOv7-tiny-Apple can meet the extremely difficult detection of small fruits and fruit overlap in dim conditions. For the four fruits in the lower right corner in the extreme dimness, their fruits were missed by (h) Faster Rcnn-VGG, (i) SSD-MobileNetV2 and two fruits were missed by (b) YOLOv7-tiny (c), YOLOv7 (d), YOLOv5s, (g) CenterNet-ResNet50, and one fruit was missed by (e) YOLOv6t, (f) YOLOx-tiny. There are two overlapping heavily occluded fruits in the upper left corner, (b) YOLOv7-tiny, (c) YOLOv7, (g) CenterNet-ResNet50, (h) Faster Rcnn-VGG did not detect them, (d) YOLOv5s, (e) YOLOv6t, (f) YOLOx-tiny, and (i) SSD-MobileNetV2 detected one fruit and the other heavily occluded fruit was not detected. At the top, there are two dim fruits, and (h) Faster Rcnn-VGG misses them all, while (b) YOLOv7-tiny, (c) YOLOv7, and (g) CenterNet-ResNet50 all omit one. This model, (a) YOLOv7-tiny-Apple, can detect all fruits, which has strong feature extraction and high localization accuracy.

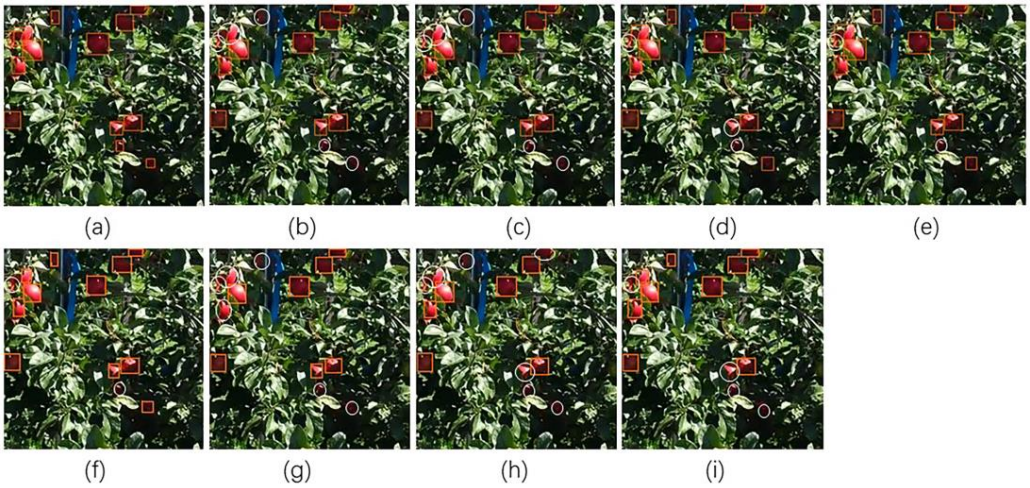


Figure 10. Detection effect of different models in dim or backlit scenes. (**a**) YOLOv7-tiny-Apple; (**b**) YOLOv7-tiny; (**c**) YOLOv7; (**d**) YOLOv5s; (**e**)YOLOv6t; (**f**) YOLOx-tiny; (**g**) CenterNet-ResNet50; (**h**) Faster Rcnn-VGG; (**i**) SSD-MobileNetV2.

4.2. Fruits under Low Light and Blurred Outline of Small Fruits in the Distance

As shown in Figure 11, YOLOv7-tiny-Apple targets fruits that are more difficult to capture with low natural light and insufficient brightness for detection. It is found that the blurring of the distant small fruit outline deepens the detection difficulty even more in the case of weak light. One to two fruit misses occur at the top of the picture, as shown in (b) YOLOv7-tiny, (c) YOLOv7, (d) YOLOv5s, (e) YOLOv6t, (f) YOLOx-tiny, (g) CenterNet-ResNet50, (h) Faster Rcnn-VGG, and (i) SSD-MobileNetV2. The loss and low resolution of the rightmost fruit outline were not detected by any of the other models, except for (a) YOLOv7-tiny-Apple, which detected it accurately. In particular, (h) Faster Rcnn-VGG, (i) SSD-MobileNetV2 miss significantly. It is clear that YOLOv7-tiny-Apple possesses higher detection performance and significantly overcomes the issue of inaccurate detection of far-off small fruits.

Figure 11. Detection effects of different models under low natural light and low brightness. (**a**) YOLOv7-tiny-Apple; (**b**) YOLOv7-tiny; (**c**) YOLOv7; (**d**) YOLOv5s; (**e**) YOLOv6t; (**f**) YOLOx-tiny; (**g**) CenterNet-ResNet50; (**h**) Faster Rcnn-VGG; (**i**) SSD-MobileNetV2.

4.3. Fruits in the Case of Bright or Overlapping Shadows and High Background Similarity

As shown in Figure 12, under the bright and overlapping shadows, there is an extreme similarity between GreenApple and the background. The fruit in the lower left corner is shaded with higher similarity to the leaves as shown in Figure (b) YOLOv7-tiny,

(e) YOLOv6t, (f) YOLOx-tiny, (g) CenterNet-ResNet50, and (h) Faster Rcnn-VGG is difficult to detect. At the same time, the contours are missing more severely, and the whole leaf covering the fruit is difficult to be detected accurately. As shown in the upper right corner of (b) YOLOv7-tiny, (c) YOLOv7, (d) YOLOv5s, (e) YOLOv6t, (f) YOLOx-tiny, (h) Faster Rcnn-VGG, (i) SSD-MobileNetV2 can be seen; all missed this fruit. For the rightmost fruit, there is also its omission. However, (a) YOLOv7-tiny-Apple can accurately detect and locate it and better extract the features of the green fruit, which greatly solves the background similarity. The problem of background similarity is considered solved, which reflects the superiority of this model.

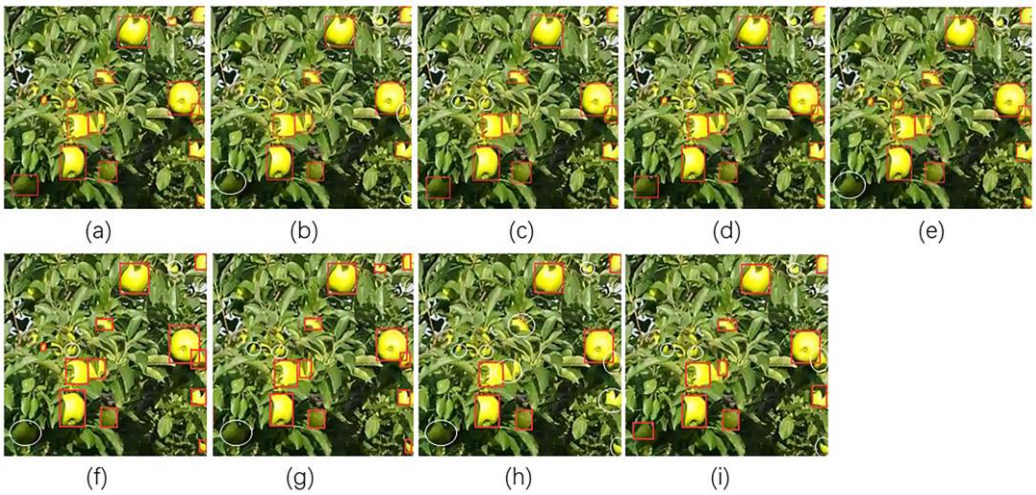


Figure 12. Detection effects of different models in bright and shadow overlapping scenes with extreme similarity to the background. (**a**) YOLOv7-tiny-Apple; (**b**) YOLOv7-tiny; (**c**) YOLOv7; (**d**) YOLOv5s; (**e**) YOLOv6t; (**f**) YOLOx-tiny; (**g**) CenterNet-ResNet50; (**h**) Faster Rcnn-VGG; (**i**) SSD-MobileNetV2.

4.4. Extreme Detection in Variable Weather

It can be seen from the above that the experiment possesses good accuracy and adaptability for the detection of apples in complex environments. To further verify YOLOv7-tiny-Apple's robustness to complex weather, we selected representative distant green fruits in rainy and foggy weather scenarios for testing. The small green apples in the distance are more difficult to detect in complex weather due to the large background similarity, high degree of leaf occlusion, and low pixels, so we verify this. As can be seen from Figure 13, YOLOv7-tiny-Apple can effectively detect most of the distant small green apples on rainy and foggy days, and only some small green apples and the part of the target feature that is not oriented are not detected. In summary, under the interference of complex weather, our detection model is still effective, with strong positioning accuracy and superior robustness.

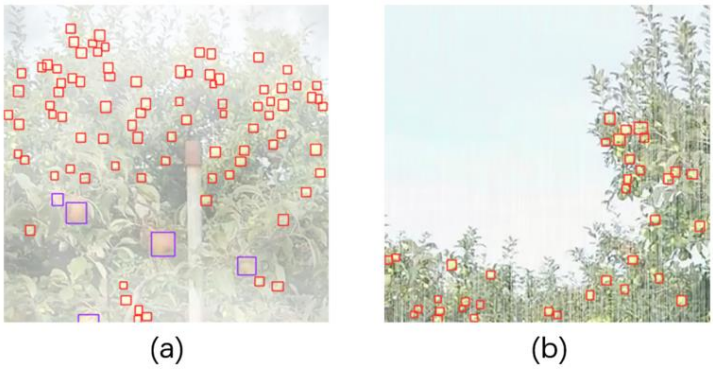


Figure 13. YOLOv7-tiny-Apple detection for rain and fog scenarios. (**a**) Foggy weather; (**b**) rainy weather.

5. Conclusions

A lightweight small-target apple detection and counting model YOLOv7-tiny-Apple, is proposed based on YOLOv7-tiny, which provides theoretical support for designing apple detection and counting models. This model can cope with the detection of small target apples under the complex weather of the natural environment, further promoting the picking and estimation of production in automated orchards, and can also provide some technical support in the deployment of equipment. Excellent potential for real-time application in orchard management. This study has a certain encouraging influence on increasing labor cost effectiveness, increasing production quality, and enhancing agricultural production efficiency. It aids in real-time apple identification and more effective orchard management by automatically identifying and counting apples. Additionally, by efficiently decreasing the use of computational resources and time and increasing the efficiency of practical application, this work plays a constructive role in promoting and using object detection technology, which has some practical implications.

(1) This model uses a skip connection integrated into the backbone network of YOLOv7-tiny to fuse the shallow features, strengthen the small target features, solve the problem of missing part of small target features, and greatly preserve the apple features;
(2) Due to the complex environment, the distant small fruit contours in the image are seriously missing, which affects the detection results. For this reason, we propose P2BiFPN added to the neck, which carries out multi-scale fusion and feature reuse, which not only improves the accuracy but also reduces the volume of the model to a certain extent;
(3) To further optimize the model, we add the ULSAM completely lightweight attention mechanism, which resolves the issue of identifying extreme similarity between the target and the background while ensuring the accuracy of the model;
(4) The majority of current research relies on the identification of complex background fruits in normal weather, which is inflexible to the effects of changing weather. In this experiment, the used dataset was processed using image enhancement techniques, and the rain and fog scenarios were simulated to verify the detection ability of the model in different weather with good detection performance.

Compared with other models, the improved model has better generalization and robustness and can be adapted to the task of small target apple detection in natural environments with complex backgrounds of variable weather. Possible techniques for using the model could also include aspects such as efficient inference techniques and algorithm optimization, which are used to improve the detection accuracy and efficiency of the model for different applications. Future work will concentrate on lightweight optimization, higher detection accuracy, and mobile device deployment to further enhance technical monitoring and control of smart orchards.

Author Contributions: The contributors are G.C. and L.Z. for conceptualization; Z.W. for methodology; J.Z. for formal analysis; Z.W. for data curation; L.M. and L.Z. for investigation/writing—original draft/supervision; L.M. and G.C. for visualization; J.Z. for writing—review/editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China–Joint Fund (u19a2061), Strategic Research and Consulting Project of Chinese Academy of Engineering (No. JL2023-03), Jilin Provincial Department of Education Project (No. JJKH20210336KJ), Jilin province science and technology development plan project (No. 20210204050YY), Jilin University Student Innovation Training Program (S202210193103).

Data Availability Statement: Data supporting the findings of this study are available from the corresponding author.

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

References

1. Lu, S.; Chen, W.; Zhang, X.; Karkee, M. Canopy-Attention-YOLOv4-Based Immature/Mature Apple Fruit Detection on Dense-Foliage Tree Architectures for Early Crop Load Estimation. *Comput. Electron. Agric.* **2022**, *193*, 106696. [CrossRef]
2. Koutsos, A.; Tuohy, K.M.; Lovegrove, J.A. Apples and Cardiovascular Health—Is the Gut Microbiota a Core Consideration? *Nutrients* **2015**, *7*, 3959–3998. [CrossRef] [PubMed]
3. Blažek, J.; Paprštein, F.; Zelený, L.; Křelinová, J. The Results of Consumer Preference Testing of Popular Apple Cultivars at the End of the Storage Season. *Hortic. Sci.* **2019**, *46*, 115–122. [CrossRef]
4. 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]
5. Zhao, C. Current situations and prospects of smart agriculture. *J. South China Agric. Univ.* **2021**, *42*, 1–7.
6. Bargoti, S.; Underwood, J. Deep Fruit Detection in Orchards. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June May 2017; pp. 3626–3633.
7. Kang, H.; Chen, C. Fruit Detection and Segmentation for Apple Harvesting Using Visual Sensor in Orchards. *Sensors* **2019**, *19*, 4599. [CrossRef] [PubMed]
8. Sun, J.; He, X.; Ge, X.; Wu, X.; Shen, J.; Song, Y. Detection of Key Organs in Tomato Based on Deep Migration Learning in a Complex Background. *Agriculture* **2018**, *8*, 196. [CrossRef]
9. Jia, W.; Wang, Z.; Zhang, Z.; Yang, X.; Hou, S.; Zheng, Y. A Fast and Efficient Green Apple Object Detection Model Based on Foveabox. *J. King Saud Univ. Comput. Inf. Sci.* **2022**, *34*, 5156–5169. [CrossRef]
10. Jia, W.; Zhang, Y.; Lian, J.; Zheng, Y.; Zhao, D.; Li, C. Apple Harvesting Robot under Information Technology: A Review. *Int. J. Adv. Robot. Syst.* **2020**, *17*, 1–16. [CrossRef]
11. Tian, Y.; Duan, H.; Luo, R.; Zhang, Y.; Jia, W.; Lian, J.; Zheng, Y.; Ruan, C.; Li, C. Fast Recognition and Location of Target Fruit Based on Depth Information. *IEEE Access* **2019**, *7*, 170553–170563. [CrossRef]
12. Lin, G.; Tang, Y.; Zou, X.; Li, J.; Xiong, J. In-Field Citrus Detection and Localisation Based on RGB-D Image Analysis. *Biosyst. Eng.* **2019**, *186*, 34–44. [CrossRef]
13. Wang, C.; Lee, W.S.; Zou, X.; Choi, D.; Gan, H.; Diamond, J. Detection and Counting of Immature Green Citrus Fruit Based on the Local Binary Patterns (LBP) Feature Using Illumination-Normalized Images. *Precis. Agric.* **2018**, *19*, 1062–1083. [CrossRef]
14. Wang, D.; Song, H.; Tie, Z.; Zhang, W.; He, D. Recognition and Localization of Occluded Apples Using K-Means Clustering Algorithm and Convex Hull Theory: A Comparison. *Multimed. Tools Appl.* **2016**, *75*, 3177–3198. [CrossRef]
15. Zhang, C.; Liu, X.; Chen, B.; Yin, P.; Li, J.; Li, Y.; Meng, X. Insulator Profile Detection of Transmission Line Based on Traditional Edge Detection Algorithm. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 27–29 June 2020; pp. 267–269.
16. Li, X.; Pan, J.; Xie, F.; Zeng, J.; Li, Q.; Huang, X.; Liu, D.; Wang, X. Fast and Accurate Green Pepper Detection in Complex Backgrounds via an Improved Yolov4-Tiny Model. *Comput. Electron. Agric.* **2021**, *191*, 106503. [CrossRef]
17. Zhang, C.; Kang, F.; Wang, Y. An Improved Apple Object Detection Method Based on Lightweight YOLOv4 in Complex Backgrounds. *Remote Sens.* **2022**, *14*, 4150. [CrossRef]
18. Tu, S.; Pang, J.; Liu, H.; Zhuang, N.; Chen, Y.; Zheng, C.; Wan, H.; Xue, Y. Passion Fruit Detection and Counting Based on Multiple Scale Faster R-CNN Using RGB-D Images. *Precis. Agric.* **2020**, *21*, 1072–1091. [CrossRef]
19. Bhattarai, U.; Karkee, M. A Weakly-Supervised Approach for Flower/Fruit Counting in Apple Orchards. *Comput. Ind.* **2022**, *138*, 103635. [CrossRef]
20. Hao, Q.; Guo, X.; Yang, F. Fast Recognition Method for Multiple Apple Targets in Complex Occlusion Environment Based on Improved YOLOv5. *J. Sens.* **2023**, *2023*, e3609541. [CrossRef]
21. Weyler, J.; Milioto, A.; Falck, T.; Behley, J.; Stachniss, C. Joint Plant Instance Detection and Leaf Count Estimation for In-Field Plant Phenotyping. *IEEE Robot. Autom. Lett.* **2021**, *6*, 3599–3606. [CrossRef]
22. Chen, J.; Liu, H.; Zhang, Y.; Zhang, D.; Ouyang, H.; Chen, X. A Multiscale Lightweight and Efficient Model Based on YOLOv7: Applied to Citrus Orchard. *Plants* **2022**, *11*, 3260. [CrossRef]
23. Sun, M.; Xu, L.; Luo, R.; Lu, Y.; Jia, W. GHFormer-Net: Towards More Accurate Small Green Apple/Begonia Fruit Detection in the Nighttime. *J. King Saud Univ. Comput. Inf. Sci.* **2022**, *34*, 4421–4432. [CrossRef]
24. Gulzar, Y. Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique. *Sustainability* **2023**, *15*, 1906. [CrossRef]
25. Mamat, N.; Othman, M.F.; Abdulghafor, R.; Alwan, A.A.; Gulzar, Y. Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach. *Sustainability* **2023**, *15*, 901. [CrossRef]
26. Hamid, Y.; Wani, S.; Soomro, A.B.; Alwan, A.A.; Gulzar, Y. Smart Seed Classification System Based on MobileNetV2 Architecture. In Proceedings of the 2022 2nd International Conference on Computing and Information Technology (ICCIT), Tabuk, Saudi Arabia, 25–27 January 2022; pp. 217–222.
27. Aggarwal, S.; Gupta, S.; Gupta, D.; Gulzar, Y.; Juneja, S.; Alwan, A.A.; Nauman, A. An Artificial Intelligence-Based Stacked Ensemble Approach for Prediction of Protein Subcellular Localization in Confocal Microscopy Images. *Sustainability* **2023**, *15*, 1695. [CrossRef]
28. Häni, N.; Roy, P.; Isler, V. MinneApple: A Benchmark Dataset for Apple Detection and Segmentation. *IEEE Robot. Autom. Lett.* **2020**, *5*, 852–858. [CrossRef]

29. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
30. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 21–37.
31. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. CenterNet: Keypoint Triplets for Object Detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6569–6578.
32. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors 2022. *arXiv* **2022**, arXiv:2207.02696.
33. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
34. Xu, D.; Wu, Y. Improved YOLO-V3 with DenseNet for Multi-Scale Remote Sensing Target Detection. *Sensors* **2020**, *20*, 4276. [CrossRef] [PubMed]
35. Ghiasi, G.; Lin, T.-Y.; Le, Q.V. NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 7029–7038.
36. Wang, K.; Liew, J.H.; Zou, Y.; Zhou, D.; Feng, J. PANet: Few-Shot Image Semantic Segmentation with Prototype Alignment. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9196–9205.
37. Chen, J.; Mai, H.; Luo, L.; Chen, X.; Wu, K. Effective Feature Fusion Network in BIFPN for Small Object Detection. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021; pp. 699–703.
38. Saini, R.; Jha, N.K.; Das, B.; Mittal, S.; Mohan, C.K. ULSAM: Ultra-Lightweight Subspace Attention Module for Compact Convolutional Neural Networks. In Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass Village, CO, USA, 1–5 March 2020; pp. 1616–1625.
39. Guo, S.; Li, L.; Guo, T.; Cao, Y.; Li, Y. Research on Mask-Wearing Detection Algorithm Based on Improved YOLOv5. *Sensors* **2022**, *22*, 4933. [CrossRef]
40. Hodson, T.O. Root-Mean-Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not. *Geosci. Model Dev.* **2022**, *15*, 5481–5487. [CrossRef]
41. Yang, B.; Gao, Z.; Gao, Y.; Zhu, Y. Rapid Detection and Counting of Wheat Ears in the Field Using YOLOv4 with Attention Module. *Agronomy* **2021**, *11*, 1202. [CrossRef]
42. Zhang, P.; Li, D. EPSA-YOLO-V5s: A Novel Method for Detecting the Survival Rate of Rapeseed in a Plant Factory Based on Multiple Guarantee Mechanisms. *Comput. Electron. Agric.* **2022**, *193*, 106714. [CrossRef]
43. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430.
44. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications 2022. *arXiv* **2022**, arXiv:2209.02976.

agronomy

MDPI

Article

Maize (*Zea mays* L.) Stem Target Region Extraction and Stem Diameter Measurement Based on an Internal Gradient Algorithm in Field Conditions

Jing Zhou [1], Mingren Cui [1], Yushan Wu [1], Yudi Gao [1], Yijia Tang [1], Zhiyi Chen [1], Lixin Hou [1] and Haijuan Tian [2,*]

[1] College of Information Technology, Jilin Agricultural University, Changchun 130118, China; 012028@jlau.edu.cn (J.Z.); 20221117@mails.jlau.edu.cn (M.C.); 20221147@mails.jlau.edu.cn (Y.W.); 1912030221@mails.jlau.edu.cn (Y.G.); 2012030102@mails.jlau.edu.cn (Y.T.); 2012030104@mails.jlau.edu.cn (Z.C.); lixinh@jlau.edu.cn (L.H.)

[2] A Jilin Province Key Laboratory of Grain and Oil Processing, Jilin Business and Technology College, Changchun 130507, China

* Correspondence: 20070552@jlbtc.edu.cn

Abstract: The target region and diameter of maize stems are important phenotyping parameters for evaluating crop vitality and estimating crop biomass. To address the issue that the target region and diameter of maize stems obtained after transplantation may not accurately reflect the true growth conditions of maize, a phenotyping monitoring technology based on an internal gradient algorithm is proposed for acquiring the target region and diameter of maize stems. Observations were conducted during the small bell stage of maize. First, color images of maize plants were captured by an Intel RealSense D435i camera. The color information in the color image was extracted by the hue saturation value (HSV) color space model. The maximum between-class variance (Otsu) algorithm was applied for image threshold segmentation to obtain the main stem of maize. Median filtering, image binarization, and morphological opening operations were then utilized to remove noise from the images. Subsequently, the morphological gradient algorithm was applied to acquire the target region of maize stems. The similarity between the three types of gradient images and the manually segmented image was evaluated by pixel ratio extraction and image quality assessment indicators. Evaluation results indicated that the internal gradient algorithm could more accurately obtain the target region of maize stems. Finally, a checkerboard was employed as a reference for measurement assistance, and the stem diameter of maize was calculated by the pinhole imaging principle. The mean absolute error of stem diameter was 1.92 mm, the mean absolute percentage error (MAPE) was 5.16%, and the root mean square error (RMSE) was 2.25 mm. The R^2 value was 0.79. With an R^2 greater than 0.7 and a MAPE within 6%, the phenotyping monitoring technology based on the internal gradient algorithm was proven to accurately measure the diameter of maize stems. The application of phenotyping monitoring technology based on the internal gradient algorithm in field conditions provides technological support for smart agriculture.

Keywords: crop phenotype; maize; stem diameter; morphological gradient; target region

Citation: Zhou, J.; Cui, M.; Wu, Y.; Gao, Y.; Tang, Y.; Chen, Z.; Hou, L.; Tian, H. Maize (*Zea mays* L.) Stem Target Region Extraction and Stem Diameter Measurement Based on an Internal Gradient Algorithm in Field Conditions. *Agronomy* **2023**, *13*, 1185. https://doi.org/10.3390/agronomy13051185

Academic Editor: Roberto Marani

Received: 29 March 2023
Revised: 20 April 2023
Accepted: 21 April 2023
Published: 22 April 2023

1. Introduction

With the global population already above 8 billion, crop yield is expected to double by 2050 to meet global food demand. Breeding research is imperative to address food crises resulting from population growth and adverse climatic factors [1–3]. In breeding research, crop genotypes and phenotyping characteristics have received widespread attention from agronomists [4–6]. Traditional plant phenotyping monitoring technologies rely on ruler-based measurements [7,8], which are inefficient, labor-intensive, and subjective. Traditional phenotyping monitoring technologies have lagged far behind the rapid development of genomics research [9,10]. Plant phenotyping technology that depends solely on manual

detection and analysis can no longer meet the demands of modern agricultural development. Therefore, it is crucial to study low-cost, high-precision, and efficient phenotyping monitoring technologies.

As one of the world's major cereal crops, maize (*Zea mays* L.) has high nutritional and medicinal value [11–13]. Stem diameter of maize is a key agronomic trait related to yield and lodging resistance [14–16], and is an important indicator measured by agronomists. Ma et al. [17] used handheld laser scanners to scan maize plants. The point cloud data was processed by 3D reconstruction techniques. The stem diameter of potted maize at multiple growth stages was calculated by fitting spheres and cylinders. Although the grid method simplified the original point cloud data, the lack of color information in the data acquired by the device is not conducive to the study of plant phenotyping color characteristics. In comparison, depth cameras are more affordable and can simultaneously capture color and depth information of images. The time-of-flight (TOF) camera is a prototypical depth imaging device, which acquires depth-related data of objects through the evaluation of round-trip time of light. This camera has extensive applications in the agricultural domain. Chaivivatrakul et al. [18] used a TOF camera to obtain 3D point clouds of maize plants, and acquired stem diameter through 3D holographic reconstruction. However, due to the insufficient resolution of the TOF camera, the error between the measured stem diameter and the true value was relatively large. In addition, TOF cameras are sensitive to ambient light, so monitoring maize phenotyping parameters with TOF cameras is typically performed in laboratory environments. In light of the TOF camera's capacity to concurrently acquire both color and depth images of plants, employing this technology within a laboratory environment facilitates a more exhaustive investigation of crop phenotypic characteristics. Although plant phenotyping analysis in controlled environments is of great significance for plant science, the results differ significantly from the actual conditions of crops in field conditions. In recent years, the superior performance of RGB-D sensors have made it possible to extract the stem diameter of maize in field conditions. These sensors have high resolution, low cost, and simple operation, and have attracted widespread attention from scholars. Fan et al. [19] used the Intel RealSense D435i sensor to obtain 3D point cloud data of maize, estimating stem diameter through point cloud convex hull and planar projection, with high measurement accuracy. Vit and Shani [20] studied the filling rate of four different RGB-D sensors in maize stems, tomatoes, and plastic balls. Plastic balls were used to evaluate the ability of RGB-D sensors in the task of calculating stem diameter. The results showed that the Intel RealSense D435 sensor had the best filling rate for plastic balls at different distances and light intensities compared to the other three sensors, making it more suitable for phenotyping analysis in field conditions.

Extracting the target region of maize stems is a critical step in measuring stem diameter, and its effect directly influences the accuracy of stem diameter measurement. Moreover, agronomists can evaluate soil moisture and nutrient supply conditions by analyzing the smoothness and diameter of the target region of maize stems [21]. Among the numerous image edge detection algorithms, the morphological gradient algorithm precisely captures object edge contours. Han and Han [22] proposed an edge detection algorithm based on morphological gradient and maximum between-class variance (Otsu) to extract edge pixels of grayscale images. The performance of the morphological gradient and Otsu-based edge detection algorithm is superior to that of the Sobel and Canny algorithms. Wu and Li [23] proposed an improved watershed color image segmentation algorithm, which improved the original algorithm by combining the morphological gradient and Otsu algorithm. The improved algorithm accurately obtained object edge information with an accuracy rate greater than 0.98, which improved the robustness and applicability of the algorithm.

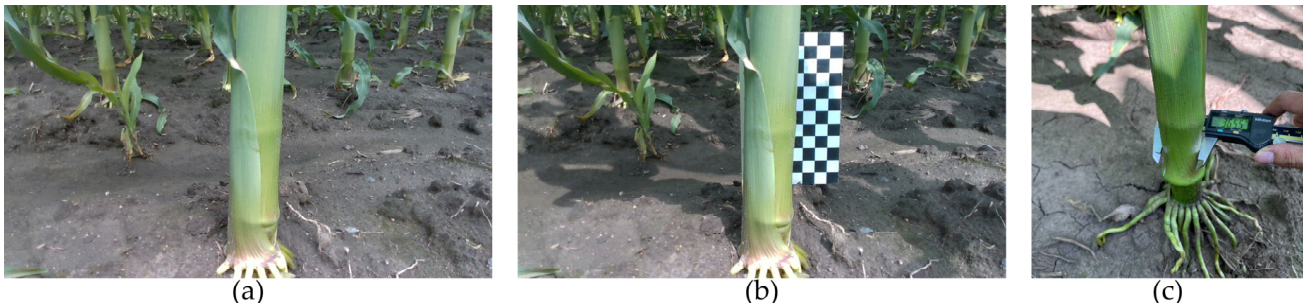
In this study, maize was as the research subject, and the Intel RealSense D435i sensor was used to obtain color images of maize stems in field conditions. The maize stems were extracted by the HSV + Otsu algorithm, and image noise was eliminated by median filtering, image binarization, and morphological opening operations. The internal gradient algorithm was applied to extract the target region of maize stems, and the stem diameter

was calculated by the pinhole imaging principle. This study combines the internal gradient algorithm with color space models, image segmentation technology, and image filtering technology to provide agronomists with more accurate phenotyping data and lays the foundation for subsequent research combining 3D point clouds.

2. Materials and Methods

2.1. Image Data Collection

Field experiments were conducted at the teaching and research base of Jilin Agricultural University in Changchun, Jilin Province, China. The maize cultivar selected for the study was Ji Nong Yu 218. Image collection occurred from 12:00 to 15:00 in July 2021 and July 2022, under clear weather conditions. The object of the study was maize stem images during the small bell stage. The maize planting region was 20 m long and 15 m wide, with a planting density of 67,000 plants per hectare. The plant spacing was 0.4 m, and the row spacing was 0.5 m. Four rows of maize were randomly selected in the experimental field for image collection. The canopy density could be neglected as there was almost no overlapping coverage in the early growth stage. The tools employed for image collection and stem diameter measurement included an Intel RealSense D435i camera, a shooting platform, a Vernier caliper, a checkerboard, and a laptop. The Intel RealSense D435i camera was placed on the shooting platform to capture maize plant images, with a distance of 0.5 m between the camera and the ground. To reduce interference from the ground and adjacent plants, the camera was tilted at a 40° angle from the vertical plane towards the maize plant. To preserve the overall morphology and appearance of the maize stem, the distance between the camera lens and the maize stem was 50 cm. The Vernier caliper was utilized to measure the true stem diameter of the maize. Based on the advice of agricultural experts, the second internode of the maize plant was designated as the region of interest [24,25]. Three random measurements of stem diameter were taken in this region, with the mean value of the three measurements considered the true stem diameter. The checkerboard, consisting of 84 alternating black and white squares with dimensions of 15 × 15 mm, was used as a reference for measurement assistance. To measure the maize stem diameter by the pinhole imaging principle, the camera captured two sets of images of the same maize plant at the identical angle: one set with the checkerboard and another without. The laptop was installed with the Intel RealSense SDK 2.0 development package under the Windows 10 operating system. The collected maize stem images were stored in the data collection terminal. The process of maize stem image data collection is shown in Figure 1.



(a) (b) (c)

Figure 1. Maize stem information collection: (**a**) field maize image; (**b**) field maize image with a checkerboard; (**c**) digital Vernier caliper measurement.

2.2. Image Segmentation and Filtering

In this study, the HSV color space model was selected. The HSV color space model has excellent robustness under varying illumination conditions [26–28]. The H value in the model effectively reflects the color information of plant stems, leaves, and other parts, serving as an important basis for separating plants from the background. The Otsu method was employed as the image segmentation algorithm in this paper, utilized for extracting the main stem of maize. The Otsu algorithm is commonly applied in processing plant images

under field conditions [29]. Since the experiment required the removal of background noise while preserving the edge information of maize stems, the median filtering was selected for the filtering algorithm for this paper. After filtering, the image was subjected a binarization process. Due to the presence of local noise in the binarized image, an opening operation was conducted on the image. First, a morphological erosion operation was applied to disconnect the weeds or other leaves from the stem. Subsequently, the image was subjected to a morphological dilation operation to fill in the disrupted regions, resulting in a denoised main stem of maize.

2.3. Morphological Gradient

The extraction of the maize stem target region is a pivotal step in maize phenotyping monitoring technology. The morphological gradient can separate the contours of the target region from the background, facilitating subsequent processing and analysis. The morphological gradient comprises the basic morphological gradient (BMG), external morphological gradient (EMG), and internal morphological gradient (IMG). The basic morphological gradient (BMG) refers to the difference image between the dilated image and the eroded image. The external morphological gradient (EMG) refers to the difference image between the dilated image and the original image. The internal morphological gradient (IMG) refers to the difference image between the original image and the eroded image. The corresponding formulas are shown in Equations (1)–(3).

$$\mathrm{BMG} = (\mathrm{A} \oplus \mathrm{B}) - (\mathrm{A} \ominus \mathrm{B}) \tag{1}$$

$$\mathrm{EMG} = (\mathrm{A} \oplus \mathrm{B}) - \mathrm{A} \tag{2}$$

$$\mathrm{IMG} = \mathrm{A} - (\mathrm{A} \ominus \mathrm{B}) \tag{3}$$

Here, A represents the input image, B represents the structuring element, $\oplus$ represents the dilation operation, and $\ominus$ represents the erosion operation.

2.4. Image Data Processing

The processing workflow of this study is illustrated in Figure 2.

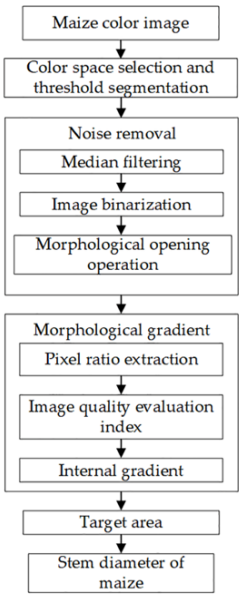


Figure 2. Flow chart of maize stem diameter measurement.

Three sets of field maize images were randomly selected from 60 samples for experimentation. Field maize images, HSV color space images, and images based on HSV and Otsu algorithms are shown in Figure 3. Median filtered images, binarized images, and images subjected to morphological opening operations are shown in Figure 4. Basic gradient images, external gradient images, and internal gradient images are shown in Figure 5.

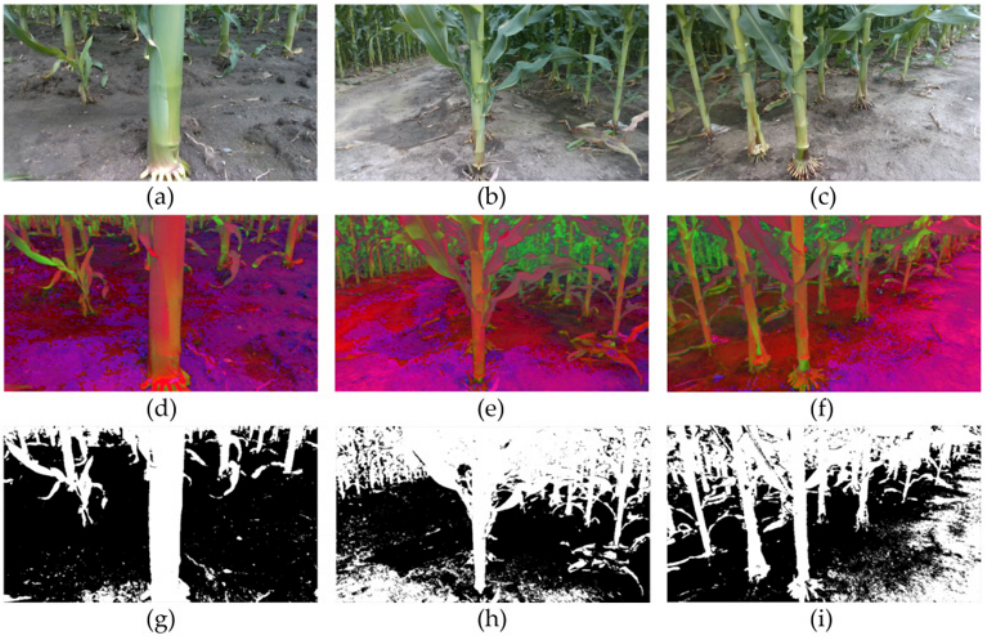


Figure 3. Image segmentation process: (**a**–**c**) original maize images; (**d**–**f**) hue saturation value (HSV) color space images; (**g**–**i**) HSV + maximum between-class variance (Otsu) images.

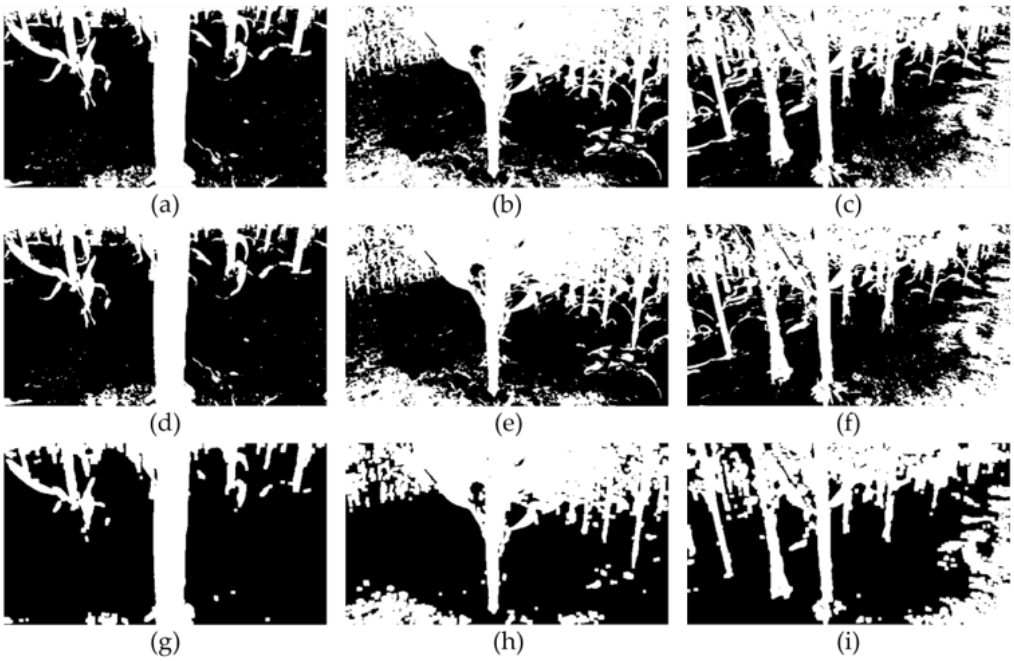


Figure 4. Image filtering process: (**a**–**c**) median filtered images; (**d**–**f**) binarized images; (**g**–**i**) images subjected to morphological opening operations.

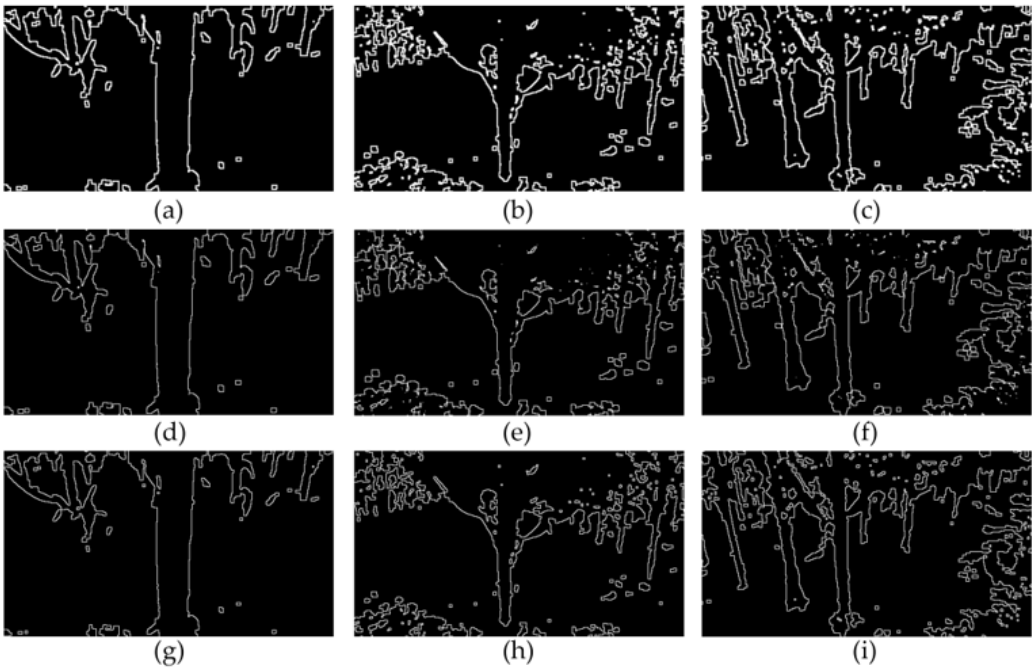


Figure 5. Image gradient process: (**a**–**c**) basic gradient images; (**d**–**f**) external gradient images; (**g**–**i**) internal gradient images.

2.5. Coordinate Extraction and Maize Stem Diameter Measurement

In maize gradient images, coordinates were semi-automatically extracted using the OpenCV library functions in Python. The maize stem diameter was extracted based on the pinhole imaging principle. A checkerboard was used as a reference object to facilitate measurements. Since both the original and processed images have a resolution of 1280 × 720, the checkerboard from the original image can be utilized to calculate the stem diameter in the processed image. This approach avoids the impact of an incomplete checkerboard after image processing on the experiment. Taking the internal gradient image, for example, the experiment was conducted in the middle of the second internode of the maize stem. A square was randomly selected on the checkerboard in the original maize image, with its left and right endpoint coordinates being (x1, y1) and (x2, y1), respectively. In the gradient image, a set of contour points was chosen with coordinates (x3, y1) and (x4, y1). Each square in the checkerboard has a side length of W, and the measured stem diameter of maize is denoted as S. S can be calculated by the proportion formula, as shown in Equation (4). This operation is repeated three times, and the average of the three measurements is used as the measured stem diameter of maize. The coordinate acquisition process is illustrated in Figure 6.

$$\frac{S}{x4 - x3} = \frac{W}{x2 - x1} \tag{4}$$

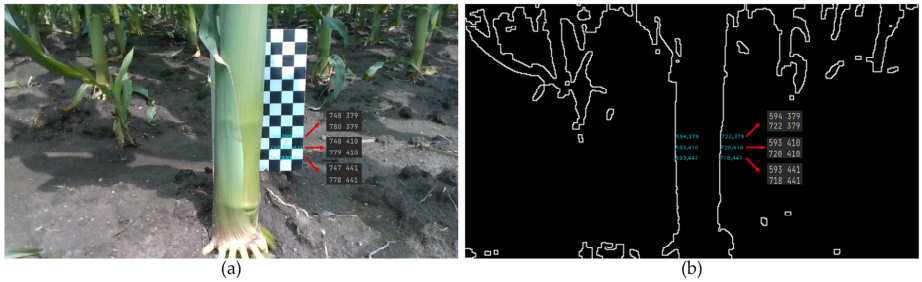


Figure 6. Coordinate acquisition process: (**a**) coordinate extraction of the checkerboard; (**b**) coordinate extraction of the maize stem contour.

2.6. Gradient Image Evaluation Method

In this study, two objective evaluation methods were employed to assess three types of gradient images. By comparing with manually segmented images, a gradient image more suitable for extracting the target region of maize stems was selected [30].

2.6.1. Pixel Proportion Extraction

In the field of computer vision, pixel proportion extraction is commonly employed to facilitate the identification of similar images [31]. This method was applied to determine the differences in pixel proportions of different color regions between gradient images and manually segmented images. If the difference between the two is small, it indicates that the gradient image can accurately detect the position and shape of the target region of maize stems. This method holds significant importance for the performance evaluation of the three gradient algorithms.

2.6.2. Image Quality Evaluation Metrics

Mean square error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM) are three metrics for evaluating image quality. They aid in assessing the similarity or differences between two images. MSE can be utilized to calculate the pixel differences between gradient images and manually segmented images, evaluating the distortion between them. The closer the MSE value is to 0, the smaller the difference between the gradient image and the manually segmented image, and the greater the similarity between the two images. PSNR is used to evaluate the relative error between gradient images and manually segmented images. The higher the PSNR, the greater the similarity between the two images. SSIM effectively simulates human perception of image quality and is employed to evaluate the structural similarity between two images. The SSIM values range from 0 to 1, representing no similarity to perfect consistency. Different metrics can evaluate model performance from various perspectives, and combining multiple metrics allows for a more comprehensive assessment of the model's performance. The formulas for MSE, PSNR, and SSIM are as follows:

$$\mathrm{MSE} = \frac{1}{\mathrm{M} \times \mathrm{N}} \sum_{\mathrm{i}=0}^{\mathrm{M}-1} \sum_{\mathrm{j}=0}^{\mathrm{N}-1} [\mathrm{I}(\mathrm{i},\mathrm{j}) - \mathrm{K}(\mathrm{i},\mathrm{j})]^2 \tag{5}$$

$$\mathrm{PSNR} = 10 \log_{10} \left(\frac{255^2}{\mathrm{MSE}} \right) \tag{6}$$

$$\mathrm{SSIM}(\mathrm{x},\mathrm{y}) = \frac{\left(2\mu_\mathrm{x}\mu_\mathrm{y} + \mathrm{c}_1\right)\left(2\sigma_{\mathrm{xy}} + \mathrm{c}_2\right)}{\left(\mu_\mathrm{x}^2 + \mu_\mathrm{y}^2 + \mathrm{c}_1\right)\left(\sigma_\mathrm{x}^2 + \sigma_\mathrm{y}^2 + \mathrm{c}_2\right)} \tag{7}$$

Here, M and N are the height and width of the test image, respectively. I(i, j) and K(i, j) are the grayscale values of the test image and the original image at point (x, y), respectively. For two images x and y, μ_x and μ_y denote the average values of x and y, respectively; σ_{xy} represents the covariance of x and y; and σ_x and σ_y are the standard deviations of x and y, respectively.

2.7. Evaluation Metrics for Stem Diameter Error

Mean absolute error, mean absolute percentage error (MAPE), root mean square error (RMSE), and coefficient of determination (R^2) serve as metrics for evaluating the accuracy of maize stem diameter measurements. The formulas for MAPE, RMSE, and R^2 are as follows:

$$\mathrm{MAPE} = \frac{1}{\mathrm{m}} \sum_{\mathrm{i}=1}^{\mathrm{m}} \left| \frac{\mathrm{L}_\mathrm{i} - \mathrm{S}_\mathrm{i}}{\mathrm{L}_\mathrm{i}} \right| \times 100\% \tag{8}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(L_i - S_i)^2} \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(L_i - S_i)^2}{\sum_{i=1}^{m}(L_i - \overline{L})^2} \tag{10}$$

Here, m represents the number of maize plant samples, S_i represents the measured stem diameter, L_i represents the true stem diameter, and $\overline{L}$ represents the average of true stem diameters.

3. Results

3.1. Analysis of Pixel Proportion Extraction Results

To reduce the interference from the ground and adjacent plants, the second internode of maize stems was selected as the evaluation region. The pixel proportion extraction results for the evaluation regions of three groups of original maize images, basic gradient images, external gradient images, internal gradient images, and manually segmented images are shown in Figure 7.

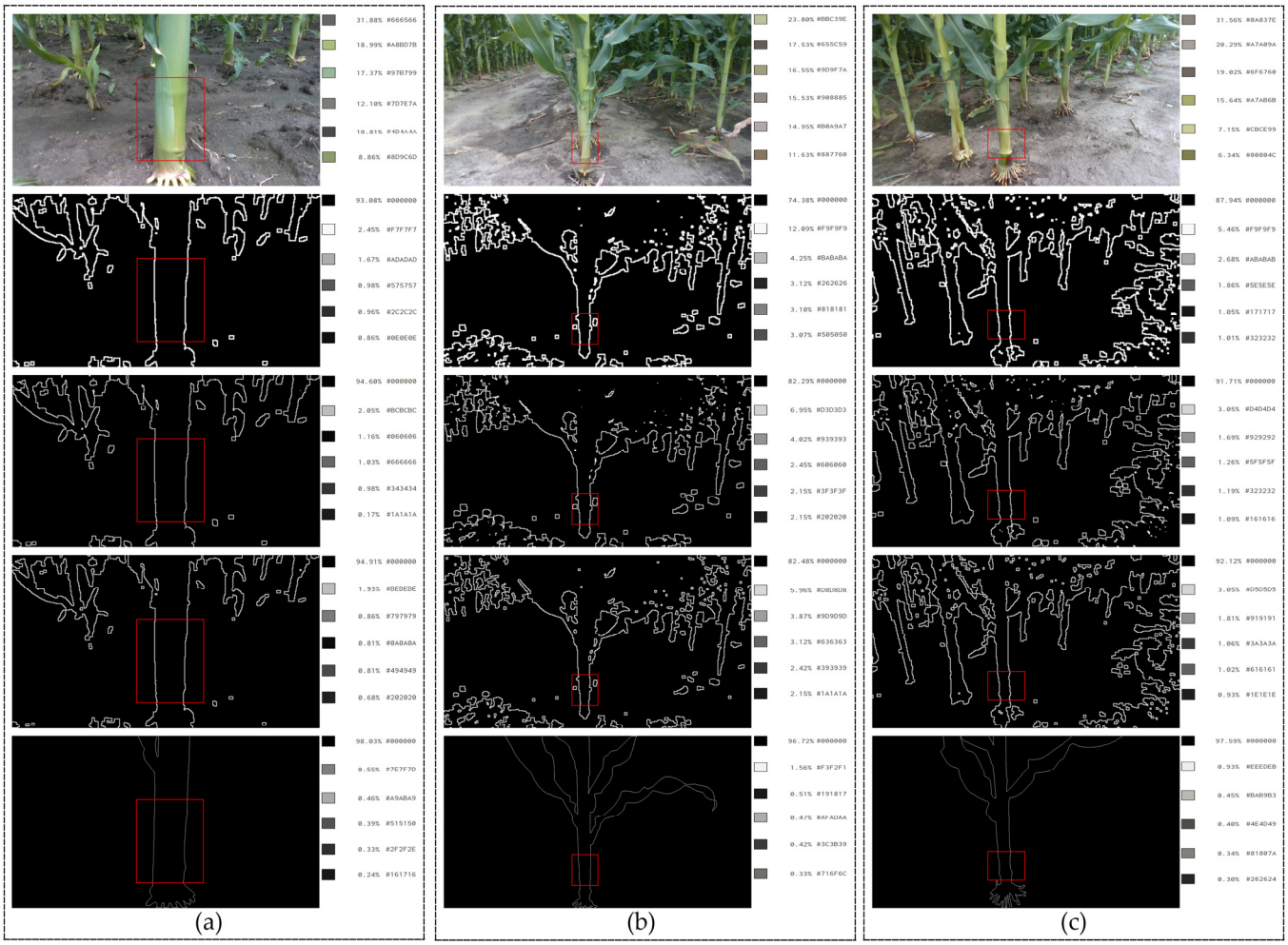


Figure 7. Pixel proportion extraction results: (**a**–**c**) Pixel proportions of three groups of maize original images, three types of gradient images, and manually segmented images.

A comparison between the three types of gradient images and manually segmented images reveals that the proportion of black pixels in the manually segmented images is 98.03%. The proportions of black pixels in the basic gradient, external gradient, and internal gradient images for the first group of maize are 93.08%, 94.60%, and 94.91%, respectively. For the second group of maize, these proportions are 87.94%, 91.71%, and 92.12%, respectively. For the third group of maize, the proportions are 74.38%, 82.29%, and 82.48%, respectively. Consequently, the proportion of non-black pixels in internal gradient images is closer to that in the manually segmented images.

3.2. Analysis of Image Quality Evaluation Results

The quality evaluation results for the three groups of maize basic gradient images, external gradient images, and internal gradient images are shown in Figure 8.

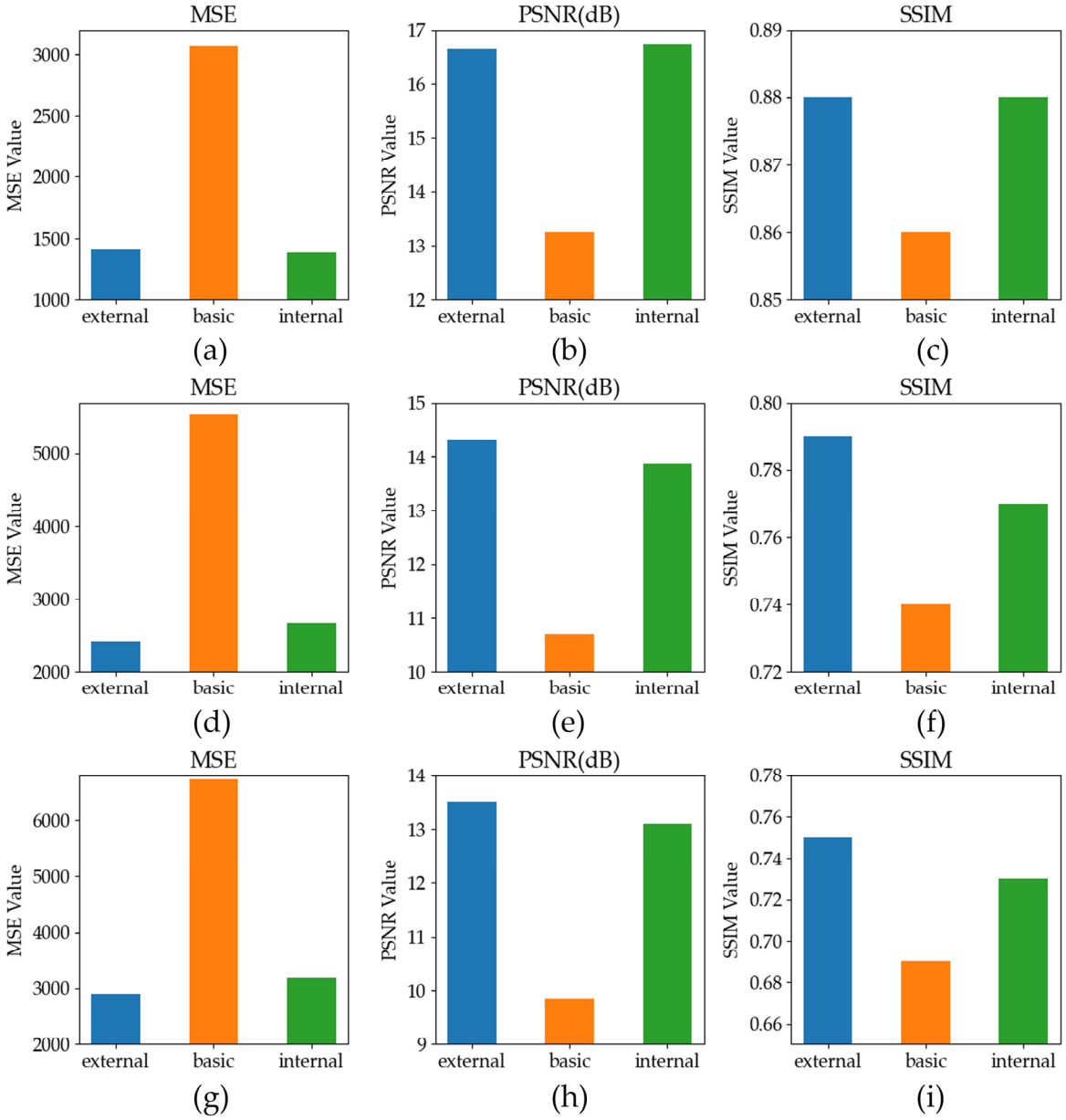


Figure 8. Image quality evaluation results based on three evaluation metrics: (**a**–**c**) Similarity between the three gradient images and manually segmented images for the first group of maize plants; (**d**–**f**) Similarity between the three gradient images and manually segmented images for the second group of maize plants; (**g**–**i**) Similarity between the three gradient images and manually segmented images for the third group of maize plants.

In the first group of maize basic gradient, external gradient, and internal gradient images, the MSE values are 3069.78, 1405.72, and 1378.63, respectively. The PSNR values are 13.26 dB, 16.65 dB, and 16.74 dB, respectively. The SSIM values are 0.86, 0.88, and 0.88, respectively. For the second group of maize, the MSE values are 5542.73, 2414.32, and 2668.17, respectively. The PSNR values are 10.69 dB, 14.3 dB, and 13.87 dB, respectively. The SSIM values are 0.74, 0.79, and 0.77, respectively. For the third group of maize, the MSE values are 6725.16, 2901, and 3187.7, respectively. The PSNR values are 9.85 dB, 13.51 dB, and 13.1 dB, respectively. The SSIM values are 0.69, 0.75, and 0.73, respectively. Thus, compared to basic gradient images, the internal gradient images and external gradient images are more similar to the manually segmented images.

Both gradient image evaluation methods have confirmed that internal gradient images and external gradient images are closer to manually segmented images, with no significant difference between them. However, the external gradient algorithm eliminates some pixels from the external edge contour of maize stems, affecting the smoothness of the maize stem edge contour and the accuracy of stem diameter measurement. In contrast, the internal gradient algorithm retains the pixels of the external edge contour of maize stems, only eliminating some pixels from the interior, without substantially altering the diameter of the maize stem target region. Therefore, the internal gradient algorithm is more suitable for extracting the target region of maize stems.

3.3. Error Analysis of Stem Diameter Measurement

In order to verify the accuracy of the internal gradient algorithm, 60 groups of field maize images were selected for error analysis, with the results shown in Table 1.

Table 1. Comparison of true and measured maize stem diameters.

Number	True Stem Diameter/ mm	Measured Stem Diameter/ mm	Absolute Error/mm	Number	True Stem Diameter/ mm	Measured Stem Diameter/ mm	Absolute Error/mm
1	36.09	36.50	0.41	31	34.40	36.32	1.92
2	34.11	33.63	0.48	32	33.06	30.00	3.06
3	31.98	28.94	3.04	33	39.23	42.19	2.96
4	27.53	28.91	1.38	34	34.06	34.74	0.68
5	37.29	38.83	1.54	35	34.36	31.76	2.60
6	38.20	37.43	0.77	36	43.80	40.00	3.80
7	34.54	39.79	5.25	37	38.16	35.71	2.45
8	35.08	36.29	1.21	38	32.87	29.32	3.55
9	32.08	33.16	1.08	39	35.12	32.05	3.07
10	30.73	31.65	0.92	40	39.61	38.48	1.13
11	38.05	38.77	0.72	41	40.86	39.29	1.57
12	36.22	34.00	2.22	42	39.05	40.43	1.38
13	37.11	34.18	2.93	43	42.10	38.89	3.21
14	32.67	36.33	3.66	44	36.48	35.22	1.26
15	44.33	45.56	1.23	45	32.98	33.26	0.28
16	35.17	34.15	1.02	46	38.31	36.88	1.43
17	36.60	35.11	1.49	47	36.32	33.13	3.19
18	28.50	28.73	0.23	48	36.21	36.88	0.67
19	41.97	40.26	1.71	49	39.16	39.55	0.39
20	45.98	48.75	2.77	50	39.67	38.28	1.39
21	38.19	39.32	1.13	51	37.23	37.86	0.63
22	47.57	51.08	3.51	52	37.18	34.29	2.89
23	41.28	42.69	1.41	53	38.95	41.25	2.30
24	51.78	52.27	0.49	54	35.56	32.14	3.42
25	39.18	35.91	3.27	55	38.87	38.75	0.12
26	53.11	54.17	1.06	56	43.38	44.17	0.79
27	41.39	43.13	1.74	57	34.57	37.89	3.32
28	43.18	40.59	2.59	58	33.26	37.17	3.91
29	40.46	42.19	1.73	59	35.18	37.83	2.65
30	43.54	46.67	3.13	60	33.47	34.62	1.15

According to the data presented in Table 1, the mean absolute error for maize stem diameter measurements is 1.92 mm, the MAPE is 5.16%, and the RMSE is 2.25 mm. The complex conditions encountered during image acquisition in field environments may yield detrimental effects on the measurement outcomes. Firstly, the irregularity of the field surface may result in discrepancies between the real and predetermined angles of the imaging equipment, subsequently influencing the precision of the stem diameter measurements. Moreover, situating the camera directly on the shooting platform to obtain maize images lacks sufficient stability. Wind interference may cause the shifts in the camera's position, resulting in increased measurement inaccuracies. To mitigate these concerns, a future design will incorporate an aluminum alloy mobile platform as a substitute for the existing imaging platform. A stable universal joint is mounted on the aluminum alloy bracket of the mobile platform, with the camera secured to the joint using screws. This not only reduces measurement errors caused by uneven ground but also ensures the stability of the imaging equipment.

In order to more intuitively present the relationship between the measured stem diameter and the true stem diameter, a linear regression analysis was conducted in this paper, and the results shown in Figure 9. The degree of linear fitting can be observed in the figure, and the coefficient of determination (R^2) was calculated to be 0.79. This result indicates that the maize stem diameter measured using the internal gradient algorithm demonstrates a relatively high level of accuracy, and there is a considerable degree of consistency between the true stem diameter and the measured stem diameter.

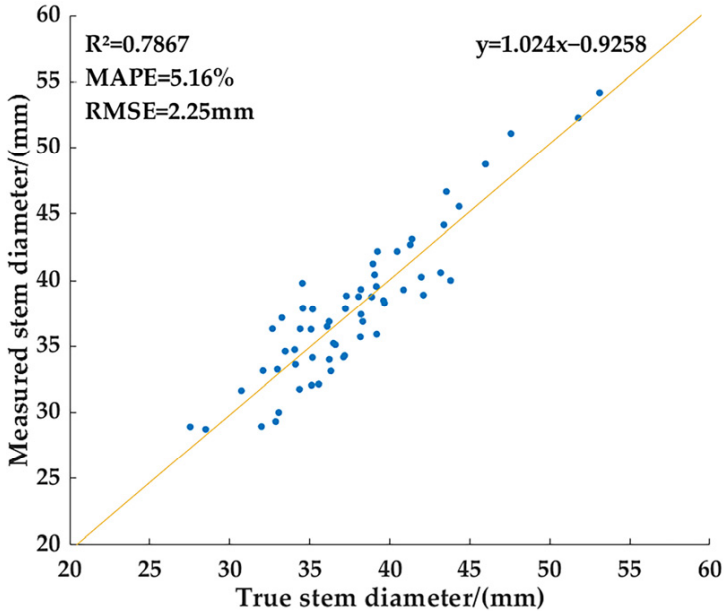


Figure 9. Linear regression between true and measured maize stem diameters.

4. Discussion

To extract the target region and stem diameter of maize stems in field conditions, this study employed the Intel RealSense D435i camera to acquire color images of maize during the small bell stage. A phenotyping monitoring technology based on internal gradient algorithm was utilized to extract the target region of maize stems. A checkerboard served as a reference for measurement assistance, and the maize stem diameter was calculated employing the pinhole imaging principle. The experimental results demonstrate the feasibility of obtaining the target region and stem diameter of maize stems by a phenotyping monitoring technology based on internal gradient algorithm.

Firstly, the current study is compared with research that utilizes LiDAR for measuring maize stem diameter. Jin et al. [32] employed LiDAR to collect point cloud data of maize plants and used a median normalized-vector growth (MNVG) algorithm to segment maize

stems and leaves, enabling the measurement of field maize height, leaf length, leaf width, stem height, stem diameter, and crown size. The R^2 were 0.91, 0.88, 0.81, 0.97, 0.65, and 0.96, respectively. The RMSE for stem diameter was 10 mm. Although this study demonstrated high accuracy in extracting other phenotypic parameters of maize, it showed relatively lower accuracy in measuring stem diameter. On the other hand, the current study achieved an R^2 of 0.79 for measuring maize stem diameter and decreased the RMSE to 5.29 mm, demonstrating higher measurement accuracy. Moreover, using the Intel RealSense D435i camera allowed the acquisition of color and texture information of maize stem that LiDAR could not obtain, thus providing more comprehensive data support for agricultural experts. Subsequently, the current study is compared with completed research. Zhou et al. [33] employed the HSV + Otsu algorithm to obtain the main stem of maize stems and measure their diameter, with mean absolute error, MAPE, and RMSE values of 4.30 mm, 10.76%, and 5.29 mm, respectively. Building on the existing algorithm, the current study employed median filtering, image binarization, and morphological opening operations to remove noise in the images. Additionally, an internal gradient algorithm was deployed to obtain the target region and diameter of maize stem. The mean absolute error, MAPE, and RMSE for stem diameter were 1.92 mm, 5.16%, and 2.25 mm, respectively. In comparison to prior research, the present study diminished measurement errors and more accurately identified the morphology and structure of maize stem.

However, several limitations of the proposed method need to be addressed. Firstly, the variety involved in this study is relatively singular. Maize includes various germplasm lines, comprising cultivated and wild varieties. Different varieties of maize exhibit differences in stem morphology and stem diameter. In this study, Ji Nong Yu 218 was selected as the experimental maize variety. However, selecting a single variety does not account for the phenotypic differences among different varieties. Therefore, future experiments will employ different maize varieties for phenotyping research and stem diameter measurements, in order to provide an important scientific foundation for maize breeding and improvement. Secondly, the observation period is relatively singular. The small bell stage is a rapid growth stage for maize and serves as a transition period between vegetative and reproductive stages. Maize at the small bell stage contains the features of both vegetative and reproductive stages. This study obtained the target region and stem diameter of maize stems only at the small bell stage using non-invasive imaging technology, without quantifying maize phenotypic parameters at other stages. However, phenotyping monitoring during vegetative and reproductive stages is crucial for improving maize yield and quality. Therefore, future research will employ the internal gradient algorithm to monitor maize phenotypes at multiple stages, assisting farmers and researchers to better formulate fertilization, irrigation, and harvesting strategies, reducing resource and labor waste, and improving maize yield and quality. Lastly, the monitored parameters are limited. This study only monitored the target region and diameter of maize stems, without obtaining growth information from other parts of the maize plant. However, phenotyping parameters such as leaf length, plant height, and leaf inclination angle are closely related to the photosynthetic efficiency, growth vigor, and water use efficiency of maize plants. Therefore, monitoring other maize phenotyping parameters will be the focus of future research. This will provide more comprehensive data support for precision agriculture and for improving the efficiency and sustainability of agricultural production.

5. Conclusions

This paper proposes a phenotyping monitoring technology based on an internal gradient algorithm for obtaining the target region and diameter of maize stems. Two image evaluation methods were used to assess the three types of target region extraction algorithms. The evaluation results indicated that the internal gradient algorithm is more suitable for obtaining the target region of maize stems. The mean absolute error for stem diameter was 1.92 mm, the MAPE was 5.16%, the RMSE was 2.25 mm, and the R^2 was 0.79. Since the R^2 value was greater than 0.7 and the MAPE was less than 6%, the stem diameter

of maize can be accurately measured. Monitoring variations in maize stem contours and stem diameter during the rapid growth period can reflect soil quality conditions. This information will help researchers and agricultural practitioners develop scientific measures to improve soil quality, such as proper fertilization, plowing, and deep loosening. These scientific management methods will enhance soil fertility, consequently improving maize lodging resistance and yield.

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

Funding: This study was supported by the National Key R&D Program of China (2022YFD2001602), the Education Department of Jilin Province (JJKH20220336KJ), the National Innovation and Entrepreneurship Training Project for University (China) (202110193017) and the Jilin Provincial Department of science and technology (20230202042NC).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We sincerely appreciate Jian Chen for his invaluable support during the field experiments.

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

References

1. Tilman, D.; Balzer, C.; Hill, J.; Befort, B.L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20260–20264. [CrossRef] [PubMed]
2. Ray, D.K.; Ramankutty, N.; Mueller, N.D.; West, P.C.; Foley, J.A. Recent patterns of crop yield growth and stagnation. *Nat. Commun.* **2012**, *3*, 1293. [CrossRef] [PubMed]
3. Ray, D.K.; Mueller, N.D.; West, P.C.; Foley, J.A. Yield trends are insufficient to double global crop production by 2050. *PLoS ONE* **2013**, *8*, e66428. [CrossRef] [PubMed]
4. Cobb, J.N.; DeClerck, G.; Greenberg, A.; Clark, R.; McCouch, S. Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theor. Appl. Genet.* **2013**, *126*, 867–887. [CrossRef]
5. Li, Y.; Sun, W.; Wang, Z.; Wan, C.; Zhang, J.; Qi, X.; Zhang, J. SDG102, a H3K36-Methyltransferase-Encoding Gene, Plays Pleiotropic Roles in Growth and Development of Maize (*Zea mays* L.). *Int. J. Mol. Sci.* **2022**, *23*, 7458. [CrossRef] [PubMed]
6. Reynolds, M.; Chapman, S.; Crespo-Herrera, L.; Molero, G.; Mondal, S.; Pequeno, D.N.; Pinto, F.; Pinera-Chavez, F.J.; Poland, J.; Rivera-Amado, C. Breeder friendly phenotyping. *Plant Sci.* **2020**, *295*, 110396. [CrossRef] [PubMed]
7. Montelatto, M.B.; Villamagua-Vergara, G.C.; De Brito, C.M.; Castanho, F.; Sartori, M.M.; de Almeida Silva, M.; Guerra, S.P.S. Bambusa vulgaris leaf area estimation on short-rotation coppice. *Sci. For.* **2021**, *49*, e3394. [CrossRef]
8. Long, W.; Li, Q.; Wan, N.; Feng, D.; Kong, F.; Zhou, Y.; Yuan, J. Root morphological and physiological characteristics in maize seedlings adapted to low iron stress. *PLoS ONE* **2020**, *15*, e0239075. [CrossRef]
9. Minervini, M.; Scharr, H.; Tsaftaris, S.A. Image analysis: The new bottleneck in plant phenotyping. *IEEE Signal Process. Mag.* **2015**, *32*, 126–131. [CrossRef]
10. Houle, D.; Govindaraju, D.R.; Omholt, S. Phenomics: The next challenge. *Nat. Rev. Genet.* **2010**, *11*, 855–866. [CrossRef]
11. Nuss, E.T.; Tanumihardjo, S.A. Maize: A paramount staple crop in the context of global nutrition. *Compr. Rev. Food Sci. Food Saf.* **2010**, *9*, 417–436. [CrossRef] [PubMed]
12. Rouf Shah, T.; Prasad, K.; Kumar, P. Maize—A potential source of human nutrition and health: A review. *Cogent Food Agric.* **2016**, *2*, 1166995. [CrossRef]
13. Siyuan, S.; Tong, L.; Liu, R. Corn phytochemicals and their health benefits. *Food Sci. Hum. Wellness* **2018**, *7*, 185–195. [CrossRef]
14. Liu, H.; Wang, H.; Shao, C.; Han, Y.; He, Y.; Yin, Z. Genetic Architecture of Maize Stalk Diameter and Rind Penetrometer Resistance in a Recombinant Inbred Line Population. *Genes* **2022**, *13*, 579. [CrossRef] [PubMed]
15. Xue, J.; Xie, R.; Zhang, W.; Wang, K.; Hou, P.; Ming, B.; Gou, L.; Li, S. Research progress on reduced lodging of high-yield and-density maize. *J. Integr. Agric.* **2017**, *16*, 2717–2725. [CrossRef]
16. Kelly, J.; Crain, J.L.; Raun, W. By-plant prediction of corn (Zea mays L.) grain yield using height and stalk diameter. *Commun. Soil Sci. Plant Anal.* **2015**, *46*, 564–575. [CrossRef]
17. Ma, X.; Zhu, K.; Guan, H.; Feng, J.; Yu, S.; Liu, G. Calculation method for phenotypic traits based on the 3D reconstruction of maize canopies. *Sensors* **2019**, *19*, 1201. [CrossRef]

18. Chaivivatrakul, S.; Tang, L.; Dailey, M.N.; Nakarmi, A.D. Automatic morphological trait characterization for corn plants via 3D holographic reconstruction. *Comput. Electron. Agric.* **2014**, *109*, 109–123. [CrossRef]
19. Fan, Z.; Sun, N.; Qiu, Q.; Li, T.; Feng, Q.; Zhao, C. In situ measuring stem diameters of maize crops with a high-throughput phenotyping robot. *Remote Sens.* **2022**, *14*, 1030. [CrossRef]
20. Vit, A.; Shani, G. Comparing rgb-d sensors for close range outdoor agricultural phenotyping. *Sensors* **2018**, *18*, 4413. [CrossRef]
21. Ismail, S.M.; Ozawa, K. Improvement of crop yield, soil moisture distribution and water use efficiency in sandy soils by clay application. *Appl. Clay Sci.* **2007**, *37*, 81–89. [CrossRef]
22. Han, L.; Han, A. An improved edge detection algorithm based on morphological operators and gradient. *J. Comput. Theor. Nanosci.* **2015**, *12*, 1121–1125. [CrossRef]
23. Wu, Y.; Li, Q. The Algorithm of Watershed Color Image Segmentation Based on Morphological Gradient. *Sensors* **2022**, *22*, 8202. [CrossRef] [PubMed]
24. Zhang, Y.; Liu, P.; Zhang, X.; Zheng, Q.; Chen, M.; Ge, F.; Li, Z.; Sun, W.; Guan, Z.; Liang, T. Multi-locus genome-wide association study reveals the genetic architecture of stalk lodging resistance-related traits in maize. *Front. Plant Sci.* **2018**, *9*, 611. [CrossRef] [PubMed]
25. Hui, F.; Xie, Z.; Li, H.; Guo, Y.; Li, B.; Liu, Y.; Ma, Y. Image-based root phenotyping for field-grown crops: An example under maize/soybean intercropping. *J. Integr. Agric.* **2022**, *21*, 1606–1619. [CrossRef]
26. Yang, W.; Wang, S.; Zhao, X.; Zhang, J.; Feng, J. Greenness identification based on HSV decision tree. *Inf. Process. Agric.* **2015**, *2*, 149–160. [CrossRef]
27. Hamuda, E.; Mc Ginley, B.; Glavin, M.; Jones, E. Automatic crop detection under field conditions using the HSV colour space and morphological operations. *Comput. Electron. Agric.* **2017**, *133*, 97–107. [CrossRef]
28. Chaves-González, J.M.; Vega-Rodríguez, M.A.; Gómez-Pulido, J.A.; Sánchez-Pérez, J.M. Detecting skin in face recognition systems: A colour spaces study. *Digit. Signal Process.* **2010**, *20*, 806–823. [CrossRef]
29. Hamuda, E.; Glavin, M.; Jones, E. A survey of image processing techniques for plant extraction and segmentation in the field. *Comput. Electron. Agric.* **2016**, *125*, 184–199. [CrossRef]
30. Meyer, G.E.; Neto, J.C. Verification of color vegetation indices for automated crop imaging applications. *Comput. Electron. Agric.* **2008**, *63*, 282–293. [CrossRef]
31. Sun, Y.; Mu, Y.; Feng, Q.; Hu, T.; Gong, H.; Li, S.; Zhou, J. Deer body adaptive threshold segmentation algorithm based on color space. *Comput. Mater. Contin.* **2020**, *64*, 1317–1328. [CrossRef]
32. Jin, S.; Su, Y.; Wu, F.; Pang, S.; Gao, S.; Hu, T.; Liu, J.; Guo, Q. Stem–leaf segmentation and phenotypic trait extraction of individual maize using terrestrial LiDAR data. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 1336–1346. [CrossRef]
33. Zhou, J.; Wu, Y.; Chen, J.; Cui, M.; Gao, Y.; Meng, K.; Wu, M.; Guo, X.; Wen, W. Maize Stem Contour Extraction and Diameter Measurement Based on Adaptive Threshold Segmentation in Field Conditions. *Agriculture* **2023**, *13*, 678. [CrossRef]

agronomy

Review

Current Optical Sensing Applications in Seeds Vigor Determination

Jian Zhang [1,2,*], Weikai Fang [3], Chidong Xu [3], Aisheng Xiong [4], Michael Zhang [5], Randy Goebel [5] and Guangyu Bo [3,*]

1 Faculty of Agronomy, Jilin Agricultural University, Changchun 130108, China
2 Department of Biology, University of Columbia Okanagan, Kelowna, BC V1V 1V7, Canada
3 Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China; weikaif@cmpt.ac.cn (W.F.); xcd@aiofm.ac.cn (C.X.)
4 Crop Genomics and Bioinformatics Center and National Key Laboratory of Crop Genetics and Germplasm Enhancement, College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China; xiongaisheng@njau.edu.cn
5 XAI Lab, Department of Computing Science, Alberta Machine Intelligence Institute, University of Alberta, Edmonton, AB T6G 2E8, Canada; mlzhang@ualberta.ca (M.Z.); rgoebel@ualberta.ca (R.G.)
* Correspondence: jian.zhang@ubc.ca (J.Z.); ayli@cmpt.ac.cn (G.B.)

Abstract: Advances in optical sensing technology have led to new approaches to monitoring and determining crop seed vigor. In order to improve crop performance to secure reliable yield and food supply, calibrating seed vigor, purity, germination rate, and clarity is very critical to the future of the agriculture/horticulture industry. Traditional methods of seed vigor determination are lengthy in process, labor intensive, and sometimes inaccurate, which can lead to false yield prediction and faulty decision-making. Optical sensing technology offers rapid, accurate, and non-destructive calibration methods to help the industry develop accurate decisions for seed usage and agronomic evaluation. In this review, we hope to provide a summary of current research in the optical sensing technology used in seed vigor assessments.

Keywords: seed vigor; spectral detection technology; image detection technology; digital agriculture

Citation: Zhang, J.; Fang, W.; Xu, C.; Xiong, A.; Zhang, M.; Goebel, R.; Bo, G. Current Optical Sensing Applications in Seeds Vigor Determination. *Agronomy* **2023**, *13*, 1167. https://doi.org/10.3390/agronomy13041167

Academic Editor: Gamal Elmasry

Received: 1 February 2023
Revised: 9 April 2023
Accepted: 17 April 2023
Published: 20 April 2023

1. Introduction

1.1. Definition of Seed Vigor

The quality of seeds is the key factor in accurately determining the quality and yield of crops. In a broad sense, seed quality is a comprehensive evaluation of single or multiple indicators such as vigor, purity, germination rate, clarity, and water content. Among the various evaluation indicators of seed quality, seed vigor can often comprehensively determine whether crop seeds can germinate normally, the degree of uniformity of emergence, and the ability of grown plants to resist diseases. Therefore, the overall term "seed vigor" is often used as a general indicator for evaluating seed quality [1]. The development of a scientific definition of seed vigor has always been closely related to the developmental level of agricultural technology and has gone through its own relatively long development stage. As early as 1950, the International Seed Testing Association (ISTA) proposed the concept of seedling vigor and initiated active research on "seed vigor". However, it was not until 1977 that ISTA formally defined the definition of seed vigor: "Seed vigor refers to the comprehensive performance of the activity intensity and seed characteristics of seeds or seeds during germination and emergence, and those that perform well are high vigor seeds, and those that perform poorly are low vigor seeds" [2]. This definition of seed vigor is essentially a comprehensive concept composed of multiple indicators such as germination rate, germination potential, field emergence rate, stress resistance, and plant hardiness.

Seed vitality is determined during the process of seed development. The basis for the formation of seed vitality is the measurement of the accumulation of a variety of

substances in the seed. With the maturity of the seed, the starch, protein, lipid, and other substances in the seed gradually accumulate, and the seed vitality increases until the seed is physiologically mature. At that point, the enzymes in the seed body are inactivated, the protein and nucleic acid are closely combined, the metabolism of the seed is continuously reduced, the embryo enters a quiescent state, and the seed obtains maximum vitality. With the typical modern practice of extending seed storage time, various metabolisms during seed storage can cause a series of changes, such as component changes, membrane structure damage, and loss of enzyme activity. The seeds enter a natural aging process, and the seeds lose their genetic integrity, thus reducing seed vigor [3].

Under a wide range of field conditions, seed vigor is the basis for sound performance crops and high yields and is the main indicator to measure field emergence and seed usage. The highly vigorous seeds typically show strong stress resistance, strong vitality, and storage resistance, which can significantly increase the germination rate of seeds, shorten their emergence time, and increase production and income. However, seeds with low vigor have poor stress resistance, long emergence time, and low seedling rate, which are likely to cause major losses to agricultural production. Therefore, in-depth research on the regulation mechanism of seed vigor, research and development of accurate and rapid seed vigor testing technology, and breeding of new varieties with high vigor are the only ways to promote the sustainable and healthy development of the seed industry.

1.2. Significance of Seed Viability Detection

The issue of seed vigor has attracted more attention from the agricultural fields of many countries across the world. This is particularly important because of the continuous growth of the global population, the continuous reduction of arable land, and the deteriorating agricultural production environment, especially in developing countries such as China and India. In China, the proportion of agricultural product yield reduction due to issues of seed vigor is between 10% and 20% per year [4]. The key to solving the above problems is to sow seeds with higher vigor on limited land resources and fulfill efficient agricultural production. Therefore, before the seeds are sown, accurate detection of seed vigor can improve the yield and quality of crops and help to achieve the objective of food production supply. It is obvious that accurate knowledge of seed vitality information is a necessary means to ensure agricultural production potential.

On the other hand, realizing the rapid identification and prediction of seed vigor will help to more deeply understand the different genetic mechanisms of high-vigor seeds, as well as the impact mechanisms of different maturity periods, storage conditions, planting conditions, and climatic conditions [5]. Improved understanding of the principles of deterioration, finding ways to improve seed vigor, and cultivating new varieties with high vigor will help realize these ideas. In production practice, it is beneficial to provide basic monitoring data in the research process of improved seed breeding, to prevent the circulation of uncertified and poor-quality seeds in the process of seed trade, and to guide seed companies to master effective storage management measures to improve seed processing and production efficiency. Confirming the vitality of this connection will improve the quality of the seeds.

In recent years, a lot of work has been done on seed vigor in the global agricultural field, and many countries have listed seed vigor as a key research and developmental topic of agricultural scientific research, which has become an active study in the field of digital agriculture. In China, colleges and research institutes such as the Institute of Botany, Chinese Academy of Sciences, Chinese Academy of Agricultural Sciences, China Agricultural University, and Jilin Agricultural University have also conducted significant research on seed vitality. To summarize, whether it is basic scientific research or the development of vigor improvements in the seed industry, there is an urgent need for high-precision and rapid detection of seed vigor. It is worth noting that the formation and development stages of seed vigor are not only affected by internal factors such as complex genetic regulation and metabolism but also easily affected by external factors such

as growth environment and storage environment. Therefore, non-destructive online or continuous detection of seed vigor has always been a technical bottleneck in the field.

1.3. Seed Vigor Detection Method

The International Seed Testing Association (ISTA) and Association of Official Seed Analysts (AOSA) have compiled the standard of "International Rules for Seed Testing" [6] and the Seed Vigor testing handbook [7], which includes at least eight kinds of seed viability determination methods such as conductivity measurement. This article mainly discusses the rapid detection methods and does not discuss the direct evaluation methods of seed vigor, such as germination assays. The non-destructive and rapid detection method of seed vitality mainly focuses on a variety of physical and chemical detection methods to measure the physical and chemical, image, spectral, and electrical characteristic parameters related to seed vitality and subsequently establishes the relationship between seed vitality and the measured characteristic parameters [8]. Related methods mainly include (1) detection methods based on the physical and chemical properties of seeds, (2) detection methods based on electrical properties [9], and (3) detection methods based on optical characteristics [10].

1.3.1. Seed Vitality Detection Technology Based on Physical and Chemical Characteristics

With the typical industry increase of seed aging time, the content of soluble protein and the activity of protective enzymes in seeds show an obvious decline. At this time, the content of hydrogen peroxide reductase, superoxide dismutase, superoxide dismutase, etc., in the seeds can be detected by conventional physical and chemical testing methods, such as tetrazolium staining, and the content of related enzymes in the seeds can be quantitatively analyzed. The detection results of physical and chemical methods have high precision and high reliability, but these methods require professionals to operate and apply special reagents, which is not conducive to rapid on-site monitoring of seed vigor but is rather considered as laboratory analysis.

1.3.2. Seed Vitality Detection Technology Based on Electrical Characteristics

Seeds are hygroscopic dielectrics. When seeds contain more water, their resistivity is relatively small, and the seeds can be measured as good conductors of electricity. When the seeds lose water and dry, they show higher resistivity, which diminishes conductivity. This research results show that the dielectric constant is negatively correlated with the vigor of seeds, and with the help of physical testing methods of the dielectric constant, the change of the dielectric constant can be used to identify the non-destructive detection of seed vigor. It should be noted that the measurement of conductivity takes more than 24 h and usually does not have an online detection capability on a production site. At the same time, the measurement of conductivity needs to follow a precise operation process, which is more suitable for laboratory analysis scenarios.

1.3.3. Seed Vitality Detection Technology Based on Optical Characteristics and Artificial Intelligence Methods

With the advancement of agricultural science and technology and the integration of multi-disciplines, emerging technologies are constantly being experimentally applied to the field of agricultural science and technology innovation. Many scholars regard the cross-integration of optics and biology as a foundation for new research for the non-destructive evaluation of seed vitality. Modern optical analysis technology is used to identify the material composition and appearance characteristics by analyzing the different reflections of the material in the visible light spectrum. For plant seeds, there are substances that determine their vigor, which can be manifested as different optical parameters. These differences can characterize the changes in the content of vigor substances during the process of seed vigor changes, and based on this, they can be measured. For grading of vitality, for example, Bi et al. used deep learning with machine vision to exploit the Swin Transformer to improve maize seed recognition [10]. From a large volume of observation

data obtained from sensing equipment (i.e., both camera and special optical devices), they develop spectral information processing and feature extraction methods, then extract those spectral and image characteristics that reflect the level of seed vitality. This provides the basis to establish an accurate mapping relationship between spectral and image features and seed vitality values. Note that improving the accuracy of seed viability prediction model analysis is the ultimate goal of seed viability detection data processing. In order to achieve these objectives, some advanced data and information processing methods are needed; for example, mature machine learning algorithms have been applied to hyperspectral data analysis in the construction of spectral feature analysis and vitality identification models of corn and habit seeds [10]. At the same time, deep learning is effective and popular for dealing with complex classification problems in large volumes of data, especially for seed viability feature construction models. Similarly, some scholars have used deep learning algorithms to analyze hyperspectral imaging data of cotton, loofah, and rice and constructed a discriminant analysis model for plant disease detection [11]. This work has produced stable and accurate vigor identification results. With the rapid development of computer technology, relevant algorithms have quickly emerged, so there will be more advanced methods developed and applied in the direction of seed vitality prediction models.

2. Progress in the Detection of Seed Vigor Based on Modern Optical Methods

Optics is an ancient science, but since the beginning of this century, with the advancement of laser technology, optical engineering, and information technology, optical detection technology has gained a resurgence. The theoretical basis for the application of modern optical methods to the detection of seed vigor is that light irradiation on the surface of seeds will produce reflection, scattering, and transmission phenomena, accompanied by changes in parameters such as frequency, phase, and polarization state. In different vigor stages, the contents and content of nucleic acid, starch, enzyme, and protein inside the seed have a predictable change, and the reflected, scattered, and transmitted light information generated by the interaction with the light field can inform the internal composition and quality of the seed.

Note that modern optical methods belong to the non-destructive detection method of seed vigor: optical methods can measure many parameters related to seed vigor. The changes in these parameters can indirectly reflect the status of seed vigor. The technical process for optical method detection research is to first collect the light field information of single or group seeds after the interaction with incident light, analyze the difference of light field information of seeds with different vigor, and then obtain accurate seed vigor through standard germination tests and other methods. By establishing a mapping between the light field information measures and seed vigor value, a seed vigor evaluation system is formed. Ultimately, this creates an overall evaluation system that can be used to predict seed vigor.

Because the traditional methods of seed vigor are based on physical, chemical, and electrical characteristics, the operation of detection equipment is complicated, the detection process is highly restricted by the external environment, and the detection accuracy is highly dependent on the operating experience of the personnel. In that context, the detection samples are easily damaged, and the detection period is long and time-consuming. The improvement trajectory of such methods can no longer meet the needs of the rapid development of the modern seed industry market. Therefore, vigor detection technologies must be developed in the direction of non-destructive, rapid, and high-precision [5]. Within this direction, optical detection methods do not require pretreatment of the seeds and, at the same time, can obtain the internal chemical information without destroying the seed. Therefore, it has unique technical advantages in the direction of accurate, non-destructive, fast, and even online detection of seed vitality. Existing optical detection methods for seed vitality mainly include machine vision detection, near-infrared spectroscopy, hyperspectral detection, Raman spectroscopy, fluorescence spectroscopy, and seed exhalation gas spectroscopy [12].

2.1. Machine Vision Method

Machine vision technology has advanced to integrate many research fields, such as image acquisition, image processing, and machine learning. During research on the detection of seed vigor, researchers have found that seed vigor is not only related to the genetic characteristics of the seeds but also closely related to the appearance and physical characteristics of the seeds, such as size, color, and texture. Machine vision technology working in the visible light band is essentially a detection method based on appearance features. It mainly simulates the human eye through image sensors such as charge-coupled imaging capture devices (CCDs), which collect images of seeds and then use computer technologies such as edge detection and feature extraction to simulate human vision. This provides for the extraction of physical information such as shape features, color features, and texture features of seeds, which can be combined with standard germination experiments of seed vigor. The combination of these techniques establishes the basis for a machine vision detection model of seed vigor to evaluate the status of seed vigor.

Since the 1990s, studies have launched in-depth research on the application of machine vision technology in agriculture. For example, McCormac et al. developed a lettuce seed vigor detection system based on machine vision technology [4]. The results of the study showed that the image analysis technique has the potential to determine the vigor of seeds. In China, scholars have successively developed similar methods for pepper seeds [5], which exploits a machine vision detection system for the vigor of perilla, wheat, and other seeds. The accuracy rate of the detection results is usually between 80 and 90%, and this system provides the basis for an online detection application for vegetable seed vigor [13].

Traditional machine vision technology mostly works in the visible light band. With the advancement of infrared devices, infrared thermal imaging technology has significantly advanced. As an extension of traditional machine vision technology, infrared methods can be applied to the field of non-destructive detection of seed vitality. Because of the influence of water absorption, respiration, and other metabolic activities of seeds during seed germination, the surface temperature will exhibit small changes which infrared thermal imaging technology can quantify. These temperature changes are closely related to seed vitality [14], thus distinguishing the seed differences in temporal vitality intervals [15].

Analyzing the existing research literature, the current problems in the application of machine vision technology in the field of seed vigor detection focus mainly on the fact that, when assessing seed vigor indicators, current machine vision methods extract vigor-related parameters from only the physical parameters of the seed surface features. However, the volume of indicative information is relatively small, and there is room for improvement. Generally speaking, with the advantage of inexpensive key equipment components and relatively mature algorithms, machine vision technology has significant promise for further development.

2.2. Infrared Spectroscopy Technology

With the help of the interaction between light and matter, spectroscopy can extract the internal vigor components and content information of seeds, so it has greater application potential in the detection of seed vigor. Among various spectroscopic techniques, near-infrared spectroscopy (NIRS) has been extensively studied. NIRS belongs to the class of imaging techniques known as molecular absorption spectroscopy. The principle is that specific functional groups in different molecules produce specific absorption indicators at specific wavelengths, which can be exhibited as absorption spectra in broadband light sources. As mentioned above, during the storage of seeds, the internal protein, starch, fat, and cellulose components will change, which will cause changes in seed vigor. The above substances are rich in hydrogen-containing groups such as C-H (aliphatic hydrocarbon), C-H (aromatic hydrocarbon), C=O (carboxyl group), OH (hydroxyl group), and N-H (amino group). Because of abundant absorption properties, one can identify obvious differences in absorption wavelength and intensity [16]. For example, Jin et al. extracted the difference

in near-infrared spectra between viable rice seeds and inactive seeds [17], which more intuitively reflects the above characteristics; the results are shown in Figure 1.

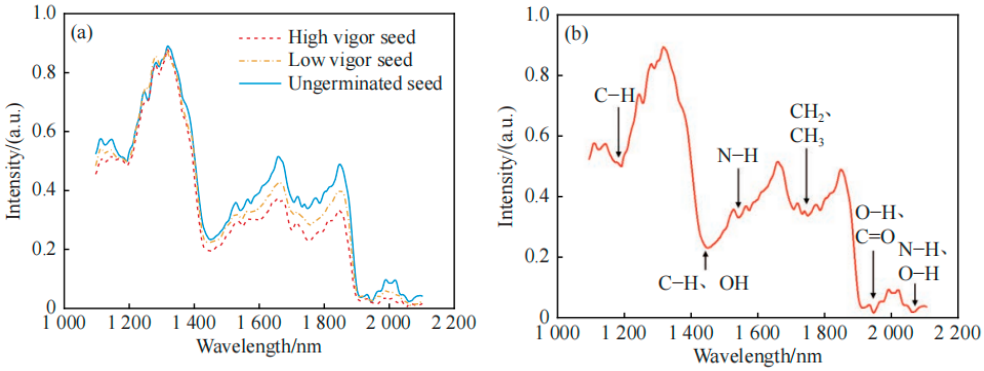


Figure 1. Near-infrared absorption spectrum of rice seeds; (**a**) The average curves of transmission spectra of rice seeds with different vigor characteristics; (**b**) Near-infrared absorption spectrum of high vigor seeds.

As shown in Figure 1, the spectra of viable and non-viable seeds have significant peak/valley differences in the bands of 1150–1220 nm, 1410–1450 nm, 1510–1540 nm, 1660–1800 nm, 1910–1950 nm, and 1980–2100 nm. The valleys represent groups such as amino N-H, hydroxyl O-H, and carbonyl C=O, respectively (Figure 1b). It can be seen that the changes in protein, starch, fat, and other components during the aging process of seeds can be characterized by this difference of infrared absorption spectrum, so as to extract identification characteristics of seed vitality.

Infrared spectroscopy, as a rapid non-destructive testing method, has previously attracted the attention of scholars earlier. For example, Tigabu et al. first applied near-infrared spectroscopy analysis technology to the detection of seed vigor. Their experiment obtained pine seeds with different vigor parameters that varied through accelerated aging. With the help of infrared spectroscopy technology, they successfully distinguished aged and unaged pine seeds. The accuracy rates of pine seed identification at the same level reached 80%, 90%, and 75%, respectively [8]. Al-Amery et al. used near-infrared spectroscopy to measure the standard germination rate and seed vigor of soybean seeds. By establishing a partial least squares prediction model, the seed vigor grades were classified, and two kinds of soybean seeds with low vigor and high vigor were identified. Note that the corrected rates were 80%~100% and 96.3%~96.6%, respectively. This is also the first non-destructive prediction model of near-infrared spectrum soybean seed vigor provided by the academic field [9]. NIRS technique has been reported for its application on seed non-destructive evaluation with pepper [12], watermelon [13], and tomato [14]. These studies helped to establish the viability of NIRS models of seed quality and have confirmed their application for seed quality evaluation.

Compared with the traditional infrared spectroscopy technology, the Fourier transform infrared spectroscopy technology that has recently emerged provides more powerful spectral analysis capabilities. For example, Larios et al. used Fourier transform infrared spectroscopy to study the vigor of soybean seeds and established a recognition model using modeling methods such as support vector machines. Their results showed that, under laboratory conditions, the correct rate of classification and identification of soybean seeds with different vigor could reach 100% [11].

In October 2006, the first near-infrared spectroscopy conference was held in China. This meeting became a milestone in the history of China's near-infrared technology development. Since then, a number of excellent research reports on the application of infrared spectroscopy to seed vigor have emerged. For example, Ling et al. used near-infrared supercontinuum laser spectroscopy to study three kinds of rice with different seed vigor, and the results showed that the accuracy of the prediction results reached more than

95% [18]. Fan et al. used near-infrared spectroscopy [19], studied the vigor of wheat seeds, and the results showed that the prediction accuracy of the model was not less than 84%, indicating that the near-infrared spectroscopy technology has the potential to predict the vigor of wheat seeds. Overall, near-infrared spectroscopy technology is widely used in wheat, rice, corn, soybean, and other crops, fir, Masson pine, Slash pine, Loblolly pine, and other forest seeds, as well as tomatoes, in the study of vigor detection of vegetable seeds [20,21].

Summarizing the existing problems in the application of infrared spectroscopy technology to the detection of seed vitality has identified the following aspects. On the one hand, infrared spectroscopy technology is susceptible to interference from working environment factors, and the extraction of spectral characteristics often needs to consider the noise caused by factors such as temperature and moisture interference. The correction of relevant environmental factors thus needs to be considered, which results in a relatively large modeling workload. On the other hand, the material components in the seeds are rich and complex, and interference factors such as the overlap and superposition of near-infrared spectral absorption peaks have always existed, which also affected the quantitative detection ability of infrared spectroscopy to a certain extent.

2.3. Hyperspectral Imaging Technology

As mentioned in Sections 2.1 and 2.2, machine vision technology and spectroscopy technology have their own outstanding advantages. How to realize the combination of the advantages of the two has naturally become the goal pursued by the developers of optical analysis instruments. Thanks to the successful development of high-performance imaging spectrometers, hyperspectral imaging (HSI) has become a reality. Each pixel of HSI can obtain a piece of multi-spectral data, which can reflect the material composition information at a certain position on the target surface; by combining a large number of pixels, seed image information can be reconstructed to achieve the effect of map integration. HSI technology can effectively integrate image technology and spectral technology, increase the dimension of information detection, and thus can analyze seed sample information more comprehensively [22].

Hyperspectral imaging technology was first applied in the field of geographic information remote sensing and gradually applied to the field of seed vigor detection in the past ten years. For example, Baek et al. used hyperspectral imaging technology to quickly and non-destructively distinguish the viability of soybean seeds, and the test accuracy was close to 100% [23]. Ambrose et al. used hyperspectral imaging technology in the 400–2500 nm band to identify two corn seeds with different vigor levels, and the recognition accuracy of the established models reached 97.6% and 95.6%, respectively [24]. Perez et al. used hyperspectral imaging technology to study the seed vigor of the Japanese juniper, and the results showed that hyperspectral imaging technology could be used to effectively predict the vigor of the Japanese juniper [25].

In China, relevant scholars have also carried out a lot of research on the application of hyperspectral imaging technology to the detection of seed vitality. The research targets include crops such as rice, wheat, soybean, and corn. Some progress has also been made in the construction of discriminant models. Pang et al. used hyperspectral imaging technology to study four kinds of corn seeds with different vigor and constructed a seed vigor identification model by using various modeling methods such as a support vector machine and extreme learning machine. The results showed that the identification of different vigor seeds was accurate. The rate reached over 90% [26]. Zhang et al. used hyperspectral imaging technology to detect wheat seed vigor and established a seed vigor detection model using spectral data sets at different positions on the seed surface. The results showed that the established vigor detection model had high accuracy [27]. Li et al. detected the vigor of rice seeds based on hyperspectral imaging technology, and the results showed that the accuracy of classification and identification of seeds with different vigor could reach 94.44% [28].

As analyzed above, compared with machine vision technology and near-infrared spectroscopy technology, hyperspectral imaging technology has the ability to extract the characteristics and internal components of samples at the same time. Therefore, hyperspectral imaging technology has gradually become an active direction for seed vigor detection. At the same time, the analysis of the existing research also found that hyperspectral imaging research generally has a series of problems, such as high spectral data dimensionality, data redundancy and interference information, and complex construction of seed vitality mapping models, which still need to be improved by domestic and foreign scholars.

2.4. Tunable Diode Laser Absorption Spectroscopy

Spectral technology can not only directly detect seed components but also approach the challenge of seed vitality detection by detecting gas components produced during seed metabolism. Respiration is a comprehensive indicator of plant seed metabolism. Most seeds need to complete life activities through respiration, such as cell division and differentiation during early seed germination. Scholars have carried out a lot of research on the correlation between seed respiration and seed vigor, and Woodstock et al. discovered there is a significant positive correlation between seed respiration rate and germination rate [29]. Kalpana et al. [30] found the respiration of seeds showed a positive correlation between strength and vigor, and the above research conclusions become the theoretical basis for judging seed vigor by using the components of seed respiration.

In the presence of oxygen, seed respiration manifests as aerobic respiration, which consumes O_2 and releases CO_2. Under anaerobic conditions, seeds decompose by storing glucose themselves to produce ethanol and CO_2. Therefore, by detecting the change in CO_2 gas concentration produced by seed respiration, the seed vitality can be judged through the information on seed respiration intensity. Scholars at home and abroad have tried to use different gas sensors to detect the CO_2 produced in the process of seed respiration and judge the vigor of seeds. Among various CO_2 sensors, Tunable Diode Laser Absorption Spectroscopy (TDLAS) has better detection accuracy and lower limit than electrochemical or semiconductor sensors and can accurately detect the CO_2 concentration produced by seed respiration, and thus has been favored by researchers. The principle of TDLAS technology exploits a distributed feedback laser (Distributed Feedback Laser, DFB) by adjusting the DFB operating temperature and the size of the driving current and adjusting the wavelength of the laser output laser so that the laser is "selectively" absorbed by the gas to be measured. Macroscopically, the laser intensity is weakened, and the concentration of the gas to be measured can be reversed by using the change of the laser intensity. TDLAS technology has excellent single gas molecule selectivity and can achieve online response; combined with optical absorption cell technology and wavelength modulation technology, its detection limit can reach hundreds of ppt levels.

TDLAS technology has been mainly used in industrial field detection, atmospheric greenhouse gas monitoring, and other fields, while there are relatively few research reports on seed breath detection in the agricultural field. However, in recent years, scholars at home and abroad have strengthened research work in related fields. In China, Jia and others used TDLAS technology [31], and the CO_2 released during the respiration process of corn seeds was detected; combined with the results of the standard seed vitality germination experiment, the results showed that the average value of the correlation coefficient between the CO_2 concentration produced by respiration and the vitality index was 0.975, and the respiration intensity strongly correlated with the vitality index. These experimental results also show that, while limited by the existing device level of TDLAS technology, it is currently difficult to accurately calculate the respiration intensity value of seeds under high dynamic resolution. This affects research progress of the detection of seed respiration intensity, which indicates the need for future research.

Note that current research on seed respiration and seed vigor is mainly qualitative, and a large number of seed respiration detection and seed germination experiments still need to be carried out to quantitatively study the correlation between seed respiration

intensity and seed vigor. On the other hand, although existing studies have shown that the respiration intensity of seeds such as corn, rice, and Chinese fir has a certain positive correlation with their germination rates, there are still some scientific issues in the process that have not been clearly understood, such as in the same genetic strain. Specific issues, such as the quantitative relationship between seed respiration intensity and seed vigor and the quantitative relationship between seed vigor and respiration intensity among different genetic lines, still need to be continuously studied.

2.5. Other Optical Detection Technologies

Fluorescence detection technology, Raman detection technology, photoacoustic spectroscopy detection technology, and other advanced optical detection methods have also been investigated for the development of non-destructive detection of seed vitality.

The application of chlorophyll fluorescence detection technology in the field of seed viability detection has also received attention in recent years. Anhui Institute of Optics and Fine Mechanics, CAS, has studied fluorescence detection techniques for years and developed a system of material composition three-dimensional platform. It has been used in digital agriculture research (Figure 2).

Figure 2. Material composition three-dimensional platform.

When the visible or near-infrared wavelength beam irradiates the seed epidermis, the chlorophyll on the seed coat will release energy in the form of fluorescence. By measuring the fluorescence spectrum released by the chlorophyll on the seed coat, relevant information on the seed vitality level can be obtained. It is confirmed that the smaller the peak signal of the fluorescence spectrum of the seed epidermis is, the lower the chlorophyll content is, and the higher the vigor of the seed is. Because this method is only specific to chlorophyll, it can reduce the influence of other optical radiation noise on the signal, so this method has certain technical advantages. Jalink and others first adopted fluorescence detection technology [32], studied the vigor of cabbage (*Brassica oleracea* L.) seeds, and confirmed that the chlorophyll fluorescence signal was negatively correlated with seed vigor. The analysis of existing research reports also found that the chlorophyll fluorescence detection method still needs to solve the technical difficulties, such as poor seed vigor grading effect due to the influence of chlorophyll content differences in seed samples.

Raman spectroscopy is a scattering spectrum technique that can provide detailed information about molecular vibrations. Since Raman peaks are usually clear and emerge in a relatively narrow band (only a few wavenumbers), it is useful in the feature identification and analysis of organic macromolecules. Raman methods have unique technical advantages and have formed the basis for a standard Raman analysis test instrument. Anhui Institute of Optics and Fine Mechanics, CAS, has carried out research on Raman Spectroscopy detection techniques. A handhold Raman Spectroscopy detection device which can be used for seed Raman spectroscopy detection is shown in Figure 3.

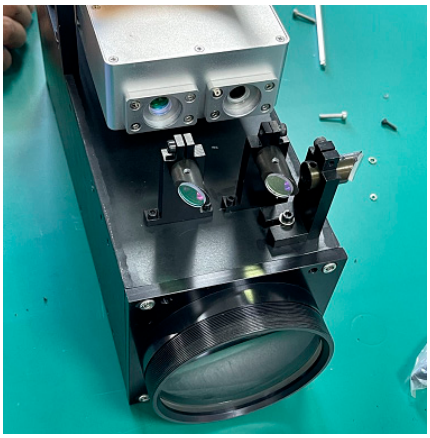


Figure 3. A handhold Raman Spectroscopy detection device.

As Raman spectroscopy shows great potential for rapid analysis, its related work in the field of general agricultural detection is also developing rapidly. More typically, Ambrose et al. used a Raman spectrometer (1700–3200 cm^{-1}) to identify viable and inactive corn seeds [33], with an accuracy rate above 93%. The analysis of existing research reports also found that the Raman scattering signal is weak, which plagues its application in the field of seed viability detection, and related problems still need breakthroughs in key devices.

Photoacoustic spectroscopy technology (Photoacoustic spectrometry technology, or PAS) is a detection technology based on a photoacoustic effect. Its detection principle is that the light source irradiates and heats the target. The ambient air is induced to vibrate weakly and generate sound waves, and the photoacoustic spectrum of the target can be obtained by detecting the weak sound wave signals. Different from traditional absorption spectroscopy, PAS does not directly detect the spectrum based on photon characteristics but detects the periodic heat flow in the non-radiative excitation phenomenon after the sample is illuminated, which changes the laser modulation frequency to adjust the light transmission. This measurement reveals the depth of the seed so that it can detect the material composition of the sample at different depths. Pardo et al. used photoacoustic spectroscopy technology [34] and carried out vegetable seed vitality detection research; their test results show that the light absorption of unaged seeds is higher, and the results show that the scheme of predicting seed vigor by photoacoustic spectroscopy is feasible. In general, photoacoustic spectroscopy is still a relatively new cutting-edge technology, and it is worth investing more resources in research. The current research developments of seed vigor detection procedures based on modern optical methods are assembled in Table 1.

Table 1. Current research developments of seed vigor detection procedures using modern optical methods.

Detection Method	Main Principle	Working Light Band	Core Device	Typical Application
Machine Vision Technology	Optical Imaging of Seeds	Visible light/near-infrared	CCD Imaging device	Perilla, Wheat, etc. [4,5,13]
Infrared Spectroscopy	Infrared Absorption Spectrum of Organic Matter	Near-infrared	Infrared spectrometer	Pine seeds, soybeans, wheat, rice, corn, etc. [8,21,35–37]
Hyperspectral Imaging Technology	Reflection/Absorption Spectrum of Seed Surface	UV to NIR	Broadband light source + imaging spectrometer	Pepper, Rice, Avocado etc. [17,28,38,39]
Turnble Laser Absorption Spectroscopy	Infrared Absorption Spectrum of Organic Matter	Near-infrared to mid-infrared	Tunable laser + lock-in amplifier Tunable laser + lock-in amplifier	Maize [29], pigeonpea [30]

Table 1. *Cont.*

Detection Method	Main Principle	Working Light Band	Core Device	Typical Application
Raman spectrometer	Raman Frequency Shift Spectroscopy of Organic Compounds	UV to Visible Light	Pulse laser + Raman spectrometer	Maize [10]
Photoacoustic spectroscopy	Photoacoustic Effect Excited by Matter Without Radiation	UV/Visible	Laser + weak acoustic wave detector	Lettuce, leaf beets, etc. [33]
Fluorescence spectroscopy	Fluorescent Effect of Organic Matter	UV to visible Light	Fluorescence Spectrometer	Cabbage and other vegetable seeds [12]

3. Development Trends of Key Common Optical Technologies

In theory, optical detection technology has powerful detection capabilities and extremely high measurement accuracy. However, in practice, it is found that the detection capability of the system is often limited by the band response range and wavelength resolution capability of the device; the detection accuracy of the system is often limited by factors such as the optical and electrical noise of the device. On the other hand, even for seeds of the same species, seeds from different planting batches or from different origins will inevitably have differences in chemical composition and phenotypic characteristics, which will affect the spectrum and image of the seeds, creating a large workload for evaluation model construction. The problems faced by optical detection of seed vigor are more complicated. The key to solving this problem lies in whether optical methods can effectively support research progress that combines key devices, modern information technology, and the correlation method of optics and chemometrics.

3.1. Development Status of Key Components

The key components of spectrum or image analysis equipment are concentrated in key units such as the laser light source, the imaging device, and the spectrum analysis device.

Traditional spectral analysis methods usually use a halogen lamp to synthesize a broadband light source, and after irradiating the seed target to be detected, a reflection, scattering, or transmission spectrum is generated. That seed image is received to achieve the purpose of extracting the apparent characteristics and composition information of the seed. Usually, the shell of the seed is thick, and the performance of the halogen light source is limited in terms of intensity, directionality, and stability. This produces a weak interaction between the light field and the seed, which restricts the detection capability of the equipment. The emergence of laser light sources has significantly improved the intensity, directionality, and stability of the light source. However, the common commercial lasers on the market are mostly single-wavelength laser sources with narrow linewidths or lasers with limited wavelength tuning capabilities near the output wavelength (no more than tens of nm), and these lasers are often only able to monitor a single material component within the seed. To meet the simultaneous detection of multiple components of the seed, the wavelength coverage of the light source is usually required to reach 400 nm~2500 nm.

Therefore, whether to increase the bandwidth of the light source while ensuring the emission power of the laser light source and then obtain a spectral signal with sharp features and a high signal-to-noise ratio or an image with high color and signal-to-noise ratio has become the key to improving the ability of optical analysis of seed vitality.

In the past 10 years, there have been breakthroughs with new optical crystal material technology, including the development of nonlinear optical technology and fiber laser technology. As a consequence, the research on supercontinuum light sources in the world has achieved rapid development, especially the infrared supercontinuum laser source, which is the key to spectral analysis. One of the most exciting new technologies to emerge in the field is the so-called Supercontinuum light source technology, which uses semiconductor lasers to pump photonic crystal fibers. These devices can generate broadband spectra covering

400~2500 nm, and the highest spectral power density can reach milliwatts/nanometers [40]. At the same time, these have the spectral range of traditional halogen lamps and monochromatic lasers. Directionality and high energy characteristics, strong light source penetration, and wide spectrum range, while satisfying the simultaneous detection of multiple material components, can also extract data from deep into the seeds and can obtain the full-band spectrum and images of the changes in the internal components of the seeds as much as possible. This kind of information can further improve the performance of optical detection of seed vitality.

While supercontinuum laser technology has made major breakthroughs in the past ten years, high-tuning precision DFB lasers for the detection of seed exhalation breath intensity [41] and quantum cascade lasers for the detection of macromolecular substances in the mid- and far-infrared bands have both achieved substantial progress. The development of this new type of light source has also made new breakthroughs in the optical detection of seed vigor and laid the foundation for key devices.

The spectral analysis component is a necessary foundational component of modern optical analysis methods. Its main function is to perform phenotype discrimination and component analysis on substances by obtaining the spectrum after the interaction between light and substances. The main technical route of common spectral analysis components is to use discrete optical devices such as gratings and prisms to separate light fields of different frequencies in space, and then use area arrays of CCD imaging devices for measurement, and finally achieve the purpose of spectral imaging measurement. The working mechanism of discrete device combination often leads to a series of problems, such as the high cost of spectral analysis components, poor integration ability, and low long-term working stability. This means the measurement requirements of high precision, high consistency, and on-chip integration of equipment are difficult to achieve. Therefore, the development of new spectral analysis components to achieve wider detection bands, finer wavelength resolution, and miniaturized on-chip integration capabilities has become an urgent breakthrough direction.

3.2. Modern Information Technology and Spectral or Image Preprocessing Algorithms

Affected by their own metabolism and storage environment, after the seed reaches its peak vitality, its vitality gradually decreases, and the change process of the vitality in a short period of time is subtle. The key parameters are usually only evident in the content of some key enzymes, hormones, and other vital substances have changed, and such changes are reflected in differences in spectral or phenotypic image information of seed vigor, often very subtle (e.g., often a few percent change in spectral intensity). On the other hand, restricted by the technical level of existing devices, the information richness of spectral and image data that can be obtained at present is limited. With the complexity of these two factors, the development of seed viability evaluation equipment has remained a significant challenge.

When combined with modern information technology, carrying out spectrum or image preprocessing research is an effective means to alleviate the lack of capability of the above-mentioned detection devices. However, the current conventional spectral preprocessing methods, such as the derivative method to extract information bursts and other algorithms, lack the ability to effectively amplify the differences in optical characteristics of different vigor seeds. Therefore, in order to achieve accurate identification of seed vigor, it is necessary to propose a more reasonable spectral preprocessing algorithm for small differences in optical characteristics and to amplify and preprocess the differences in spectral or image information so as to effectively reflect the small changes in seed vigor. The above spectral or image-preprocessing technologies often rely on the development of modern information technology.

3.3. Correlative Methods of Optics and Chemometrics

Optical detection technology is an indirect analysis method, so it is necessary to establish an effective "spectrum-vitality" evaluation model as a bridge between the spectrum and seed vigor information. However, the existing optical detection of seed vigor often directly correlates the measurement results of optical information with standard seed germination experiments. The "spectrum-vitality" evaluation model established by these methods fails to expose the required "spectrum-omics-seed" insight. The complete research chain of vitality makes the mechanism of the model unclear, and the accuracy is limited. However, the chemometric analysis methods to further obtain the material components in the process of seed vigor changes can be combined with an exploration of more refined differences in material components of different vigor seeds. This coupling will provide a material basis for determining seed vigor so that a more quantitative evaluation model can be established. It can be seen that strengthening the research on the correlation method between modern optical technology and chemometrics is also a key potential breakthrough to improve the accuracy of optical detection technology [42].

4. Conclusions and Perspective Discussion

Due to the wide variety of seed types and traits, traditional seed viability detection methods have difficulty meeting the needs of rapid and non-destructive detection of seed vigor. Optical technology can provide different vigor detection methods in a targeted manner according to different vigor expressions of seeds. For example, for seeds with a strong correlation between the vitality index and the degree of browning of the seeds, machine vision technology provides the basis for differential assessment; for seeds whose components have undergone drastic changes after deterioration, near-infrared spectroscopy or hyperspectral technology can be used to judge the vigor of the seeds; and for seeds that need to be detected in batches, TDLAS technology can be used to detect their respiration intensity and evaluate the vitality of the seed population, for some seeds that require extraction of parameters that are revealed only by looking into the depth of seeds. This provides the basis for detecting seed vitality components, and photoacoustic spectroscopy is an effective method to detect this component of seed vitality.

With the development of key devices and detection methods, optical technology has received extensive attention in the field of assessing seed viability. However, the analysis of current research reports is still based on laboratory research. There is still a lack of commercial seed viability optical detection equipment in the market. The relevant problems are mainly reflected in the following three aspects: (1) Although the non-destructive detection method of seed vigor has the advantage of being fast and non-destructive, it is restricted by factors such as the development level of key photoelectric detection devices and many problems still need to be overcome in the current research stage. For example, machine vision technology can judge the vigor of seeds through the information of seed phenotype, but not all kinds of seed vigor are related to the physical information of seed phenotype; near-infrared spectroscopy and hyperspectral technology can obtain the internal composition spectrum information of seeds, therefore, spectral components contain a lot of redundant information and noise, and breakthroughs in image noise reduction and fast image processing technology are urgently needed. (2) The current optical detection methods for seed vigor require a standard germination test for each batch or variety of seeds, followed by subsequent correlation analysis with the data obtained through optical technology detection to obtain a vitality evaluation model. In the process of model construction, the focus is on the mapping relationship between optical features and the apparent features of seed vigor, but there is a lack of an effective correlation process between chemometrics and optical features. In addition, there is relatively little research on the mechanism of seed vigor, and it is impossible to establish an effective relationship mapping. The quantitative relationship between optical features and viability grades affects the validity of evaluation models. (3) The cost of laser light sources and precision optical devices involved in optical detection methods is generally high, especially the

high price of key devices for hyperspectral imaging detection, which limits its commercial application prospects.

In view of the above problems, the research direction that needs to be strengthened for the modern optical detection of seed vigor includes: (1) Strengthen the theoretical research on the correlation between optical characteristics and seed components through in-depth research on the optical characteristics, omics, and vigor of seeds, obtaining a clear mechanism related to optical properties and seed vigor, and finally establish a refined and more adaptable seed vigor detection model. (2) Due to the different advantages of optical detection technologies with different principles, it is often difficult to fully obtain seed vigor information with a single technology, which affects detection accuracy. It is undoubtedly one of the future research directions to explore the fusion of multi-sensing technologies, realize multi-method detection and multi-parameter combination, and improve the effectiveness of crop seed vigor detection. On the other hand, although hyperspectral imaging and other detection methods are powerful, they are limited by the cost of devices, and their commercial application is currently underdeveloped. The combination of low-cost infrared absorption spectroscopy technology and machine vision technology can also increase the dimension of information and achieve the purpose of further improving the accuracy of seed vitality detection, thus having better commercial application prospects. (3) Strengthen engineering application research and promote the early industrialization of detection technology research. Internationally, giants such as Monsanto Company of the United States have carried out application attempts of online seed viability detection devices, which have initially demonstrated the commercial value of the equipment.

In order to further promote the application of optical seed viability detection technology, the field of engineering research needs to comprehensively consider the three elements of performance, cost, and reliability and select the technology with high commercial feasibility for engineering. On the other hand, with the development of electronic information and automation technology, it is necessary to strengthen the combination of detection technology and intelligent classification devices, develop integrated detection and sorting equipment, and promote the commercial application of related equipment as soon as possible.

Author Contributions: Conceptualization: J.Z. and G.B.; software, G.B.; validation, C.X., W.F. and A.X. resources, G.B. and J.Z.; data curation, C.X. and G.B.; writing—original draft preparation, J.Z. and G.B.; writing—review and editing, J.Z., M.Z. and R.G.; visualization, G.B.; supervision, J.Z. and G.B.; project administration, C.X.; funding acquisition, A.X., G.B., C.X. and J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: Jilin Agricultural University high-level researcher grant (JLAUHLRG20102006).

Data Availability Statement: Not applicable.

Acknowledgments: Authors would like to thank Jianyu Lu for the reference checking for this manuscript.

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

References

1. Seo, Y.-W.; Ahn, C.K.; Lee, H.; Park, E.; Mo, C.; Cho, B.-K. Non-Destructive Sorting Techniques for Viable Pepper (*Capsicum annuum* L.) Seeds Using Fourier Transform Near-Infrared and Raman Spectroscopy. *J. Biosyst. Eng.* **2016**, *41*, 51–59. [CrossRef]
2. Meike, D.; Chen, W.J.; Beachy, R.N. Study on rice fat content based on near infrared reflectance spectroscopy. *GEain Oil Process.* **2007**, *2*, 62–64.
3. Ao, Y.; Wang, J.; Liu, M.; Wang, J.; Knops, J.M.H.; Mu, C. Larger Seed Size Shows Less Germination and Seedling Growth Decline Caused by Seed Ageing under Na_2CO_3 Stress in *Leymus chinensis*. *Agron. J.* **2019**, *111*, 2326–2331. [CrossRef]
4. Wang, Y.H. *Seed Quality Tracing System of China Research*; Chinese Academy of Agricultural Sciences: Beijing, China, 2016. (In Chinese)
5. Filho, J.M. Seed vigor testing: An overview of the past, present and future perspective. *Sci. Agric.* **2015**, *72*, 363–374. [CrossRef]
6. International Seed Testing Association (ISTA). *Seed Vigor Testing*; International Rules for Seed Testing: Zurich, Switzerland, 2014.

7. Association of Official Seed Analysts (AOSA). *Seed Vigor Testing Handbook*; Association of Official Seed Analysts: Springfield, IL, USA, 1983.
8. Pereira, D.F.; Bugatti, P.H.; Lopes, F.M.; Souza, A.L.; Saito, P.T. Contributing to agriculture by using soybean seed data from the tetrazolium test. *Data Brief* **2019**, *23*, 103652. [CrossRef] [PubMed]
9. Fatonah, K.; Suliansyah, I.; Rozen, N. Electrical conductivity for seed vigor test in sorghum (Sorghum bicolor). *Cell Biol. Dev.* **2017**, *1*, 6–12. [CrossRef]
10. Bi, C.; Hu, N.; Zou, Y.; Zhang, S.; Xu, S.; Yu, H. Development of Deep Learning Methodology for Maize Seed Variety Recognition Based on Improved Swin Transformer. *Agronomy* **2022**, *12*, 1843. [CrossRef]
11. Ma, L.; Yu, Q.; Yu, H.; Zhang, J. Maize Leaf Disease Identification Based on YOLOv5n Algorithm Incorporating Attention Mechanism. *Agronomy* **2023**, *13*, 521. [CrossRef]
12. Xia, Y.; Xu, Y.; Li, J.; Zhang, C.; Fan, S. Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artif. Intell. Agric.* **2019**, *1*, 35–47. [CrossRef]
13. Yuan, J.; Zheng, W.; Qi, H.; Gao, L.; Hu, X.; Zhao, G.; Shi, J.; Jia, L. Progress in research of optical non-destructive test technology for seed vigor. *Crops* **2020**, *5*, 9–16. [CrossRef]
14. McCormac, A.C.; Keefe, P.D.; Draper, S.R. Automated vigor testing of field vegetables using image analysis. *Seed Sci. Technol.* **1990**, *1*, 103–112.
15. Li, Z.; Liao, T.Q.; Feng, Q.C.; Zhang, D.Y.; Wang, X. Study on vegetable seed vigor index detection algorithm and system realization based on machine vision. *Acta Agric. Zhejiang* **2015**, *27*, 2218–2224. (In Chinese)
16. Men, S.; Yan, L.; Liu, J.; Qian, H.; Luo, Q. A Classification Method for Seed Viability Assessment with Infrared Thermography. *Sensors* **2017**, *17*, 845. [CrossRef] [PubMed]
17. Mo, C.; Kim, G.; Lee, K.; Kim, M.S.; Cho, B.-K.; Lim, J.; Kang, S. Non-Destructive Quality Evaluation of Pepper (*Capsicum annuum* L.) Seeds Using LED-Induced Hyperspectral Reflectance Imaging. *Sensors* **2014**, *14*, 7489–7504. [CrossRef]
18. Li, H.; Liang, Y.; Xu, Q.; Cao, D. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Anal. Chim. Acta* **2009**, *648*, 77–84. [CrossRef]
19. Jin, W.L.; Cao, N.L.; Zhu, M.D.; Chen, W.; Zhang, P.G.; Zhao, Q.L.; Liang, J.Q.; Yu, Y.H.; Lv, J.G.; Kan, R.F. Nondestructive GEading test of rice seed activity using near infrared super-continuum laser spectrum. *Chin. Opt.* **2020**, *13*, 1032–1043. (In Chinese) [CrossRef]
20. Fan, Y.; Ma, S.; Wu, T. Individual wheat kernels vigor assessment based on NIR spectroscopy coupled with machine learning methodologies. *Infrared Phys. Technol.* **2020**, *105*, 103213. [CrossRef]
21. Belin, E.; Douarre, C.; Gillard, N.; Franconi, F.; Rojas-Varela, J.; Chapeau-Blondeau, F.; Demilly, D.; Adrien, J.; Maire, E.; Rousseau, D. Evaluation of 3D/2D Imaging and Image Processing Techniques for the Monitoring of Seed Imbibition. *J. Imaging* **2018**, *4*, 83. [CrossRef]
22. Yasmin, J.; Ahmed, M.R.; Lohumi, S.; Wakholi, C.; Kim, M.S.; Cho, B.-K. Classification Method for Viability Screening of Naturally Aged Watermelon Seeds Using FT-NIR Spectroscopy. *Sensors* **2019**, *19*, 1190. [CrossRef]
23. Baek, I.; Kusumaningrum, D.; Kandpal, L.M.; Lohumi, S.; Mo, C.; Kim, M.S.; Cho, B.-K. Rapid Measurement of Soybean Seed Viability Using Kernel-Based Multispectral Image Analysis. *Sensors* **2019**, *19*, 271. [CrossRef]
24. Ambrose, A.; Kandpal, L.M.; Kim, M.S.; Lee, W.-H.; Cho, B.-K. High speed measurement of corn seed viability using hyperspectral imaging. *Infrared Phys. Technol.* **2016**, *75*, 173–179. [CrossRef]
25. Lee, H.-S.; Jeon, Y.-A.; Lee, Y.-Y.; Lee, G.-A.; Raveendar, S.; Ma, K.H. Large-Scale Screening of Intact Tomato Seeds for Viability Using Near Infrared Reflectance Spectroscopy (NIRS). *Sustainability* **2017**, *9*, 618. [CrossRef]
26. Pang, L.; Men, S.; Yan, L.; Xiao, J. Rapid Vitality Estimation and Prediction of Corn Seeds Based on Spectra and Images Using Deep Learning and Hyperspectral Imaging Techniques. *IEEE Access* **2020**, *8*, 123026–123036. [CrossRef]
27. Zhang, T.; Wei, W.; Zhao, B.; Wang, R.; Li, M.; Yang, L.; Wang, J.; Sun, Q. A Reliable Methodology for Determining Seed Viability by Using Hyperspectral Data from Two Sides of Wheat Seeds. *Sensors* **2018**, *18*, 813. [CrossRef]
28. Li, M.L. *Research on Rice Seed Vigor Detection Technology Based on Hyperspectral*; Zhejiang A&F University: Hangzhou, China, 2014.
29. Woodstock, L.W.; Grabe, D.F. Relationships Between Seed Respiration During Imbibition and Subsequent Seedling Growth in *Zea mays* L. *Plant Physiol.* **1967**, *42*, 1071–1076. [CrossRef]
30. Kalpana, R.; Madhava, R.K. Ultrastructural and physiological changes associated with loss of seed viability in pigeonpea. *Indian J. Plant Physiol.* **1993**, *36*, 86.
31. Jia, L.Q.; Qi, H.N.; Hu, W.J.; Zhao, G.W.; Kan, R.F. CO_2 concentration detection system for seed respiration. *Opt. Precis. Eng.* **2019**, *27*, 1397–1404. [CrossRef]
32. Jalink, H.; van der Schoor, R.; Frandas, A.; van Pijlen, J.G.; Bino, R.J. Chlorophyll fluorescence of *Brassica oleracea* seeds as a non-destructive marker for seed maturity and seed performance. *Seed Sci. Res.* **1998**, *8*, 437–443. [CrossRef]
33. Ambrose, A.; Lohumi, S.; Lee, W.-H.; Cho, B.K. Comparative nondestructive measurement of corn seed viability using Fourier transform near-infrared (FT-NIR) and Raman spectroscopy. *Sens. Actuators B Chem.* **2016**, *224*, 500–506. [CrossRef]
34. Pardo, G.P.; Pacheco, A.D.; Tomás, S.A.; Orea, A.C.; Aguilar, C.H. Characterization of Aged Lettuce and Chard Seeds by Photothermal Techniques. *Int. J. Thermophys.* **2018**, *39*, 118. [CrossRef]
35. Tigabu, M.; Odén, P.C. Discrimination of viable and empty seeds of Pinus patula Schiede & Deppe with near-infrared spectroscopy. *New For.* **2003**, *25*, 163–176. [CrossRef]

36. Al-Amery, M.; Geneve, R.L.; Sanches, M.F.; Armstrong, P.R.; Maghirang, E.B.; Lee, C.; Vieira, R.D.; Hildebrand, D.F. Near-infrared spectroscopy used to predict soybean seed germination and vigour. *Seed Sci. Res.* **2018**, *28*, 245–252. [CrossRef]
37. Larios, G.; Nicolodelli, G.; Ribeiro, M.; Canassa, T.; Reis, A.R.; Oliveira, S.L.; Alves, C.Z.; Marangoni, B.S.; Cena, C. Soybean seed vigor discrimination by using infrared spectroscopy and machine learning algorithms. *Anal. Methods* **2020**, *12*, 4303–4309. [CrossRef] [PubMed]
38. Díaz, J.J.V.; Aldana, A.P.S.; Zuluaga, D.V.R. Prediction of dry matter content of recently harvested 'Hass' avocado fruits using hyperspectral imaging. *J. Sci. Food Agric.* **2020**, *101*, 897–906. [CrossRef]
39. Mukasa, P.; Wakholi, C.; Mohammad, A.F.; Park, E.; Lee, J.; Suh, H.K.; Mo, C.; Lee, H.; Baek, I.; Kim, M.S.; et al. Determination of the viability of retinispora (*Hinoki cypress*) seeds using shortwave infrared hyperspectral imaging spectroscopy. *J. Near Infrared Spectrosc.* **2020**, *28*, 70–80. [CrossRef]
40. Ranka, J.K.; Windeler, R.S.; Stentz, A.J. Visible continuum generation in air–silica microstructure optical fibers with anomalous dispersion at 800 nm. *Opt. Lett.* **2000**, *25*, 25–27. [CrossRef]
41. Sun, Z.; Su, Z.; Xiao, R.; Wang, Y.; Liu, K.; Wang, F.; Liu, Y.; Fang, T.; Chiu, Y.-J.; Chen, X. Tunable Laser via High-Density Integration of DFB Lasers With High Precision Wavelength Spacings. *IEEE Photonics Technol. Lett.* **2022**, *34*, 467–470. [CrossRef]
42. Wang, Z.; Huang, W.; Tian, X.; Long, Y.; Li, L.; Fan, S. Rapid and Non-destructive Classification of New and Aged Maize Seeds Using Hyperspectral Image and Chemometric Methods. *Front. Plant Sci.* **2022**, *13*, 849495. [CrossRef]

agronomy

MDPI

Communication

Hybrid CNN-SVM Classifier Approaches to Process Semi-Structured Data in Sugarcane Yield Forecasting Production

Debnath Bhattacharyya [1], Eali Stephen Neal Joshua [2], N. Thirupathi Rao [3] and Tai-hoon Kim [4,*]

[1] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur 522302, Andhra-Pradesh, India
[2] Department of Computer Science & Engineering, Gitam (Deemed to be University), Gandhi Nagar, Rushikonda, Visakhapatnam 530045, Andhra-Pradesh, India
[3] Department of Computer Science & Engineering, Vignan's Institute of Information Technology (A), Visakhapatnam 530049, Andhra-Pradesh, India
[4] School of Electrical and Computer Engineering, Yeosu Campus, Chonnam National University, 50, Daehak-ro, Yeosu-si 59626, Jeollanam-do, Republic of Korea
* Correspondence: taihoonn@chonnam.ac.kr; Tel.: +82-10-8592-4900

Abstract: Information communication technology (ICT) breakthroughs have boosted global social and economic progress. Most rural Indians rely on agriculture for income. The growing population requires modern agricultural practices. ICT is crucial for educating farmers on how to be environmentally friendly. It helps them create more food by solving a variety of challenges. India's sugarcane crop is popular and lucrative. Long-term crops that require water do not need specific soil. They need water; the ground should always have adequate water due to the link between cane growth and evaporation. This research focuses on forecasting soil moisture and classifying sugarcane output; sugarcane has so many applications that it must be categorized. This research examines these claims: The first phase model predicts soil moisture using two-level ensemble classifiers. Secondly, to boost performance, the proposed ensemble model integrates the Gaussian probabilistic method (GPM), the convolutional neural network (CNN), and support vector machines (SVM). The suggested approach aims to correctly anticipate future soil moisture measurements affecting crop growth and cultivation. The proposed model is 89.53% more accurate than conventional neural network classifiers. The recommended models' outcomes will assist farmers and agricultural authorities in boosting production.

Keywords: Information communication technology; agriculture; ensemble learning; Gaussian probabilistic method function; convolutional neural network; support vector machines

Citation: Bhattacharyya, D.; Joshua, E.S.N.; Rao, N.T.; Kim, T.-h. Hybrid CNN-SVM Classifier Approaches to Process Semi-Structured Data in Sugarcane Yield Forecasting Production. *Agronomy* **2023**, *13*, 1169. https://doi.org/10.3390/agronomy13041169

Academic Editors: Jian Zhang, Randy G. Goebel and Zhihai Wu

Received: 4 March 2023
Revised: 10 April 2023
Accepted: 12 April 2023
Published: 20 April 2023

1. Introduction

Precision agriculture uses cutting-edge technology to improve crop yield and quality while meeting growing food demand [1]. Farmers may optimize the efficacy of their seeds, water, fertilizers, and other inputs by employing advanced farming techniques. The goal is to maximize financial benefit while reducing expenditures and the effects on the natural environment by using the correct parameters at the right time and place.

Farmers must monitor soil moisture. This entails obtaining a more precise image of the soil's needs and available water. An amount of 70% of the world's water is utilized for agriculture, 20% is utilized for industry, and 10% is utilized for residences, according to UN figures (http://www.worldometers.info/) (accessed on 10 February 2022) [2]. To monitor and plan irrigation, you must know the soil's water content. Sugarcane is a popular crop in India. It is the critical raw material for the nation's second-largest agricultural sector, behind only the textile industry.

Additionally, it serves as the foundation for producing every significant sweetener in the country. Long-term growth does not need a specific soil type, but it requires much water. The soil must have enough water during the growing season, and cane growth depends on how much cane sheds. In addition to categorizing sugarcane yield, the research aims to anticipate soil moisture.

Hybrid Model for Better Sugarcane Yield Production

Agriculture has since incorporated soft computing to manage crops, anticipate agricultural yields, and monitor environmental conditions [3]. In India, sugarcane is one of the most important crops [4].

Sugarcane businesses provide significant benefits to rural areas in terms of the extra cash they generate and the job possibilities they make available. The proposed study will investigate how different soft computing technologies may be used to improve the categorization and forecasting of agricultural data to find better solutions to existing problems.

This section offers a condensed summary of the critical issues and contributions that this study has made.

- In order to provide predictions about the amount of moisture present in the soil, a mixed model was used. In order to do this, the most advantageous aspects of the GPM, CNN, and SVM algorithms may be merged.
- CNN is used to sort the data on the production of sugarcane.

According to a thorough literature survey, CNN analysis is helpful for crop management and yield research in precision agriculture. They improve over time by uncovering new patterns, not in the training datasets.

Ensemble learning's power comes from its capacity to train several algorithms and combine their results to obtain more accurate conclusions. The hybrid model, an ensembling technique, produces accurate predictions. To make hybrid solutions for real-world problems, you must combine at least two methods to compensate for each other's flaws. Because agriculture generates so much data, precision agriculture applications require various data processing technologies.

Academics have created algorithms for agricultural data applications. These algorithms will help farmers and agronomists manage crops and adopt agricultural strategies to guarantee a continual crop supply. This study uses neural networks, GPM, and SVM algorithms to improve agricultural data prediction and categorization.

The primary objective of this research is to enhance the precision of agricultural yield in sugarcane production in agriculture.

- Sugarcane production has been the most prominent crop in Andhra Pradesh, India.
- Certain soil moisture methods may be deployed for better crop yield.
- Developing forecasting models for soil moisture data sets using convolutional neural network ensembles was performed. To do this, the attributes of SVM and Gaussian function were combined. Hybrid model (CNN + SVM + Gaussian) approaches are being used to train the model's parameters to increase the accuracy of a model that detects sugarcane yields.

2. Related Works

Precision agriculture and smart farming need rapid disease diagnosis [5]. Farmers may benefit from using plant pictures and existing machine-learning algorithms to detect crop illnesses early. Modern diagnostics may help rural farmers save their harvests. Yet, current technologies cannot accurately identify diseases. Academics and software developers are interested in how to pick a machine-learning model and scale it down to run efficiently on a device at the network's perimeter. This study yields a model for the Plant-Village and Plant-Doc datasets [6]. The technique relies on deep learning model performance assessment research. The analysis found that the proposed model is similar to the most

sophisticated option on the market. Quantization ensures that the reduction in model size to suit an edge device will not affect system performance.

The N nutrition index helps farmers determine whether their crops have enough nitrogen (N) for safe harvesting. Nitrogen utilization efficiency (NUE) [7] will rise, reducing environmental damage. However, active canopy sensor data cannot accurately detect the crop's N condition or topdressing N quantity. Reflectance data from several active crop sensors is unpredictable because many environmental and management elements vary by location. To ensure food security and long-term growth, further study is needed to develop ways for monitoring and prescribing N levels throughout the growing season in various contexts.

The average phosphorus surplus in the Andhra-Pradesh and Telangana regions was 0.8 kg per acre per year. However, this varied by location. When inputs were lowered in regions with big phosphorus surpluses and easy-to-access soil, the region's yearly phosphorus budget revealed plenty of potential for improvement [8,9].

Agriculture is important to India's economy [10]. Most farmers know little about soil preservation. Grow the right crops and use the right fertilizers to maximise land production. Modern agriculture depends on technology. More productive and efficient agriculture practices are replacing older ones. A rising number of individuals are concerned about fertilizer consumption. Farmers may use several machine learning algorithms to determine fertilizer quantities. Fertilizer consumption rates vary according to crop-specific fertilizer needs. Machine learning (ML) researchers aim to create computer programs that can explore massive amounts of data for patterns and use those patterns to anticipate fresh data. Machine learning (ML) methods assist in solving prediction and classification issues. Machine learning, utilized in many agricultural contexts, helps farmers overcome their biggest challenges. This experiment tested the support vector machine (SVM), decision tree (DT), and CNN for soil nutritional analysis. The CNN algorithm is 4% more accurate than other algorithms [11,12].

New technologies affect farming, textiles, vehicles, engineering, and education. AI has shaped educational practices throughout time. This study [13] examines how current AI advances may influence new educational methods. The research created a useful random forest algorithm teaching methods learning companion. The proposed random forest technique performed inconsistently during training, but, thereafter, it performed better. The recommended strategy outperformed the support vector machine, a hesitant fuzzy set, and the overall expert score in the 30–35 range of the final weighted evaluation. 7%, 6%, and 17% were the results of this strategy.

Villupuram is a rural Tamil Nadu city. This research will focus on future climate and agriculture harvest advice [14]. The seasonality of agricultural production is another aspect that must be considered. This work computed 1990 agricultural production using Villupuram district data from 1990–1991 and 2019–2020. Moreover, agricultural production patterns will be examined. IMD forecasts weather and environment. Seasons alter the maximum temperature and lowest rainfall. SVM is the most reliable machine-learning method for seasonal output forecasting. We will also evaluate the SVM machine algorithm's efficiency and accuracy. A support vector machine (SVM) [15] showed that the seasonal temperature scale will accurately predict the global average temperature in 20 years. Researchers used agronomic practices to show when and how to grow different crops throughout this research.

Humans are endangering Madagascar's unique vegetation and fauna. This article presents data and analysis on the most prevalent and vital threats, collections, and programs, as well as protected area extent and breadth, to evaluate Madagascar's terrestrial and freshwater biodiversity's conservation status. An amount of 10.4% of Madagascar is preserved as terrestrial parks and reserves. Ninety-seven percent of freshwater fish, amphibians, reptiles, birds, and mammals, and 67% of native plants, live inside its boundaries. Rarer species exist (at least 79.6 percent of endangered plant species and 97.3 percent of endangered animal species reside in at least one protected location). Overexploitation

of biological resources and unsustainable farming are the most significant risks to biodiversity, according to Bayesian neural network plant research and International Union for Conservation of Nature (IUCN) Red List evaluations. We concentrate on five action levels to ensure that conservation and ecological restoration goals, methods, and activities address complex underlying and interrelated causes and benefit Madagascar's biodiversity and people. Due to cassava's weekly price increases, it is hard to anticipate agricultural expenditures and similar group inputs. Machine learning may be used to calculate cassava costs in the future [16,17].

The Office of Agricultural Economics under the Ministry of Agriculture and Cooperatives studied cassava prices from January 2005 to February 2022, which represents strategy development using the data. The support vector machine approach, incorporating potato and garlic prices, had the lowest root mean squared error (RMSE) [18] of 0.10 compared to the previous methodology that utilized the same information. Comparing the support vector machine to the previous approach revealed this. The recommended technique is 0.61% more successful, despite its 3.45% MAPE. The penultimate phase of price clustering, K-mean analysis, gave a December peak pricing period of 14.08%. The study would help the agricultural firm maximize profits by guiding its planting strategies.

It is essential when formulating food policy and ensuring there is enough food for everyone, to accurately predict what the winter wheat yields will be in each region ahead of time. This is especially important when considering the growing population and the changing environment. Artificial intelligence (AI) and large amounts of data in agriculture are the foundations of intelligent agriculture, which aims to help the agricultural industry find effective and cheap solutions to the challenges it faces [19]. Profound learning-based crop production predictions have emerged in recent years as an essential tool for controlling agricultural operations' output. In this study, we offer a Bayesian optimization-based strategy for predicting winter wheat production by fusing data from several sources using a long-term and short-term memory model. Our method can provide more accurate results than previous methods (BO-LSTM) [20]. Using multi-source data as input variables, analysis of BO-LSTM, support vector machine (SVM), and most minor absolute shrinkage and selection operator (Lasso) give the capability to make predictions. The results showed that Bayesian optimization might be used to adjust deep learning hyperparameters successfully.

The recent advancements in information and communication technology have impacted many aspects of the global economy. The rise of artificial intelligence (AI) and data science, in addition to the ubiquitous availability of digital technology, are likely to be credited as the primary drivers behind the arrival of digital agriculture. Farmers now have access to new farming practices that may increase productivity without affecting the ecological health of the land they work on as a residue of total agricultural improvement. The most recent developments in digital technology and data science have made it possible for agronomists, farmers, and other experts to acquire and analyze large amounts of agricultural data. This allows them to gain insight into farming operations and make better decisions. [21] examines how data mining methods may be used in contemporary agricultural technologies. [22] offers a high-level overview of agricultural yield management, focusing on crop production and monitoring as the two most important aspects to discuss. When the data mining approaches for monitoring agricultural production have been sorted into their basic categories, several studies that focus on analytics are discussed and evaluated. After that, a comprehensive analysis and discussion of big data's influence on agriculture follows.

The algorithm for predicting time series in decision support systems has seen substantial advancements in recent years. Automating and improving prediction models by using machine learning (ML) methods to analyze time series may be possible. This would come about when managing enormous amounts of data becomes more accessible. The most important thing this research brings is a hybrid model that performs better than previous forecasting models. To do this, its innovative take on the traditional Kalman filter is performed by incorporating machine learning techniques, such as support vector

regression (SVR) [23] and nonlinear autoregressive (NAR) [24] neural networks into the mix. The proposed hybrid model addresses the problem of high error fluctuation in time series data convergence by combining an updated Kalman filter strategy with a machine learning technique as a correction factor to estimate model error. The ML technique serves as a correction factor to estimate model error. The findings indicate that, compared to traditional and alternative Kalman filter models [25], our hybrid models have a goodness of fit of more than 0.95 and much lower root mean square and absolute mean errors. The method's applicability to various settings was shown further in two different ways.

Gap Analysis

The thorough literature review [5–25] indicates the need to improve crop yield production. The authors offer a heavyweight model that calls for extended training and generates results that vary, depending on the computer. No matter how complicated the machine learning model is, domain experts cannot understand how it works. They have a hard time maintaining their consistency. Recent research has shown that all models, particularly machine learning models, can be accurate. However, it is pointless to utilize them if the pitching expert disagrees with the model. To clarify the decision-making process, to increase farmers' trust, and to normalize attitudes towards the concept, all of which would increase the likelihood of its eventual adoption, the goal is to use the lightweight process model and methods, such as integrating the SVM, as well as the radial basis function, with the ensemble learning techniques. This will be accomplished by using the proposed classifier.

The soil's moisture content is determined by its current water content. This factor affects air temperature, precipitation, and other weather aspects. Agriculture, precision farming, and drought monitoring benefit from their usage. When deciding when and how much to irrigate, growers must monitor soil moisture [26]. This increases output, improves product quality, and reduces disease and irrigation costs. Most Indians make their living via agriculture. Population growth will continue as a significator for water conservation for human consumption, with the rest used for agricultural irrigation. You must understand the crop's water needs to meet this growing need. In locations with little water, monitoring soil moisture for effective irrigation management is increasingly important. Soil moisture explains how water and energy are moved from the ground surface to the atmosphere.

Real-time soil moisture data is needed to predict the weather, monitor floods, identify droughts early, and calculate crop growth. Forecasting is constructing a model to predict the future based on current knowledge. The capacity to accurately anticipate soil moisture affects crop growth, cultivation, and soil temperature. The challenge is creating a system that can continuously evaluate soil moisture levels. Researchers studied several statistical techniques, algorithms, and tools to construct ensemble models for weather forecasting and assessing soil moisture. The created ensembles' performance is estimated using a statistical method. A neural network-based annealing method was used to anticipate soil moisture. The suggested technique outperforms BPN and GPM networks [27]. A network-based model that can predict tomorrow's ground temperature. The model received ground and air temperatures. A good prediction helps navigate the unpredictable future. Random forest, bagging, boosting, voting, stacking, and boosting are popular data-gathering methods.

To precisely measure moisture in the soil, we must also have excellent and reliable sensors, such as quartz sensors that work on the principle of open capacitive sensitive elements pushed to different depths into the ground, such as those described in ref: [28,29].

3. Materials and Methods

3.1. The Proposed Ensemble Model for Soil Moisture (Hybrid Model)

Introduction to the Model Deployment

A network of wireless soil moisture sensors is placed to assess how wet the top layer of soil is. The base station stored the sensor's time series data. The soil moisture data from the base station are analyzed to fix the problem.

The study presented aims to enhance soil moisture forecasting. The suggested ensemble model comprises GPM, CNN, and SVM individual learners. The ZeroR classifier provides predictions and is a meta-learner. The ZeroR classifier is the easiest to use and can predict the most frequent category. It helps to compare classifiers to a standard.

3.2. Gaussian Probabilistic Method for CNN-SVM

The Gaussian probabilistic functions were used with the CNN. The network's output is a linear blend of the inputs' GPM functions [30] and neuronal characteristics. Because it uses supervised learning, it can represent nonlinear data and help find the proper application. Input, concealed, and output are GPM network layers (Figure 1). GPM allows faster learning, better approximation, and more accessible network design than other CNNs. We will utilize least-squares mean-squared to determine optimum weights.

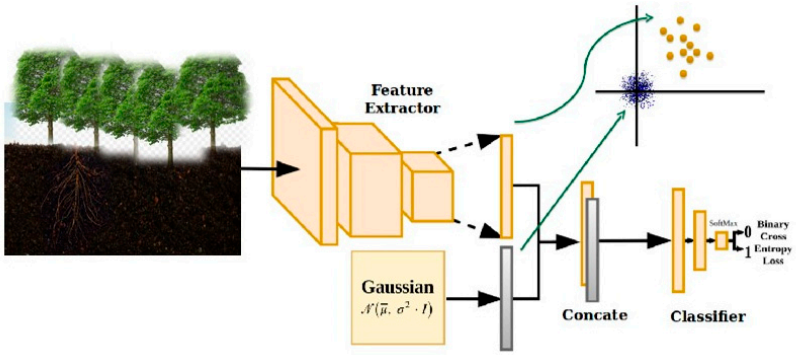


Figure 1. Gaussian probabilistic for the Soil Perception for adding nonlinearity.

3.3. Convolutional Neural Networks (CNN)

A feed-forward neural network is similar to a CNN. Data are processed to provide valuable results. CNN uses backpropagation, as well as semi-structured data [31], to train networks. Except for the input nodes, every node in the web is a processing element or a neuron. During "training," the output and hidden layer weights are changed. Two successive layers cannot be fully connected. Unnecessary consequences may be removed after or during learning. Figure 2 shows a CNN network with hidden layers. Multilayer perceptrons offer two advantages over other varieties.

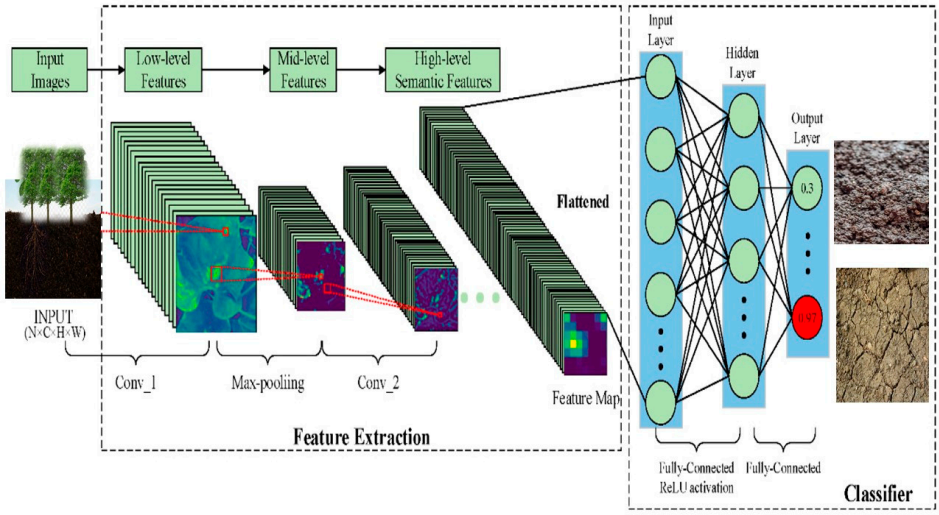


Figure 2. A CNN with N Inputs for precision agriculture.

Discriminant analysis approaches multilayer perceptrons to make no assumptions about how data are arranged or how covariance matrices of categorizable data sets are

presented. Second, multilayer perceptrons allow for the creation of disconnected and nonsequential regions.

3.4. Support Vector Machines

SVMs are a "supervised machine learning" approach used for classification. This is machine learning. It develops an objective function using Vapnik-Chervonenkis (VC) [32] theory and structural risk reduction, and then it finds a partition hyperplane that meets class requirements. A SVM training approach organizes incoming data into one of two classes using a series of training datasets. Using two classifications, this model contains incoming data. SVM operates binary division to establish the border between two classes, such as Positive-class 1 and Negative-class 2. Figure 3 shows two classes, with the hyperplane in between. This minimizes the distance between the plane and each data point. The nearest data points are used to create the support vectors.

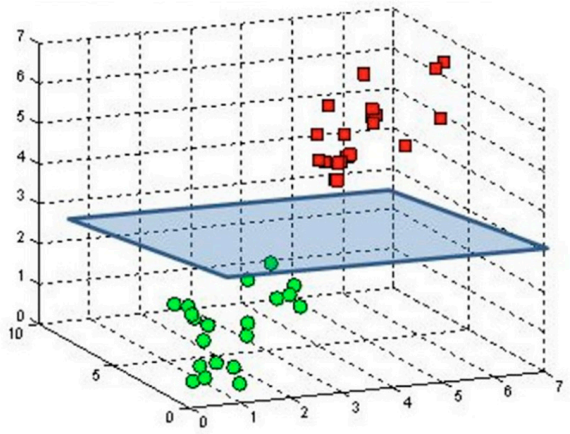


Figure 3. Classification of SVM with Hyperplanes as Class 0 and Class 1.

3.5. Two-Hierarchical Hybrid Ensemble Model

Figure 4 illustrates how the two-level soil moisture forecasting ensemble model presented is constructed from individual blocks and how it operates. And the working of the proposed algorithm is elaborated below.

The algorithm of the proposed approach is given below.
Input: Original data set Do and the Preprocessed Dataset Dp.
Set of heterogeneous first level learning classifiers Lf = {GPM, CNN, SVM}
Second-level learning algorithm Ls = {ZeroR}
Start
Prepare the dataset for ensemble training and testing.
Training dataset—Dataset used for ensemble training (TR). This
dataset contains soil moisture data.
Testing dataset—Dataset used for ensemble testing (TE). This
dataset contains soil moisture data.
for i = 1 to n
//where n represents the number of heterogeneous classifiers
hti = Lfi (TR)
Train an individual first-level learner by applying the corresponding
learning algorithm and the training data set
end
The output of base classifiers
Test the output from step 4 by using the second-level learning
algorithm Ls.
Output the predicted soil moisture value
Stop

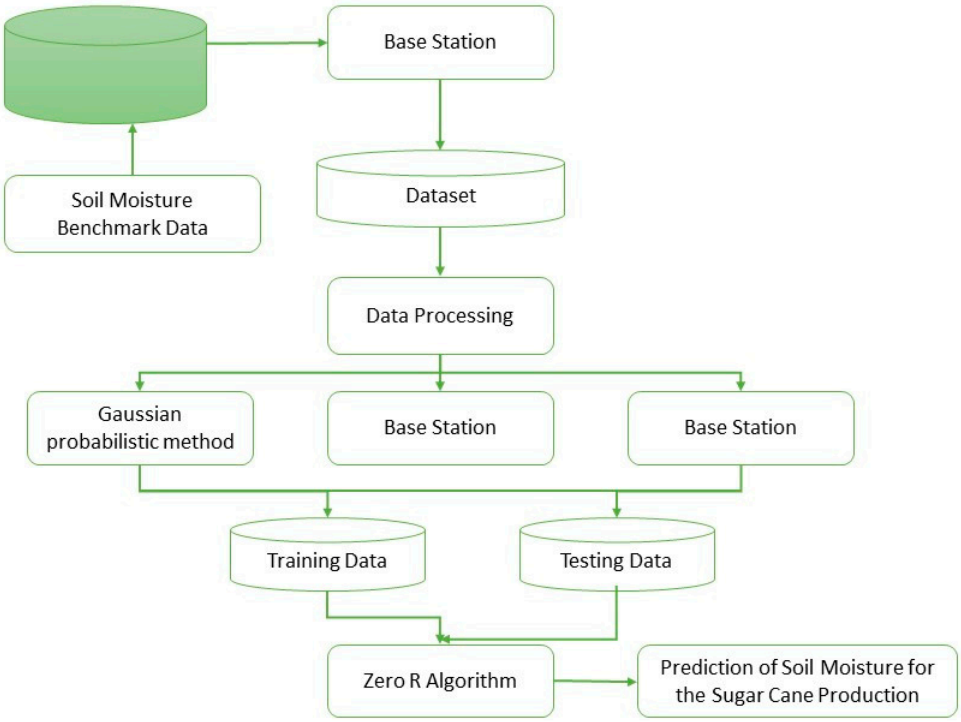


Figure 4. The proposed model for soil precision in agriculture using Hybrid Algorithm.

3.6. Godavari Plateau Delta Region Dataset

The research used soil moisture measurements from the Godavari River Plateau (https://apwrims.ap.gov.in) (accessed on 10 November 2022) and Krishna River Plateau (https://krmb.gov.in/krmb/home) [33], which involved online databases for both. The Krishna River plateau Soil Moisture Monitoring Network provides long-term soil moisture time data from 18 stations. The average of twice-monthly measurements is used. Figure 5 shows monthly data from 2016 to 2021 from a 10–30 cm layer at Station 1. Godavari river plateau Creek drains 600 km^2. Fourteen stations detect soil moisture in the Godavari River plateau Creek's basin. At 30-min intervals, soil moisture is monitored and averaged. Figure 6 shows the average daily soil moisture in the top 30 cm of station K4's profile from 3 January 2022, to 31 August 2022.

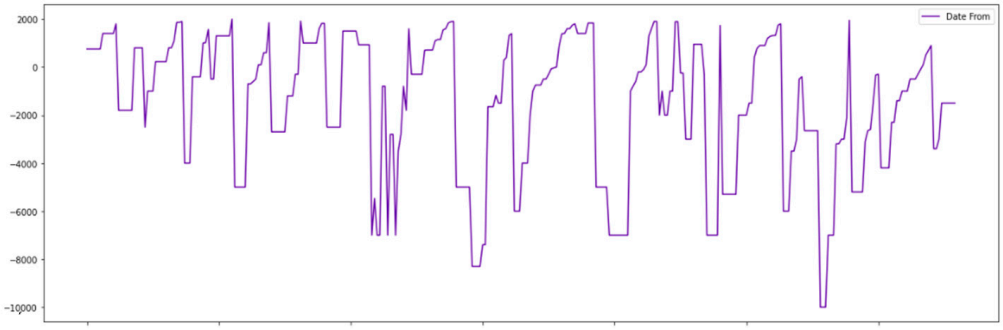


Figure 5. Shows the soil moisture of the Godavari Plateau Region.

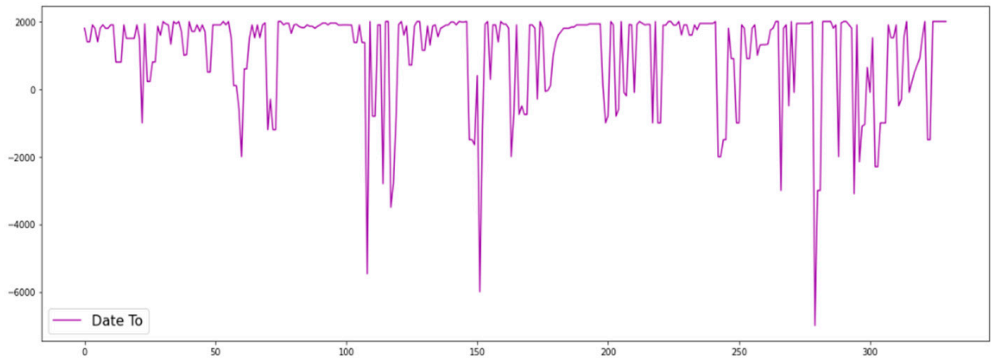


Figure 6. Krishna River plateau Soil Moisture.

3.7. Performance Evaluation Metrics of the Proposed Model

To determine how well the proposed model functions, statistical metrics, such as mean absolute error (MAE), root mean square error (RMSE), and mean absolute % error (MAPE), which are shown in Equations (1)–(3), are utilized.

$$MAE = \frac{\sum_{i=1}^{n} |xi' - xi|}{n} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i-1}^{n} (xi' - xi)^2}{n}} \tag{2}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{(xi' - xi)}{xi} \right| \times 100 \tag{3}$$

$Where$
$xi = actual\_value$
$xi' = \mathrm{P}redicted\_Value$
$n = number\_of\_observations$

The main objective of the function was to decrease the cost function between the actual and predicted values.

4. Results and Discussions

The Keras toolkit was used in the simulation-based testing that was carried out. This open-source program can predict time series and provides a machine-learning library that may be used for data mining jobs. As a consequence of this, this instrument is utilized to evaluate how well the proposed plan works. It is essential to partition the initial dataset into training and test sets before attempting any model evaluation. In this scenario, 70% of the dataset is put toward the training process, while the remaining 30% is used for testing.

Concerning the soil moisture datasets from the Krishna River plateau and the Godavari River plateau, tests were run with GPM, CNN, SVM, single-level GPM ensemble, single-level CNN ensemble, single-level SVM ensemble, and the proposed two-level soil moisture prediction model. Additionally, GPM, CNN, and SVM ensembles were tested individually.

The bagging approach generates single-level ensembles from GPM, CNN, and SVM classifiers. Examples of these ensembles are the GPM ensemble, the CNN ensemble, and the SVM ensemble. Figures 7 and 8 show the intended training and testing data projections for the Krishna River plateau data set when it was in the training phase and while it was in the testing phase, respectively. Figures 9 and 10 illustrate how a prediction was derived from the training and test data in the Godavari River plateau dataset.

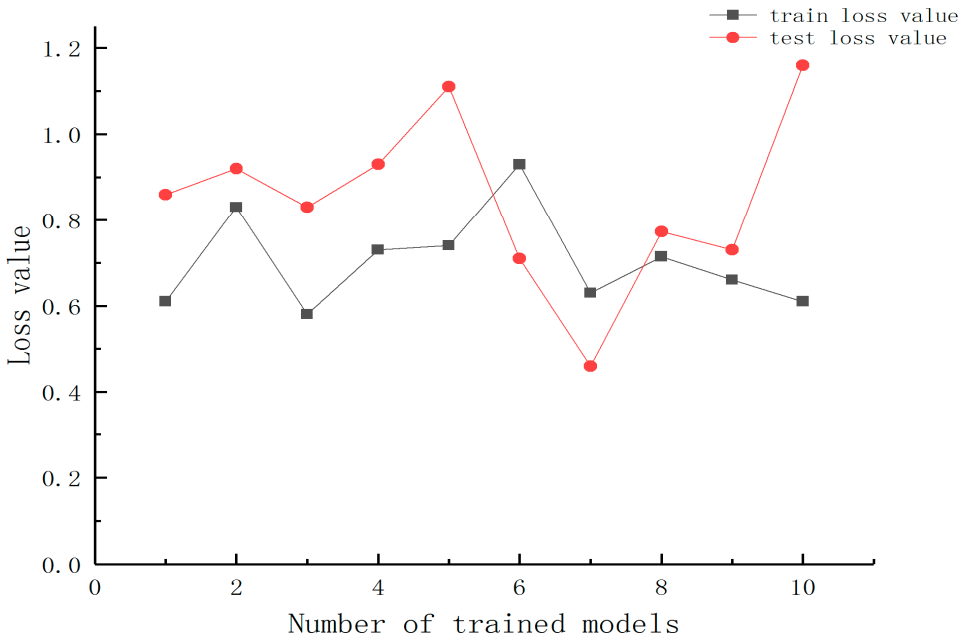


Figure 7. Forecast of training data—Godavari River plateau dataset.

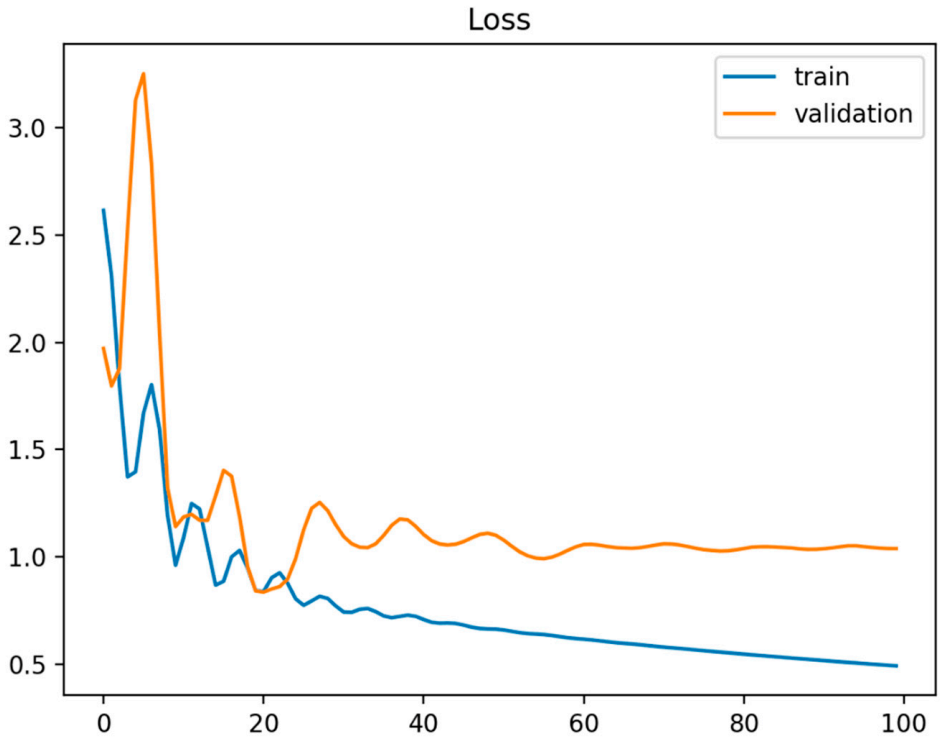


Figure 8. Forecast of testing data—Krishna River plateau dataset.

The one-step forward estimates of the future are shown in Figures 11 and 12 for two different datasets.

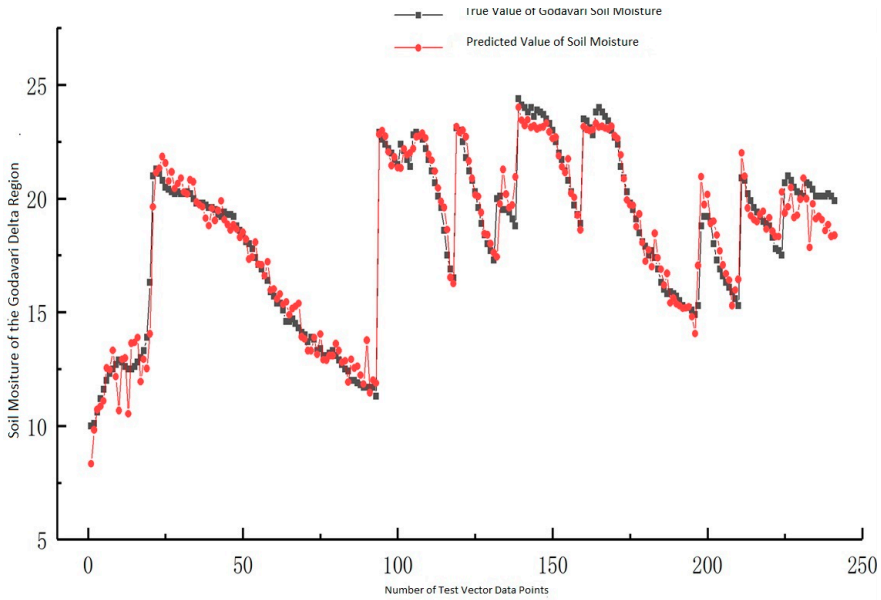


Figure 9. Forecast of training data—Godavari soil Moisture Dataset.

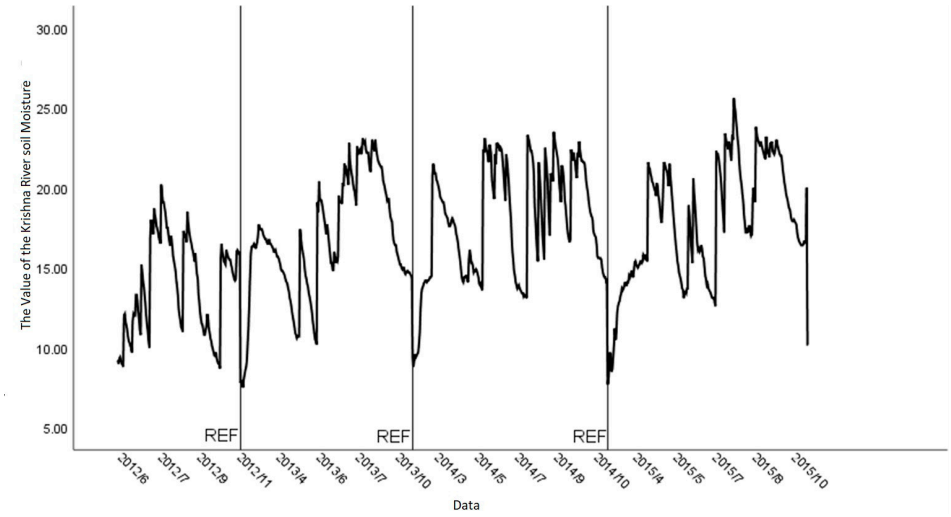


Figure 10. Forecast of testing data—Krishna Soil Moisture Dataset.

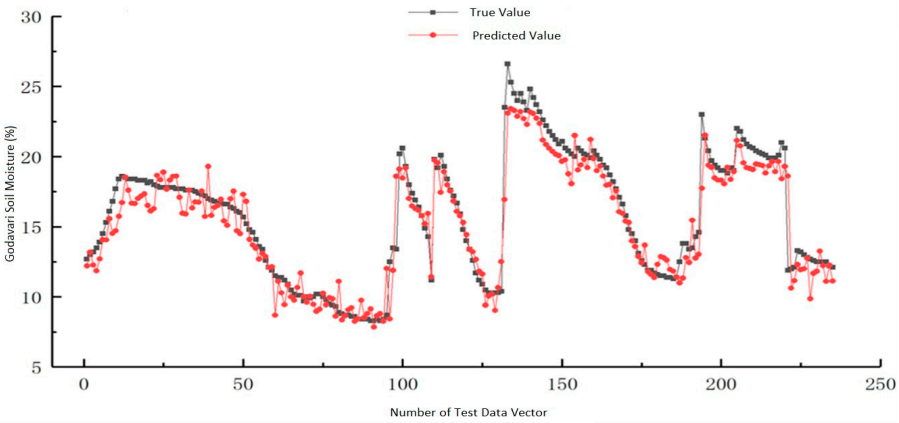


Figure 11. Future forecast of testing data—Godavari Soil Moisture Dataset.

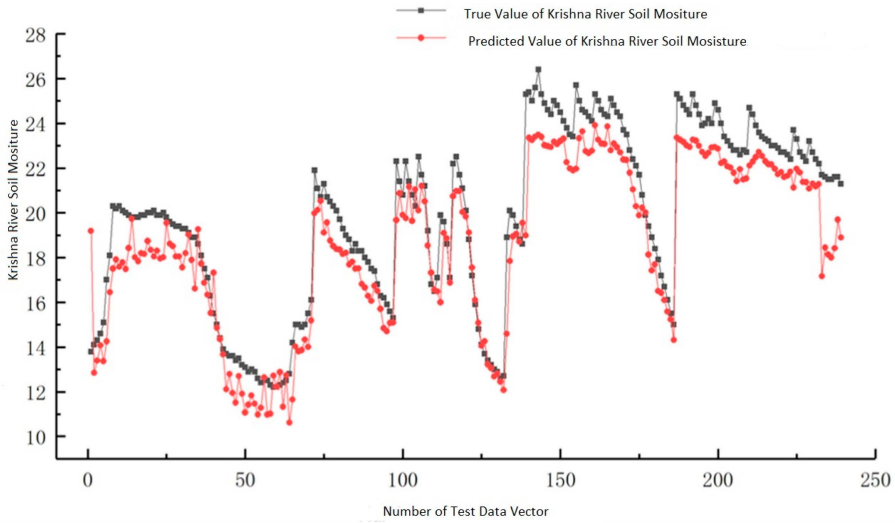


Figure 12. Future forecast of testing data–Krishna Soil Moisture Dataset.

Each machine learning method was applied to the same datasets to determine how accurately the suggested ensemble model could categorize items. Examples of individual classifiers are GPM, CNN, and SVM, in addition to the ensembles that go along with each of these. Using prediction accuracy, mean absolute error (MAE), root mean squared error (RMSE), and mean absolute % age error, the performance of the proposed model was compared to that of current machine learning approaches. Table 1 compares the effectiveness of several classifiers about the version of the one selected for use with the soil moisture data from the Krishna River plateau and Godavari River plateau. This is seen in Figures 13 and 14.

Table 1. Performance of the various classifiers with the proposed model.

	Krishna River Plateau Dataset			**Godavari River Plateau Dataset**		
Method	**MAE**	**RMSE**	**MAPE**	**MAE**	**RMSE**	**MAPE**
SVM	3.51	4.62	12.65	0.72	0.84	3.36
CNN	3.77	4.57	13.29	0.77	0.86	3.82
GPM	3.47	4.35	13.6	0.76	0.84	3.75
GPM-Bagging	3.43	4.42	12.61	0.77	0.98	4.06
CNN-Bagging	3.26	4.27	12.42	0.78	0.79	3.36
SVM-Bagging	3.24	4.12	12.52	0.69	0.81	3.29

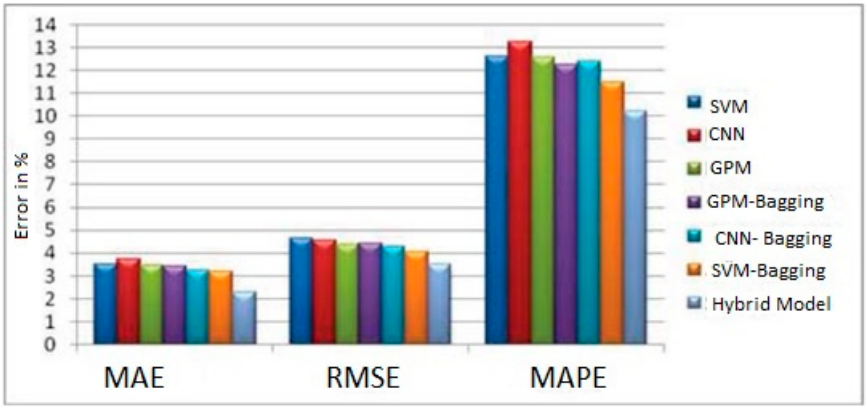


Figure 13. Performance of the various classifiers with the proposed model.

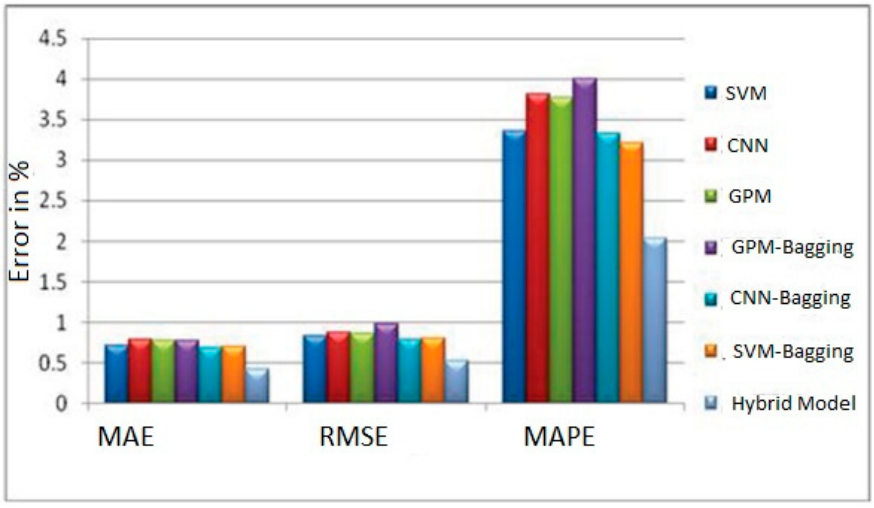


Figure 14. Performance of the various classifiers with the proposed model.

The ability of the two sets of data presented in Table 2 to accurately anticipate measurements is seen in Figures 15 and 16. The number of various classifiers in the testing set that could predict the one displayed accurately is one way to determine how accurate a measurement's prediction is.

Table 2. Prediction accuracy of various classifiers.

Classifier/Result	Prediction Accuracy	
	Krishna River Plateau Data	**Godavari River Plateau Data**
SVM	76.89%	81.23%
CNN	74.22%	71.98%
GPM	75.36%	77.78%
GPM-Ensemble	77.96%	78.92%
CNN-Ensemble	76.96%	75.69%
SVM-Ensemble	79.86%	83.59%
Hybrid Model	86.97%	92.69%

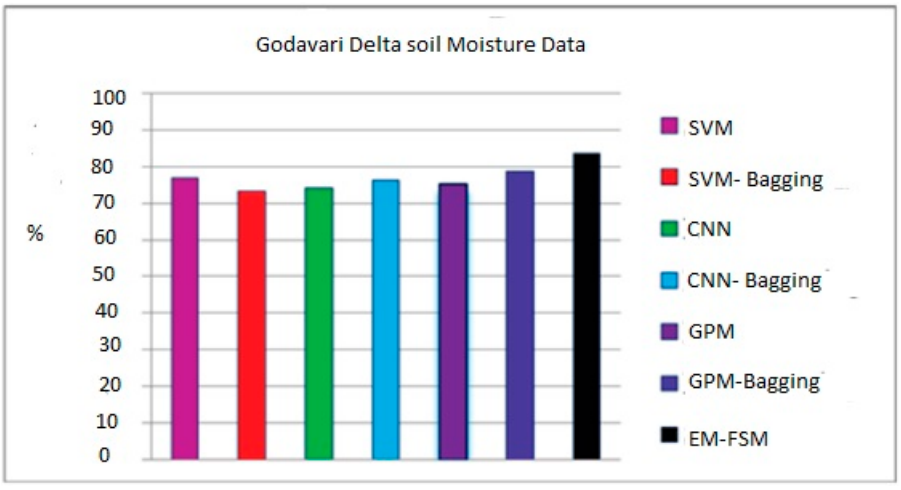


Figure 15. Prediction accuracy of different classifiers.

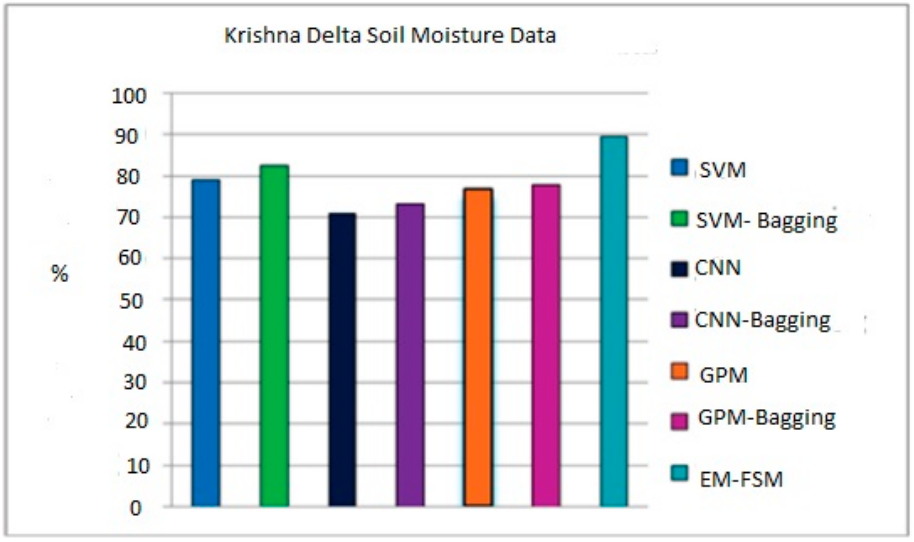


Figure 16. Prediction accuracy of different classifiers for the Godavari River plateau.

Compared to other individual classifiers, the SVM classifier has an accuracy ranging between 76.85% and 79%. The accuracy of the SVM-bagging ensemble is 78.81% and 82.56% higher than that of other bagging ensembles. Still, the accuracy of the proposed two-level ensemble model is 83.52% and 89.56% higher than that of all other ensemble models. The previously used solo classifiers and the single-level ensemble classifiers constructed using GPM, CNN, and SVM outperform the newly presented ensemble model. According to this research's findings, ensemble neural networks to forecast soil moisture data may result in more accurate forecasts. Using this model will assist farmers and others responsible for making policies in creating future irrigation plans. In this particular piece of research, a method for the long-term collection of data on soil moisture is investigated. The GPM, CNN, and SVM classifiers each provide their own set of helpful rules to the ensemble model that is being proposed. The experiment's findings demonstrate that, when several models are effectively connected, they are superior to individual learners in predicting soil moisture. Farmers can profit from this strategy, since it provides them withlinebreak reliable information.

Alterations to the methods used to cultivate crops, as well as new knowledge on the moisture content of the ground in the years to come. The findings of the studies indicate that the suggested ensemble model accomplishes its objective of increasing prediction accuracy. The most fantastic accuracy of prediction was found to be 83.53% for both of the datasets utilized and 89.53%, respectively. In addition to these applications, soil moisture estimates may be used to plan irrigation, calculate agricultural productivity, provide early warning of drought conditions, and forecast runoff.

5. Conclusions

It has been observed that sugarcane growth can be forecasted by looking at the pedogenetic traits of the soil moisture. The GPM helps to learn more about the different qualities of soil and the quickest ways to handle sugarcane. Using machine learning in sugarcane output was interesting because it allowed farmers to take advantage of traits that could be taught rapidly and cheaply.

Regarding the classification process, the hybrid approach recommended for the hybrid model for learning has an accuracy rate of 89.53%. The proposed system's efficiency was evaluated and contrasted with the efficiency of various other machine learning methods. The findings demonstrate that the strategy that was recommended is effective.

6. Future Study

The following alterations are currently in the works. Installing real-time sensor nodes across agricultural areas to monitor various factors for research purposes is important. More study is required to determine which route design will be most effective for farming data processing applications. Using the methodology outlined in this research work, researchers have conducted studies on crops that are more economically significant. To investigate and assess the performance of various hybrid evolutionary computing technologies to find a solution to the existing challenge is important.

Author Contributions: Conceptualization, E.S.N.J.; methodology, D.B.; software, E.S.N.J. and D.B.; validation, T.-h.K. and E.S.N.J.; formal analysis, N.T.R.; writing—original draft preparation, E.S.N.J., T.-h.K., D.B. and N.T.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are publicly available and Benchmarked.

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

References

1. Bhargavi, G.; Arunnehru, J. An extreme machine learning model for evaluating landslide hazard zonation in nilgiris district, causative factors and risk assessment using earth observation techniques. In *Lecture Notes on Data Engineering and Communications Technologies*; Springer Nature Singapore: Singapore, 2023. Available online: www.scopus.com (accessed on 23 November 2022). [CrossRef]
2. United Nations World Meter. Available online: https://www.worldometers.info (accessed on 10 April 2023).
3. Huang, Y.; Lan, Y.; Thomson, S.J.; Fang, A.; Hoffmann, W.C.; Lacey, R.E. Development of soft computing and applications in agricultural and biological engineering. *Comput. Electron. Agric.* **2010**, *71*, 107–127. [CrossRef]
4. Asha, R.; Babu, G.S.K.; Teja, T.S. Production and marketing of sugarcane in Visakhapatnam district of Andhra Pradesh. *J. Res. Angrau* **2019**, *47*, 69–77. Available online: https://epubs.icar.org.in/index.php/TJRA/article/view/133728 (accessed on 23 November 2022).
5. Singh, P.; Pandey, P.C.; Petropoulos, G.P.; Pavlides, A.; Srivastava, P.K.; Koutsias, N.; Deng, K.A.K.; Bao, Y. Hyperspectral remote sensing in precision agriculture: Present status, challenges, and future trends. In *Earth Observation, Hyperspectral Remote Sensing*; Pandey, P.C., Srivastava, P.K., Balzter, H., Bhattacharya, B., Petropoulos, G.P., Eds.; Elsevier: Amsterdam, The Netherlands, 2020; pp. 121–146. ISBN 9780081028940. [CrossRef]
6. Singh, D.; Jain, N.; Jain, P.; Kayal, P.; Kumawat, S.; Batra, N. PlantDoc: A Dataset for Visual Plant Disease Detection. *arXiv* **2019**, arXiv:1911.10317. [CrossRef]
7. Anas, M.; Liao, F.; Verma, K.K.; Sarwar, M.A.; Mahmood, A.; Chen, Z.-L.; Li, Q.; Zeng, X.-P.; Liu, Y.; Li, Y.-R. Fate of nitrogen in agriculture and environment: Agronomic, eco-physiological and molecular approaches to improve nitrogen use efficiency. *Biol. Res.* **2020**, *53*, 47. [CrossRef] [PubMed]
8. Kai, P.M.; de Oliveira, B.M.; de Costa, R.M. Deep Learning-Based Method for Classification of Sugarcane Varieties. *Agronomy* **2022**, *12*, 2722. [CrossRef]
9. Nihar, A.; Patel, N.R.; Danodia, A. Machine-Learning-Based Regional Yield Forecasting for Sugarcane Crop in Uttar Pradesh, India. *J. Indian Soc. Remote Sens.* **2022**, *50*, 1519–1530. [CrossRef]
10. Militante, S.V.; Gerardo, B.D. Detecting Sugarcane Diseases through Adaptive Deep Learning Models of Convolutional Neural Network. In Proceedings of the 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), Kuala Lumpur, Malaysia, 20–21 December 2019; pp. 1–5. [CrossRef]
11. Wang, J.; Guo, J.Y. Sugarcane cultivation management technology. *Agric. Technol. Serv.* **2013**, *30*, 321–322.
12. Everingham, Y.; Sexton, J.; Skocaj, D.; Inman-Bamber, G. Accurate prediction of sugarcane yield using a random forest algorithm. *Agron. Sustain. Dev.* **2016**, *36*, 27. [CrossRef]
13. Talaviya, T.; Shah, D.; Patel, N.; Yagnik, H.; Shah, M. Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. *Artif. Intell. Agric.* **2020**, *4*, 58–73. [CrossRef]
14. Blackmore, I.; Rivera, C.; Waters, W.F.; Iannotti, L.; Lesorogol, C. The impact of seasonality and climate variability on livelihood security in the Ecuadorian Andes. *Clim. Risk Manag.* **2021**, *32*, 100279. [CrossRef]
15. Kok, Z.H.; Shariff, A.R.M.; Alfatni, M.S.M.; Khairunniza-Bejo, S. Support Vector Machine in Precision Agriculture: A review. *Comput. Electron. Agric.* **2021**, *191*, 106546. [CrossRef]
16. Tavva Srinivas, M. Nedunchezhiyan The nexus between adoption and diffusion of production technologies with yield: Evidence from sweet potato farmers in India. *Technol. Soc.* **2020**, *60*, 101208. [CrossRef]
17. Sequeros, E.; Schreinemachers, P.; Depenbusch, L.; Shwe, T.; Nair, R.M. Impact and returns on investment of mungbean research and development in Myanmar. *Agric. Food Secur.* **2020**, *9*, 5. [CrossRef]

18. Uyeh, D.D.; Iyiola, O.; Mallipeddi, R.; Asem-Hiablie, S.; Amaizu, M.; Ha, Y.; Park, T. Grid Search for Lowest Root Mean Squared Error in Predicting Optimal Sensor Location in Protected Cultivation Systems. *Front. Plant Sci.* **2022**, *13*, 920284. [CrossRef] [PubMed]
19. Solomon, S.; Swapna, M. Indian Sugar Industry: Towards Self-reliance for Sustainability. *Sugar Tech* **2022**, *24*, 630–650. [CrossRef] [PubMed]
20. Tian, H.; Wang, P.; Tansey, K.; Zhang, J.; Zhang, S.; Li, H. An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China. *Agric. For. Meteorol.* **2021**, *310*, 108629. [CrossRef]
21. Kai, P.M.; Oliveira, B.M.; Vieira, G.S.; Soares, F.; Costa, R.M. Effects of resampling image methods in sugarcane classification and the potential use of vegetation indices related to chlorophyll. In Proceedings of the 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), Madrid, Spain, 12–16 July 2021; pp. 1526–1531.
22. Panwar, E.; Singh, D.; Sharma, A.K. Exploring the Possibility of Assessing Biochemical Variables in Sugarcane Crop with Sentinel-2 Data. In Proceedings of the 2020 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2020), Virtual, 26 September–2 October 2020; pp. 3712–3715.
23. Shi, L.; Duan, Q.; Ma, X.; Weng, M. The Research of Support Vector Machine in Agricultural Data Classification. In Proceedings of the Computer and Computing Technologies in Agriculture V. CCTA 2011. IFIP Advances in Information and Communication Technology, Beijing, China, 29–31 October 2011; . Li, D., Chen, Y., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 370. [CrossRef]
24. Abdel-Sattar, M.; Aboukarima, A.M.; Alnahdi, B.M. Application of artificial neural network and support vector regression in predicting mass of ber fruits (*Ziziphus mauritiana* Lamk.) based on fruit axial dimensions. *PLoS ONE* **2021**, *16*, e0245228. [CrossRef] [PubMed]
25. Lv, M.; Wei, H.; Fu, X.; Wang, W.; Zhou, D. A Loosely Coupled Extended Kalman Filter Algorithm for Agricultural Scene-Based Multi-Sensor Fusion. *Front. Plant Sci.* **2022**, *13*, 849260. [CrossRef]
26. Gałęzewski, L.; Jaskulska, I.; Jaskulski, D.; Lewandowski, A.; Szypłowska, A.; Wilczek, A.; Szczepańczyk, M. Analysis of the need for soil moisture, salinity and temperature sensing in agriculture: A case study in Poland. *Sci. Rep.* **2021**, *11*, 16660. [CrossRef]
27. Lu, J.; Tan, L.; Jiang, H. Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification. *Agriculture* **2021**, *11*, 707. [CrossRef]
28. Matko, V.; Koprivnikar, J. Quartz sensor for water absorption measurement in glass-fiber resins. *IEEE Trans. Instrum. Meas.* **1998**, *47*, 1159–1162. [CrossRef]
29. Eze, P.N.; Kumahor, S.K. Gaussian process simulation of soil Zn micronutrient spatial heterogeneity and uncertainty—A performance appraisal of three semivariogram models. *Sci. Afr.* **2019**, *5*, e00110. [CrossRef]
30. Cravero, A.; Pardo, S.; Galeas, P.; López Fenner, J.; Caniupán, M. Data Type and Data Sources for Agricultural Big Data and Machine Learning. *Sustainability* **2022**, *14*, 16131. [CrossRef]
31. Gey, S. Vapnik–Chervonenkis dimension of axis-parallel cuts. *Commun. Stat. Theory Methods* **2018**, *47*, 2291–2296. [CrossRef]
32. Andhra Pradesh Water Resources Management. Available online: https://apwrims.ap.gov.in/ (accessed on 10 November 2022).
33. Krishna River Board Management. Available online: https://krmb.gov.in/krmb (accessed on 10 November 2022).

agronomy

MDPI

Article

Hyperspectral Non-Imaging Measurements and Perceptron Neural Network for Pre-Harvesting Assessment of Damage Degree Caused by Septoria/Stagonospora Blotch Diseases of Wheat

Sofia V. Zhelezova [1], Elena V. Pakholkova [1], Vladislav E. Veller [1,2], Mikhail A. Voronov [1], Eugenia V. Stepanova [1], Alena D. Zhelezova [1,3,*], Anton V. Sonyushkin [4], Timur S. Zhuk [1] and Alexey P. Glinushkin [1]

[1] Federal State Budgetary Scientific Institution All-Russian Scientific Research Institute of Phytopathology (VNIIF), Institut Street, 5, Bolshie Vyazemy, Moscow Region 143050, Russia; soferrum@mail.ru (S.V.Z.); epaholkova@mail.ru (E.V.P.); vellervladislav@gmail.com (V.E.V.); mvoronov97@gmail.com (M.A.V.); step@vniif.ru (E.V.S.); timurjuk2000@gmail.com (T.S.Z.)
[2] BASF, Department of Russian Federation, Leningradsky Ave, 37A Building 4, Moscow 125167, Russia
[3] Dokuchaev Soil Science Institute, Department of Soil Biology and Biochemistry, Pyzhyovskiy Lane 7 Building 2, Moscow 119017, Russia
[4] Institute of Geography, Russian Academy of Sciences, Staromonetniy Lane, 29, Moscow 119017, Russia; asonyushkin@opendatabox.info
* Correspondence: zhelezova_ad@esoil.ru

Abstract: The detection and identification of plant diseases is a fundamental task for sustainable crop production. Septoria tritici and Stagonospora nodorum blotch (STB and SNB) are two of the most common diseases of cereal crops that cause significant economic damage. Both pathogens are difficult to identify at early stages of infection. Determining the degree of the disease at a late infection stage is useful for assessing cereal crops before harvesting, as it allows the assessment of potential yield losses. Hyperspectral sensing could allow for automatic recognition of Septoria harmfulness on wheat in field conditions. In this research, we aimed to collect information on the hyperspectral data on wheat plants with different lesion degrees of STB&SNB and to create and train a neural network for the detection of lesions on leaves and ears caused by STB&SNB infection at the late stage of disease development. Spring wheat was artificially infected twice with Septoria pathogens in the stem elongation stage and in the heading stage. Hyperspectral reflections and brightness measurements were collected in the field on wheat leaves and ears on the 37th day after STB and the 30th day after SNB pathogen inoculation using an Ocean Insight "Flame" VIS-NIR hyperspectrometer. Obtained non-imaging data were pre-treated, and the perceptron model neural network (PNN) was created and trained based on a pairwise comparison of datasets for healthy and diseased plants. Both statistical and neural network approaches showed the high quality of the differentiation between healthy and damaged wheat plants by the hyperspectral signature. A comparison of the results of visual recognition and automatic STB&SNB estimation showed that the neural network was equally effective in the quality of the disease definition. The PNN, based on a neuron model of hyperspectral signature with a spectral step of 6 nm and 2000–4000 value datasets, showed a high quality of detection of the STB&SNB severity. There were 0.99 accuracy, 0.94 precision, 0.89 recall and 0.91 F-score metrics of the PNN model after 10,000 learning epochs. The estimation accuracy of diseased/healthy leaves ranged from 88.1 to 97.7% for different datasets. The accuracy of detection of a light and medium degree of disease was lower (38–66%). This method of non-imaging hyperspectral signature classification could be useful for the identification of the STB and SNB lesion degree identification in field conditions for pre-harvesting crop estimation.

Keywords: septoriosis; Septoria tritici blotch; hyperspectral signature; hyperspectral disease detection; data science; neural network; wheat

Citation: Zhelezova, S.V.; Pakholkova, E.V.; Veller, V.E.; Voronov, M.A.; Stepanova, E.V.; Zhelezova, A.D.; Sonyushkin, A.V.; Zhuk, T.S.; Glinushkin, A.P. Hyperspectral Non-Imaging Measurements and Perceptron Neural Network for Pre-Harvesting Assessment of Damage Degree Caused by Septoria/Stagonospora Blotch Diseases of Wheat. *Agronomy* **2023**, *13*, 1045. https://doi.org/10.3390/agronomy13041045

Academic Editors: Jian Zhang, Randy G. Goebel, Zhihai Wu and Alberto San Bautista

Received: 17 January 2023
Revised: 13 March 2023
Accepted: 30 March 2023
Published: 1 April 2023

1. Introduction

Sustainable crop production is a basis for food security. Wheat is one of the most important cereal crops, and its productivity could be significantly constrained by several pathogenic species. Septoria tritici blotch and Stagonospora nodorum blotch (STB and SNB), or septoriosis, are widely spread and common diseases of cereal crops. Under this name, several variants of the disease are combined that are caused by different pathogens with the similar mechanism of action. This pathogen complex affects wheat, rye, triticale, barley and many other crops, as well as about 20 types of wild cereals [1]. Septoriosis is a common disease in all small-grain-cereal-growing territories on all continents, especially in the medium- and high-rainfall zones [1–6].

Septoriosis causes significant economic damage due to crop losses and increased fungicide application. Annual economic losses from septoriosis were estimated at EUR 0.28–1.2 billion per year in Europe and more than USD 275 million per year in the United States [2,7]. In Europe, cereal crop losses in epiphytotic years could range from 20 to 40% and losses of wheat yields could be up to 50% [2]. In Russia, septoriosis is one of the most spread leaf–stem fungal diseases of wheat. Mass outbreaks of this disease are noted every 5 years out of 10; crop losses in epiphytotic years can range from 20 to 40% [8–10].

On wheat crops in Russia, two types of STB pathogens are the most common: *Zymoseptoria tritici* (Desm.) Quaedvlieg and Crous (synonym of *Septoria tritici* Rob. et Desm) and *Parastagonospora nodorum* (Berk.) Quaedvlieg, Verkley and Crous (synonym of *Stagonospora nodorum* (Berk.) Castellani and E. G. Germano) [8–10]. The first pathogen is the causative agent of septoria leaf blotch and it affects mainly leaves. The second pathogen is the causative agent of septoria glume blotch and it affects both leaves and ears equally. The Septoria pathogens affect plants at any time of the growing season, in several stages of vegetation, causing the spotting of leaves, stems and ears. The main damage from septoriosis, as well as other plant leaf diseases, is a reduction in plants' photosynthetic assimilation surface area.

Traditionally, visual assessment is used to determine the degree of leaf damage using the nominal, ordinal and ratio scales [11]. In recent years, the automatic detection of plant diseases using images has also become widely applied. Many databases on plant diseases are represented by sets of digital photographic images [12,13]. The results of multi- and hyperspectral imaging are also used to determine some problems of plant health, such as salinity stress, drought and fungal and viral diseases [11,14–19]. Various methods are used to process sets of digital, multi- and hyperspectral images, from non-parametric statistical methods (principal component analysis, cluster analysis and others) to automatic data processing based on neural networks. The basis of automatic methods for determining plant diseases is a comparison of photographs of healthy and diseased plants with clearly presented disease signs [13,20,21]. With a sufficient number of images in the training sample, neural networks (NNs) could easily and with high accuracy identify various plant diseases from digital photographs and hyperspectral survey results.

Currently, different approaches and models of neural networks are used to extract data and select a classification model. The extraction of an image's features is the first important step for successful image classification. The precision of classification depends on (1) quality of input image (or hyperspectral) information and (2) the advantages and disadvantages of machine learning algorithms [18,22]. A convolutional neural network (CNN) is an example of a deep learning neural network. It uses computer vision for image recognition and classification. For automatically determining plant diseases, the number of CNN models can be used, for example, AlexNet, GoogLeNet, VGG19, ResNet50 and others [13,21,23]. CNN is a common method used for image classification and computer vision. Another method of the neural network is multilayer perceptron (MLP). This method can be used for non-imaging hyperspectral data [16]. In our research, non-imaging spectral curves were used instead of images, and there were none applied to the CNN class.

It is necessary to find the effective methods of plant disease detection in plants for timely diagnostics and the prevention of notable damage. Optical imaging techniques such

as RGB, thermal, fluorescence, multi- and hyperspectral imaging are non-invasive methods of plant disease detection [11,15,24,25]. RGB with hyper-spectral imaging was successfully applied to identify various diseases of sugar beet leaves and other crops [14,25], yellow rust and Fusarium head blight in cereal crops [26], and STB in wheat [27–29]. Both hyperspectral imaging and non-imaging measurements are effective for the determination of various visible symptoms of plant stress or infection [16,28–32]. A specialized library of spectral characteristics (signatures) of the main foliar diseases of wheat was created for pathogen detection and identification [32]. Many studies are dedicated to obtaining and analyzing spectral images in controlled laboratory conditions [33]. The method of disease development assessment depends on the scale of research: the cellular level, the level of the leaf, the individual plant, the canopy of the crop, the field and the landscape [11,14,15,34,35]. Field conditions make constraints for studies using hyperspectral imaging: the light intensity can change significantly during the shooting, which makes it difficult to interpret the result. Some works emphasize that the methods developed for the laboratory for the detection of plant diseases cannot be transferred to field conditions [33].

The detection of STB pathogens on leaves and SNB on ears in the field is difficult due to the long latent stage of disease. The visual diagnostic characteristic of Septoriosis on the leaves and ears is the presence of black fruiting bodies (pycnidia) within pale grey to dark brown spots. Leaf discoloration is associated with chlorosis and tissue necrosis. The necrotic phase and the first chlorotic tissue become visible 12–13 days after infection (dai). These signs can be easily visually detected and by using optical and spectral cameras [11,36]. The possibilities of hyperspectral imaging and machine learning for assessing the development of Septoriosis were shown in laboratory conditions at the level of leaves and ears [32,35], in field conditions at the level of leaves and the crop canopy [27,28] and at the level of plots and fields [19,29].

All previous studies have stated the indisputable fact that Septoria disease is difficult to identify in the initial stages of development. To one degree or another, this problem of disease identification has been solved with the use of hyperspectral surveys. In the last stages of this disease's development, it can be easily identified using visual detection, but researchers can only state the fact that the disease caused weak, medium or severe damage. However, this knowledge is useful for special aims of inventory. The assessment of the damage caused by the disease in the pre-harvesting stage of crop development can be used to predict crop yield losses from septoriosis, to estimate the effectiveness of protective measures carried out or to compare the cultivars' sensitivity to this disease. Visual pre-harvesting crop inspection in fields is limited because of subjective reasons, and automatic inspection can be more convenient and faster. This study shows the possibility of automatic assessment of the degree of wheat damage by septoriosis based on the use of non-imaging hyperspectral analysis and the perceptron neural network.

The aim of the study was to examine the results of automatic identification of the Septoriosis degree in the pre-harvesting stage of wheat on the basis of hyperspectral non-imaging measurements in the field condition and a comparison of the results of visual and automatic STB&SNB estimation.

2. Materials and Methods

2.1. Wheat Cultivation and Pathogen Inoculation

For the dataset creation it was necessary to observe healthy and diseased leaves and ears with different degrees of Septoria tritici and Stagonospora nodorum blotch. Two moderately susceptible cultivars of spring wheat (*Triticum aestivum*) were used: Stepnodar-90 (originating from Kazakhstan) and Aekada-282 (originating from Tatarstan, Russia).

For the collection of spring wheat cultivars (*Triticum aestivum*), they were grown according to standard growth technology in the field condition in the Moscow region, Russia. The soil of the site was sod-podzolic medium-loamy. The standard technology used for the growing of the spring wheat cultivar collection included (1) pre-winter soil plowing after the harvesting of the previous crop sunflower, where the depth of conventional

plowing was 24 cm; (2) fertilization with cultivation in the spring, two days before sowing, where the doses of complex fertilizer NPK were 60, 36 and 36 kg·ha^{-1}, respectively; (3) sowing with the density equivalent to 4.5 million seeds per hectare, where the inter-row distance was 15 cm and (4) weeding as needed. Since this was a collection of varieties for studying resistance to STB&SNB, no pesticide treatments were carried out on this collection. Seeds of the wheat cultivar collection were sown at the 24th of May 2022, and the plants developed normally during the whole growing season. Inoculation with spores of Septoria pathogens was carried out twice: for STB on 27th of June in the growth stage BBCH-35-37 (flag leaf just visible, still rolled) and for SNB on 4th of July in earing stage BBCH-55 [37].

To create an artificial infection background, we used pathogenic strains of *Parastagonospora nodorum* and *Zymoseptoria tritici* from the core facility center "State Collection of Phytopathogenic Microorganisms" of the Federal State Budgetary Scientific Institution All-Russian Scientific Research Institute of Phytopathology (VNIIF).

Inoculation of wheat plots was carried out in the most sensitive phases for each type of fungus: in the stem elongation phase for *Z. tritici*, and in the heading phase for *P. nodorum*. Inoculums were evenly applied to the plants with a spray gun in the evening after dew had fallen. The concentrations of suspensions were 1×10^6 spores/mL for *P. nodorum* and 1×10^7 spores/mL for *Z. tritici*. A total of 100 mL of suspension was applied to 1 m^2 of the crop's canopy. The lesions of STB were firstly detected visually on 9th of July, on the 12th dai (day after infection for STB). Up until the end of July, wheat had reached the growth stage of the maturity beginning (BBCH-70). On 29th of July (32nd dai of STB and 25th dai of SNB), wheat leaves and ears were investigated using a non-imaging hyperspectrometer sensor. Because the early detection of STB&SNB symptoms via non-imaging sensors is limited, especially for low severity [11], this was a reason to determine the degree of STB&SNB severity at the mature stage of cultivar–pathogen interaction. Hyperspectral reflection brightness measurements were carried out on live plants in the field (in situ) at the leaf and ear level. The degree of disease had previously been assessed visually. Healthy leaves and leaves with different degrees of lesion (light, medium or severe) were chosen and observed in situ in the collection for hyperspectral imaging (Figure 1a–c). The same degrees of disease severity were selected for the wheat ears (Figure 1d–f). For each degree of STB&SNB severity, 3 replicates of leaves and ears were taken, respectively. After this, an individual unique dataset of non-imaging hyperspectral signatures was compiled for each replicate.

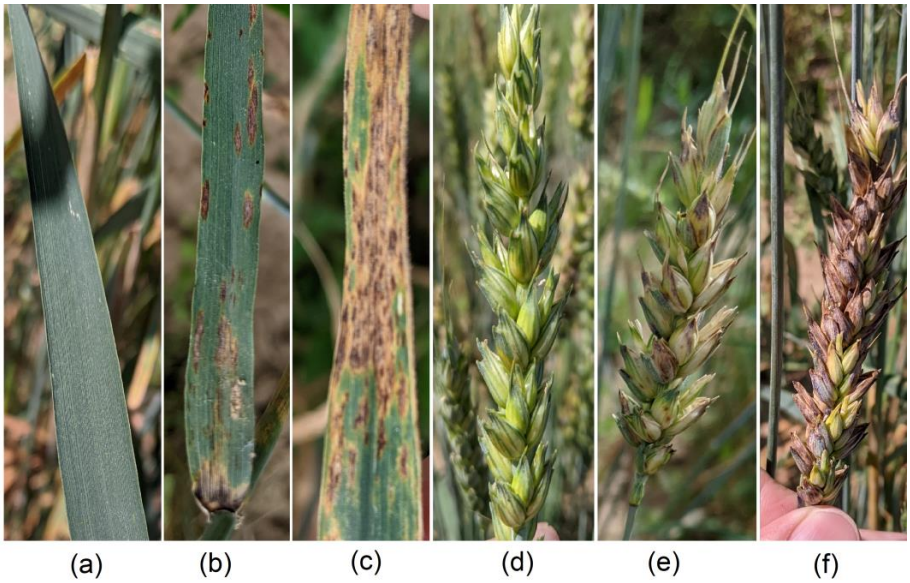


Figure 1. The degrees of septoriosis disease severity on spring wheat leaves and ears: (**a**) healthy leaf; (**b**) medium lesion on leaf; (**c**) severe lesion on leaf; (**d**) healthy ear; (**e**) medium lesion on ear; (**f**) severe lesion on ear.

At the end of the vegetation season, the biological yield of all wheat cultivars was recorded. The regression equation of the relationship between Septoriosis disease degree and crop yield was calculated.

2.2. Hyperspectral Non-Imaging Measurements

Hyperspectral non-imaging reflection signatures of wheat leaves and ears were obtained using the Ocean Insight "Flame" VIS-NIR hyperspectrometer [38]. This passive optical spectrometer registers the intensity of electromagnetic radiation (EMR) in the wavelength range from 338 to 1018 nm, including visible (VIS) and near infrared (NIR) diapasons. The spectral resolution (step) of the measurements was 0.3 nm. The data of reflection intensity were recorded in a text file format from which a spectral reflection curve (signature) could be designed. One signature was extracted from a unit *.txt file. The curve of each measurement file had an input of 2048 single indicators of spectral reflection for each 0.3 nm wavelength gradation in the range from 338 to 1018 nm. Immediately before the start of the non-imaging measurements, the spectrometer device was calibrated via reflection from a white rough fluoroplastic plate (standard white).

The size of the spectrometer field of vision depends on the distance from the collimator to the leaf surface; in our experiment, it was a spot with size of about 5 mm $\times$ 5 mm. The leaf necrosis and chlorosis spots can be either larger or smaller than the field of vision size. Measurements were carried out continuously when the eyepiece of the spectrometer moved near the surface of the leaves at the distance of 2.5–3 cm. Within the visible area, spectral information at each wavelength was averaged.

There were 8 unique datasets of leaves, and 8 datasets of ears were created. Every dataset includes series of non-imaging measurements according to the degree of STB for leaves and SNB for ears. Leaves and ears were chosen manually according to visual estimation. There were 2 datasets of healthy leaves, 2 datasets of healthy ears and the other 12 datasets included different stages of STB&SNB severity of leaves or ears. Every dataset included 2000 non-imaging hyper-spectral signatures for leaves, and 2000 or 4000 signatures for ears. The volume of the datasets on the ears was larger, because the surface of the ear is uneven. In such a condition, the raw data had more noise and irrelevant values related to the absorption and reflection of light by the ear surface. So, increasing the size of the dataset was necessary, because increasing the sample size reduces the standard error, and the training of neural networks is expected to be more reliable.

The time-line period for one dataset containing 2000 values depended on the intensity of the external light flux and usually ranged from 2 to 10 min. Measurements were performed in the field conditions on midday under natural bright sunlight, and it took 1–2 min time-lines for one survey (unique dataset). In total, 16 datasets (each of 2000 or 4000 records) were collected in half an hour; so, the change in the angle of sun inclination and the illumination conditions could be neglected.

2.3. Hyperspectral Data Analysis and Neural Network

Raw hyperspectral reflection data vary in a wide range of values, especially in the NIR diapason. Data preprocessing and the design of spectral brightness curves were carried out in a specially created Python script, using numpy, scipy and matplotlib standard libraries. Noise, outliers and irrelevant values were removed in each unique dataset (raw spectral reflection data). There were discarded values, whereby deviations exceeded 3σ from the average value of the dataset for a given wavelength. Hyperspectral brightness reflection curves were designed in a software module and inspected visually and by using classic statistics methods. The spread distribution of values over the entire spectrum range was estimated, and a quantile analysis and comparison of the spectral brightness curves were performed based on the values averaged for each dataset.

The neural network was created and trained on the basis of pairwise comparison of standard datasets of "healthy leaf"/"severely damaged leaf" and "healthy ear"/"severely damaged ear" for further classification of plants with signs of disease. We used the

perceptron model of the neural network, PNN (Figure 2) [39]. Two datasets were used to train the neural networks: hyperspectral characteristics of healthy (1) and severely damaged, diseased (2) plants (Supplementary File S1). The training datasets were divided into proportions of 80% and 20% for training and validation parts, respectively. Several perceptron neural network models were created and tested, and a model with two layers of neurons was chosen. There were 5 neurons on the first layer and 2 neurons on the second layer; the number of training epochs was 5000 for the leaves datasets and 10,000 for the ears datasets. In such a configuration of the neural network, the predicted percentage of correct answers was 89% for leaves and 78% for ears. After the training and testing of the neural network, every new dataset of leaves and ears with different damage degrees was tested in this neural configuration. The metrics of the PNN model after 10,000 learning epochs were 0.99 accuracy, 0.94 precision, 0.89 recall and 0.91 F-score.

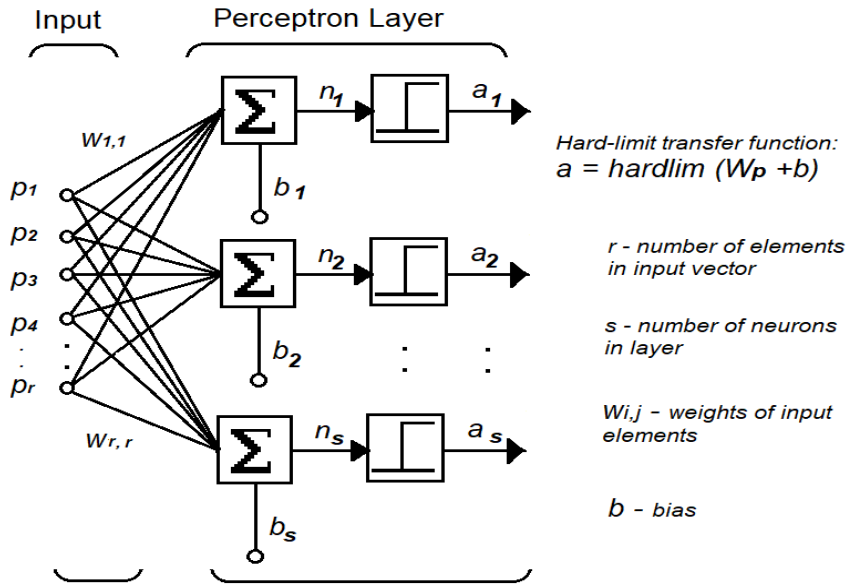


Figure 2. The perceptron neural network architecture. Inputs p1, p2 . . . pr are the hyperspectral signatures from one dataset.

Other traditional convolution neural network models (CNNs) earlier proposed for plant disease classification by images could not be used in our investigation because we worked with non-imaging signatures. So, we used only one PNN model and the results were compared with the visual estimation of STB severity via a nominal scale [11].

3. Results

3.1. Hyperspectral Data Visualization and Distribution Quantile Analysis

A total collection of spectrograms (spectral reflection curves) in the range from 330 to 1018 nm was created. For each dataset of hyperspectral measurement survey, 2000 curves for leaves (or 4000 for ears) were obtained. Every set of curves, designed using the computing module of Python, illustrates a wide range of values of spectral reflection (Figure 3). Due to a comparable size of the device's field of vision and necrosis-chlorosis spots on leaves, every shooting txt-file included individual spectral reflections that differed from each other—the hyperspectral signature. Shooting an individual hyperspectral reflectance could include a damaged spot, a healthy part of the leaf or a part that was partially damaged/healthy. Every individual curve is unique, but in the cases of healthy leaves, the shapes of the curves were approximately the same (Figure 3a). In the cases of damaged leaves, the shapes of the curves demonstrated significant variation (Figure 3b). Therefore, a very large spread of spectrogram curves appears and a large error appears during averaging in the visible (VIS) range.

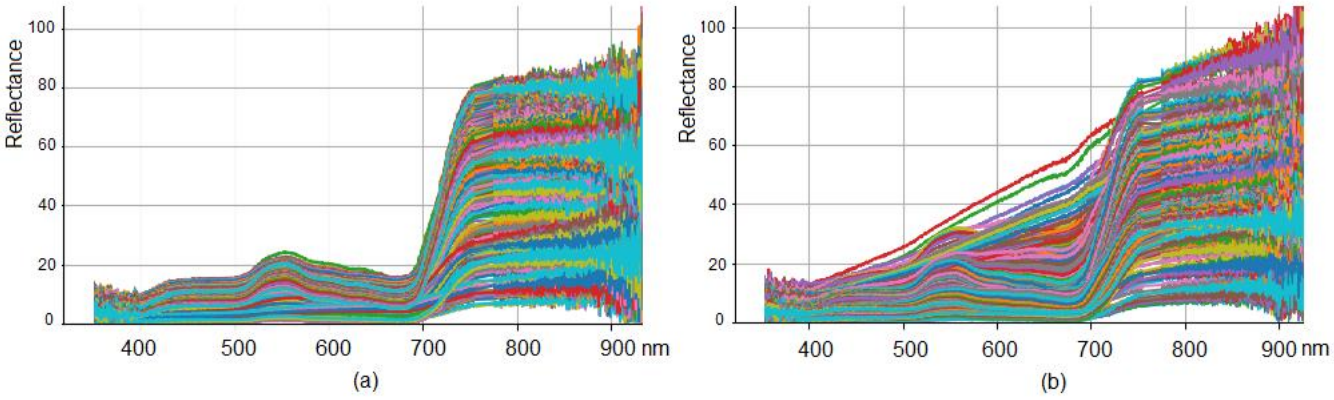


Figure 3. Hyperspectral curves of one dataset after preprocessing of raw data (each chart includes 2000 curves): (**a**) curves of healthy leaf, (**b**) curves of leaf in medium stage of Septoria disease.

For the determination of the curves obtained for healthy and damaged leaves, the data clusterization procedure was carried out, and twenty-four data clusters were formed for every dataset. The same clustering procedures were carried out for datasets with data on leaves and ears in the light, medium and severe stages of the disease. Clusters of curves of healthy and damaged plants were visually different according to the shape of the curves. After clustering, the selection of data for further analysis was carried out manually according to the correspondence of the curves' shapes and the number of curves in the clusters. To classify clusters, an approach based on two indicators was applied: (1) the shape of the curve and (2) the number of curves in every cluster. The shape of the curves was evaluated visually and was compared with previously published data, which provide examples of spectral brightness curves for healthy plants and for those affected by diseases [14,15,22,25,35,40]. The number of spectral curves in one dataset was 2000 or 4000. The number of curves in clusters ranged from 1 to 1680. Small clusters with fewer than 20 curves were excluded from the analysis. Uncharacteristic curves selected visually were excluded from the analysis. From each dataset with a volume of 2000 or 4000 curves, a combined cluster with a volume of at least 1000 spectral reflection curves was formed for further statistical analysis. After this classification, for every dataset, two clusters were formed: (1) cluster of typical curves and (2) cluster of invalid curves that were not reflecting the condition of leaves (Figure 4).

The approximate data ratio was as follows: typical curves were no less than 70% of the total primary volume of the dataset, and invalid curves ranged from 5 to 30%, depending on the dataset and the degree of the disease development. The most difficult classification of curves was for plants, especially ears, damaged by septoriosis in the middle degree.

Further distribution quantile analysis was carried out with relevant data curves, excluding the invalid ones.

At the beginning of the VIS range (wavelength 350 to 450 nm), a minimal spread in the values of the spectral brightness was observed at any stage of the development of the disease, i.e., the reflection in this range was similar for healthy and damaged plants. The biggest difference in average meanings was found in the red radiation spectrum on the wavelength of about 680 nm (Figure 5a,c,e). The change in the spectral brightness of reflection in the diapason of 550–680 nm was approximately −10% of the entire range of the reflection scale for healthy plants, approximately −2% for a medium lesion and +20% for a severe lesion. Additionally, in damaged and healthy plants, the difference in reflections in the red slope area (Red-Edge, 705–745 nm) was pronounced: the stronger the degree of disease infestation, the smaller the slope of the curves and the larger the total quantile range (from 1-quantile to 99-quantile)in this area (Figure 5b,d,f). At 700–705 nm, the total quantile range was 10% for healthy leaves, 26% for a medium lesion and 44% for a severe lesion. The same trend was observed for ears, an increase in the total quantile range of 35,

38 and 41%, respectively (Figure 6). Thus, the more pronounced the disease, the greater the total quantile range of values at the beginning of the red edge.

When comparing the ears with different degrees of lesion caused by Septoria, we observed higher variability in reflectance meanings at the beginning of the spectrum (wavelength 350–450 nm) compared to those observed for leaves (Figure 6). The total quantile range was also higher. This suggests that the ear itself is an object with a much more uneven coloration than the leaf. However, in general, the characteristics of the curves of the spectral reflection of the ears are similar in shape to the characteristics of the curves of the reflection of the leaves (Figures 5 and 6).

As a measure of the differences in the type of spectral brightness curves (hyperspectral signatures), the angle between the branches of the graph can be proposed (Figure 7). The left branch of the graph is a reflection in the VIS diapason of the spectrum. The reflection at wavelengths of 550 to 680 nm corresponds, respectively, to the maximum and minimum reflection of a healthy plant. The right branch of the graph is the reflection of the red edge and NIR diapason. For our investigation, the attention was focused on the red edge. The red edge slope is always more gentle in plants under stress [41]. Therefore, the angle between the branches of the graph for healthy and diseased plants differs significantly. Since the angle in degrees will depend on the scale of the axes, the numerical angle is represented in % of 100, where 100% is an angle of 180 degrees, or a straight line without a break. In healthy wheat plants, the average value of the angle between the branches of the graph was 59–65%. With a medium disease severity degree, the value was 76–78%, and with a high severity degree the value was 86–93%. When comparing spectral signatures, this pattern can be used to assess the degree of the STB disease.

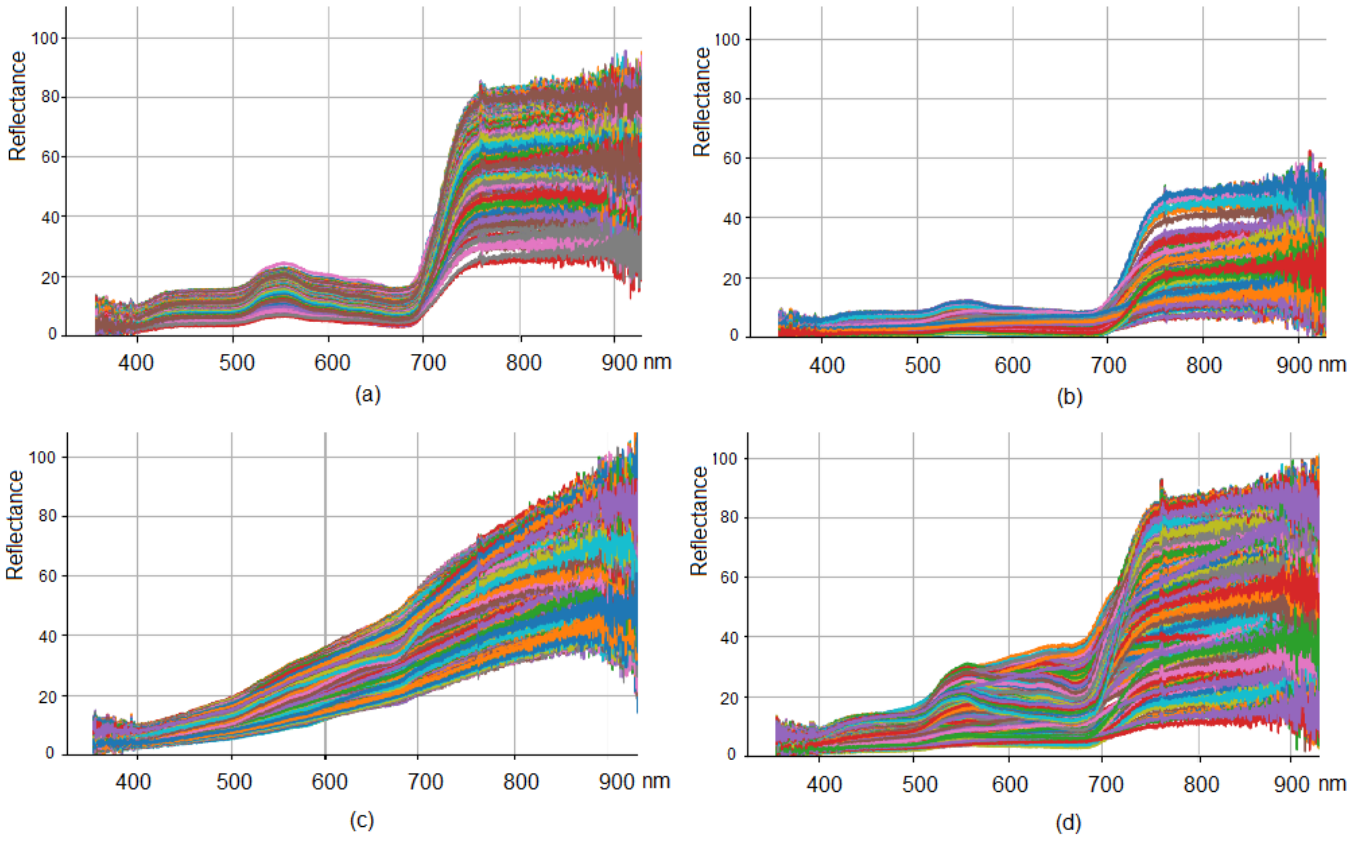


Figure 4. Two examples of curve dataset clusterization: (**a**) cluster of typical curves of healthy leaves; (**b**) cluster of invalid curves of healthy leaves; (**c**) cluster of typical curves of leaves with severe lesions caused by Septoria; (**d**) cluster of non-typical (invalid) curves of leaves with severe lesions caused by Septoria.

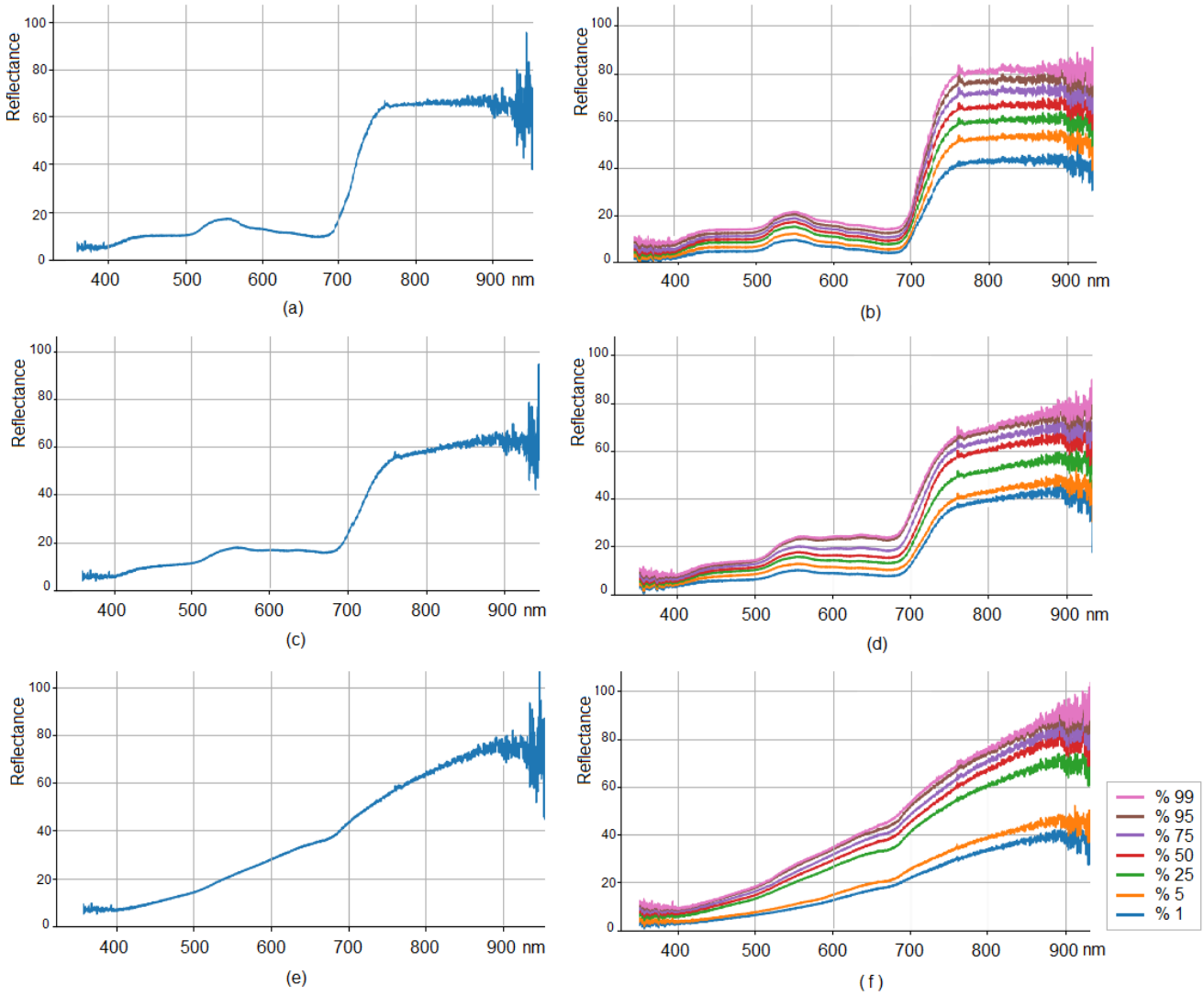


Figure 5. Datasets of means and quantiles of variety of hyperspectral curves of leaves: (**a**,**c**,**e**)—means of healthy, medium-damaged and severely damaged leaves, respectively; (**b**,**d**,**f**) quantile distribution of the same datasets.

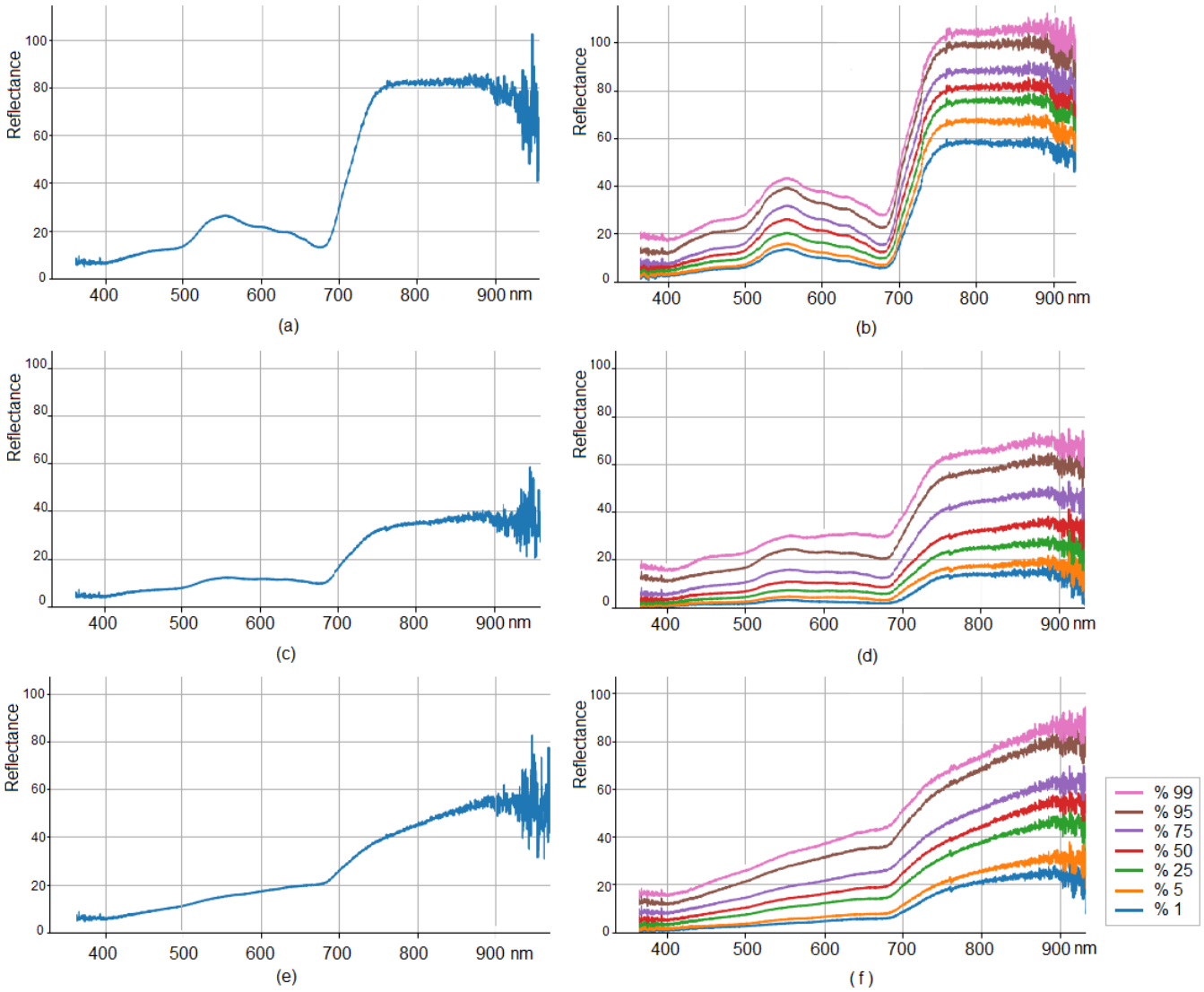


Figure 6. Dataset means and quantiles of variety of hyperspectral curves of ears: (**a**,**c**,**e**)—means of healthy, medium-damaged and severely damaged ears, respectively; (**b**,**d**,**f**) quantile distribution of the same datasets.

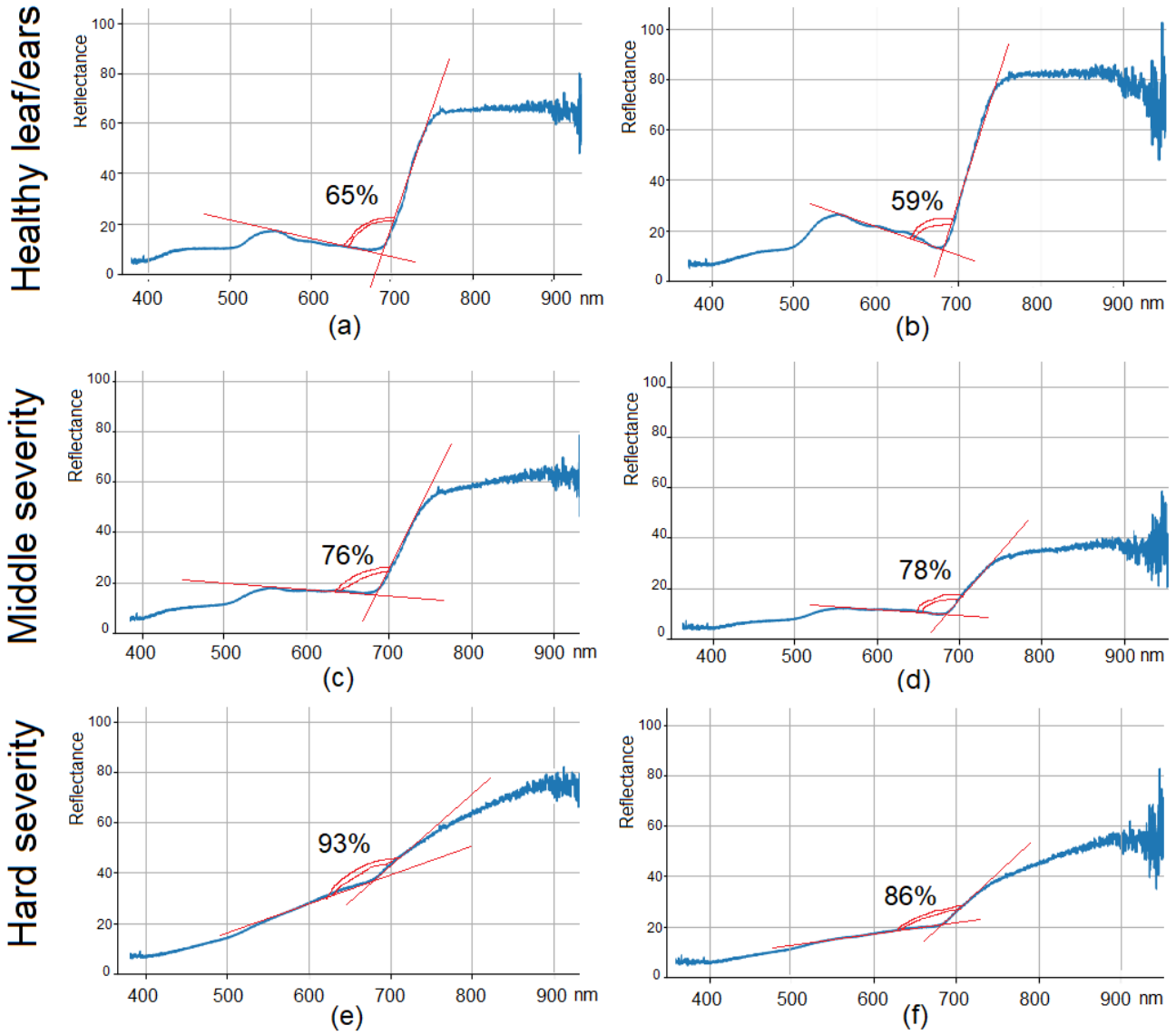


Figure 7. Hyperspectral curves of reflectance of leaves (**a**,**c**,**e**) and ears (**b**,**d**,**f**)—means of healthy, medium-damaged and severely damaged leaves and ears, respectively, and values of the angle (%) between the branches of the graph.

3.2. Classification of Diseased Plants by Non-Imaging Hyperspectral Signatures with Neural Network

The task of creating a neural network was to divide healthy and diseased wheat plants by spectral characteristics in the automatic mode. The first step was to recognize disease on leaves; the second step was to recognize STB severity on ears. The algorithm for creating a neural network was the same, but the training samples (datasets) for leaves and ears were different. Each dataset of leaves and ears contained 2000 or 4000 files, respectively. Every file included the numerical values of the spectral reflection curve for the range 338–1018 nm, with a spectral resolution step of 0.3 nm. Each of the 2000 files contained 2048 numeric values for a given individual curve. The algorithm for preparing primary raw data included (1) reducing the hyperspectral scale range from 338–1018 nm to 350–950 nm, (2) smoothing the data using the sliding window technique with a step (window) of 6 nm and (3) forming a new dataset after limiting the range boundaries and smoothing procedure. Limiting the range was necessary to reduce the influence of noise at the ends of the spectral brightness curve, since the largest measurement errors were in the region of less than 350 nm and more than 950 nm. The sliding window smoothing also

reduced the impact of individual outliers in the primary raw data. The size of the sliding window was chosen at 6 nm, so 16 consecutive measurements in 0.3 nm increments were placed into this window. Using a sliding window method, every 6 nm of the spectrum, the average mean of spectral brightness was calculated and recorded. When forming a new dataset, the raw primary information was minimized and optimized: from a set of 2048 numeric values of hyperspectral brightness with a step of 0.3 nm, a set of 100 values with a step of 6 nm was formed. There were still 2000 such sets in each dataset of leaves and 4000 sets for each dataset of ears.

The neural network was trained on two datasets of 2000 files which were converted from 2048 numeric to 100 consecutive values of the hyperspectral brightness. These datasets were the reference for the neural network training. The first training dataset corresponded to healthy plants, and the second contained data for plants severely damaged by septoriosis disease. Two independent neural networks were trained and tested for leaves and ears. The neural network had two layers of neurons. On the first layer there were five neurons, and the second layer contained two neurons. The number of training epochs for the neural network for leaves was 5000, and for ears it was 10,000. Each neuron on the first layer contained one dataset of 100 spectral brightness values in the total range of the investigated spectrum in increments of 6 nm. To train the network, data were fed to the neural network in the form of a training sample (80% of the entire dataset) and a test part (20% of the dataset). This ratio of 80 and 20% was formed randomly from a set of 2000 files for leaves or 4000 for ears. As a result of the training and testing of the neural network, a forecast of the operability of this network configuration was obtained: the expected number of correct answers was 100% with an average error of 27%. During subsequent testing on other datasets, when leaves and ears were affected by septoriosis, it was revealed that the neural network worked with higher accuracy on leaf materials than on ears. Thus, out of 100% of the total data of a sample of healthy leaves, 88.1% of the data was classified as healthy, 10.6% of the data was classified as damaged and about 1% of the data was not classified. The stronger the degree of damage to the leaves and ears, the higher the accuracy of the neural network in determining it. The most difficult task was classifying the medium stage of the disease manifestation. From the entire dataset, samples with an average lesion of 33 to 65% of all values could be classified as being both healthy and damaged (Table 1).

Table 1. Checking the results of the neural network determining the septoriosis diseases of wheat leaves and ears: the average value of the feature definition in % with a confidence interval of the average (detection percentage).

	Defined with Neural Network			**Cohen's Kappa**
	Healthy, %	**Damaged, %**	**Uncertain, %**	
Nominal scale				
Leaves				
Healthy leaves	88.1 ± 2.4	10.6 ± 2.5	1.1 ± 0.1	0.90
Light lesion of Septoria	84.0 ± 5.5	10.8 ± 6.1	5.0 ± 0.1	0.64
Medium lesion of Septoria	60.5 ± 5.6	38.3 ± 5.9	1.2 ± 0.3	0.13
Severe lesion of Septoria	7.1 ± 2.9	92.7 ± 3.1	0.1 ± 0.1	0.98
Dead leaves	2.3 ± 3.0	97.7 ± 3.2	0.1 ± 0.1	0.99
Ears				
Healthy ears	94.9 ± 4.8	4.9 ± 4.7	0.1 ± 0.1	0.96
Light lesion of Septoria	70.2 ± 6.5	29.2 ± 6.3	0.6 ± 0.2	0.40
Medium lesion of Septoria	33.3 ± 7.6	65.6 ± 7.9	1.2 ± 0.2	0.24
Severe lesion of Septoria	0.6 ± 1.0	99.4 ± 1.0	0.0 ± 0.0	0.99

The quality of disease identification by the neural network was good. In the automated mode, the precision of determining diseased/healthy leaves ranged from 88.1 to 97.7% for different datasets. It was more difficult to identify the presence of the disease on wheat ear. Only severely damaged plants were precisely identified (up to 99.4%). Light and medium

stages of septoriosis disease were classified with a higher error. This is due to the uneven ear surface. A comparison of visual and automatic (PNN) assessment was carried out according to Cohen's kappa for each gradation of the disease (Table 1, column Cohen's kappa). A strong coincidence of estimates for healthy and STB&SNB highly damaged plants was observed, and the Kappa coefficient was 0.90–0.98. For light and medium degrees of the disease, the visual and automatic assessments coincided less, and the Kappa coefficient was 0.13–0.64. The average value of Cohen's kappa together, according to all of the data, was 0.69.

3.3. Estimation of Yield Losses Due to Septoriosis

The biological yields of every wheat cultivar were determined during harvesting. Yield losses were recorded for different degrees of STB&SNB lesion. A strong positive relationship was shown between the lesion degree and yield losses (Figure 8). Yield losses from disease were described using the logarithmic equation. For evaluation according to leaves, the relationship was stronger than for evaluation according to ear: $R^2 = 0.7993$ and 0.4941, respectively ($p < 0.05$). A severe lesion caused by STB on flag-leaf (more than 70%) led to 35–45% yield losses. A severe lesion caused by SNB on ears (more than 70%) led to 20–40% yield losses.

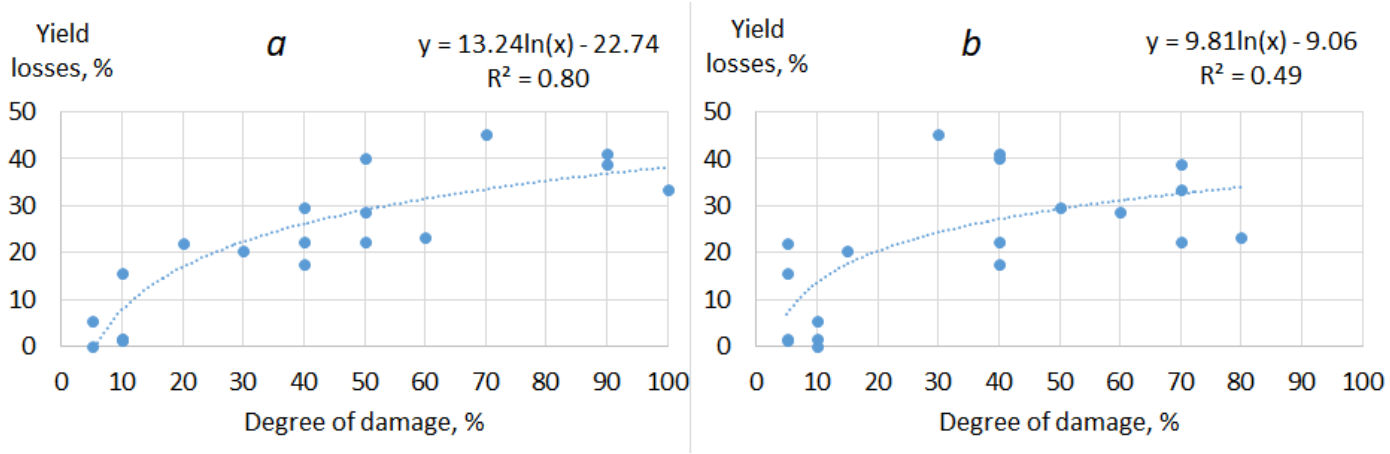


Figure 8. Regression relationship between the degree of septoriosis lesion and yield loss: (**a**) STB lesion of flag-leaf; (**b**) SNB lesion of ears. Every circle is an individual measurement from the plot with studied wheat cultivar.

The loss of biological yield is usually associated with the loss of the photosynthetic surface area of the leaves and ears under the influence of the disease. The main reason for crop loss in this case was a decrease in the mass of 1000 grains.

4. Discussion

The identification of septoriosis symptoms in the beginning of the vegetation season is difficult because of a long latent period. In the second half of the growing season, at the beginning of the wheat maturation phase, the lesion of leaves and ears becomes obvious; however, at this time, it is too late to protect wheat crops. However, it is possible to predict yield losses due to STB by the amount of affected plants and the degree of their lesion. Under septoriosis, plant tissues with spots necrotize and die prematurely, so the leaf surface area decreases and the active growing season becomes shorter. With a degree of leaf damage of 30%, the yield is reduced by 10%; with leaf damage of 51–75%, the yield is reduced by 30%; and when leaf and ear damages come to 75% and more, the yield is reduced by 40% and more [10]. The disease has the strongest effect on grain size and on its quality, which is expressed by a decrease in gluten [35]. With an increase in the lesion of the ear by 10%, the gluten content decreases by 2.5% [42]. With the chemical parameters of the grain change, in particular, the content of protein nitrogen decreases. The sowing qualities of seeds, such

as germination energy and field germination, are noticeably degraded. Different wheat cultivars could demonstrate susceptibility or resistance to STB, and the use of cultivar mixtures may help to reduce the damage caused by this pathogen [43–46]. It has also been shown that the mass of 1000 grains becomes lower with the development of ear disease caused by STB [35].

If the investigation is carried out for the collection of crops of different varieties, it could allow for the assessment of the cultivar resistance to Septoria tritichi blotch. Visual, manual and laboratory assessment of STB development is time-consuming, even for an experienced expert in phytopathology [35,46–48]. Therefore, the use of optical sensors, including hyperspectral non-imaging cameras, could be useful and time-saving for assessing the state of cereal crops.

The main difficulty in our study was the processing and interpretation of field hyperspectral survey data for Sertoriosis lesion identification and the comparison of the degrees of lesion. In the field, the development of the disease is usually estimated via points of marking damage scale or percentages [11]. The results of hyperspectral imaging could not be presented in points; instead, it was necessary to compare the shape of the spectral brightness curves (signatures) for different degrees of lesion. The raw data, as the basis for curve creation, had a large range of values; so, a two-step procedure for preparing the initial data was required, including automatic clustering and the manual culling of clusters with an unsuitable shape of curves. Based on the results of data preprocessing and processing, curves characteristic of different degrees of the disease were constructed. Thus, for diseased plants, the spectral reflection in the visible range (350–690 nm) was approximately 20% higher than that of healthy plants. In addition, the spectral reflection curves of the diseased plants had a flat shape in the region of the red edge (705–745 nm). As a measure of STB&SNB severity on green plants, we suggest using values of the angle (%) between the branches of the hyperspectral reflection curve graph.

Our results are consistent with a previous study on STB and SNB detection using hyperspectral measurements, when a 350–1150 nm wavelength range was used, and several spectral vegetation indices (calculated from the reflectance measured for two or more wavelengths) were indicative of the disease development on the 15th–17th dai [27]. These spectral vegetation indices were related to the different levels of chlorophyll, anthocyanin and carotenoid content in the infected leaves, and the variation in these levels could be detected in the visible range. The changes in the reflections in the NIR range were also proven to be indicative of various plant stresses and diseases [49–53].

However, there are other options for STB&SNB detection using hyperspectral measurements. In the study of Iori et al., where a spectrograph worked in a range from 1000 to 1700 nm, the variation in reflectance on the wavelength 1650 nm was proclaimed as being indicative of Septoria nodorum blotch, even in the early stages (1st–3rd dai) of disease development [35].

The neural network could successfully recognize the degree of STB&SNB disease on the leaves and ears in the pre-harvesting period. Hyperspectral survey raw data were loaded into the neural network for both training and testing. The results of determination of STB&SNB via the perceptron neural network are comparable with the nominal scale of the STB&SNB disease. Thus, the created neural network based on the processing of raw hyperspectral survey data successfully coped with the assessment of the presence of STB on wheat leaves and SNB on ears in the late stage.

These studies were carried out at the level of individual plants and at the level of an experimental plot. For food production crops, studies should be conducted at the level of the crop cover area (a single field or a group of fields). This can be useful for predicting crop yield losses in the fields [54,55]. In this case, the use of a manual spectrometer is not applicable, and it is necessary to use remote assessment methods using an unmanned aerial vehicle or satellite imagery.

5. Conclusions

This study suggests an automatic approach to detect the degree of STB&SNB disease damage on spring wheat in the pre-harvesting period in field conditions using hyperspectral non-imaging measurements and the trained perceptron neural network. With all of the possibilities of modern data processing, the results largely depend on the quality of the input information. Septoriosis is a difficult-to-identify disease in the early stages of development. Therefore, for our research, we chose a rather late detection period in the field in the pre-harvesting stage of wheat growth, when the disease could be unmistakably determined visually, and the training dataset was relevant. The proposed perceptron neural network model successfully coped with the identification of healthy leaves and ears, and also determined the % of occurrence of affected leaves and ears. The results were consistent with the visual expert assessment of the development of the disease; the average value of Cohen's kappa was 0.69 (substantial agreement). In the case of healthy plants and a high degree of STB&SNB severity, Cohen's kappa reached 0.90–0.98.

Looking ahead, further research is needed for STB severity degree estimation in different cultivars with varying degrees of resistance to STB. Hyperspectral reflection measurements of healthy/diseased plants in the field condition and with machine learning algorithms are prospective for the determination of disease.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/agronomy13041045/s1: File S1: A list of links to zip-archived raw hyperspectral data files.

Author Contributions: Conceptualization, S.V.Z., V.E.V., E.V.S. and A.P.G.; methodology, S.V.Z. and V.E.V.; software, A.V.S. and T.Z; validation, S.V.Z., A.V.S. and T.S.Z.; formal analysis and investigation (field measurement), S.V.Z., V.E.V. and M.A.V.; resources (wheat collection and inoculation), E.V.P.; data curation, S.V.Z. and E.V.S.; writing—original draft preparation, S.V.Z. and A.D.Z.; writing—review and editing, S.V.Z. and A.D.Z.; visualization, S.V.Z. and A.V.S.; supervision, A.P.G.; project administration, S.V.Z., E.V.S. and A.P.G.; funding acquisition, S.V.Z., E.V.S. and A.P.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Science and Higher Education of the Russian Federation, project № 075-15-2021-1409.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The strains of pathogens cultivated in the core facility center "State Collection of Phytopathogenic Microorganisms" of the Federal State Budgetary Scientific Institution All-Russian Scientific Research Institute of Phytopathology (VNIIF) were used in this research.

Conflicts of Interest: On behalf of all of the authors, the corresponding author states that there are no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript or in the decision to publish the results.

References

1. Suffert, F.; Sache, I.; Lannou, C. Early stages of septoria tritici blotch epidemics of winter wheat: Build-up, overseasoning, and release of primary inoculum. *Plant Pathol.* **2011**, *60*, 166–177. [CrossRef]
2. Fones, H.; Gurr, S. The impact of Septoria tritici Blotch disease on wheat: An EU perspective. *Fungal Genet. Biol.* **2015**, *79*, 3–7. [CrossRef] [PubMed]
3. Chedli, R.B.H.; M'barek, S.B.; Yahyaoui, A.; Kehel, Z.; Rezgui, S. Occurrence of Septoria tritici blotch (*Zymoseptoria tritici*) disease on durum wheat, triticale, and bread wheat in northern Tunisia. *Chil. J. Agric. Res.* **2018**, *78*, 559–568. [CrossRef]
4. Turgay, E.B.; Büyük, O.; Yıldırım, A.F.; Ölmez, F.; Akci, N.; Karakaya, A. Pathotype diversity of *Zymoseptoria tritici* (Desm. Quaedvlieg & Crous) isolates collected from Central Anatolia, Turkey. *J. Phytopathol.* **2022**, *170*, 588–597. [CrossRef]
5. Ponomarenko, A.; Goodwin, S.B.; Kema, G. Septoria tritici blotch (STB) of wheat. *Plant Health Instr.* **2011**. [CrossRef]

6. McDonald, M.C.; Renkin, M.; Spackman, M.; Orchard, B.; Croll, D.; Solomon, P.S.; Milgate, A. Rapid parallel evolution of azole fungicide resistance in Australian populations of the Wheat Pathogen *Zymoseptoria tritici*. *Appl. Environ. Microbiol.* **2019**, *85*, e01908-18. [CrossRef]
7. O'Driscoll, A.; Doohan, F.; Mullins, E. Exploring the utility of Brachypodium distachyon as a model pathosystem for the wheat pathogen *Zymoseptoria tritici* Plant Biology. *BMC Res. Notes* **2015**, *8*, 132. [CrossRef]
8. Sanin, S.S.; Korneva, L.G.; Polyakova, T.M. Prognosis of the risk of epiphytotic septoria of wheat leaves and ears. *Plant Prot. Quar.* **2015**, 33–36.
9. Kononenko, O.S.; Shishkin, N.V.; Derova, T.G. Winter wheat septoria blotch (*Zymoseptoria tritici*) (literature review). *Grain Econ. Russ.* **2021**, *6*, 104–108. [CrossRef]
10. Sanin, S.S.; Zhohova, T.P. The influence of diseases and plant protection products on the quality of wheat grain. Защита и карантин растений **2012**, *11*, 16–20. (In Russian)
11. Bock, C.H.; Barbedo, J.G.A.; Del Ponte, E.M.; Bohnenkamp, D.; Mahlein, A.-K. From visual estimates to fully automated sensor-based measurements of plant disease severity: Status and challenges for improving accuracy. *Phytopathol. Res.* **2020**, *2*, 9. [CrossRef]
12. Hughes, D.P.; Salathe, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv* **2015**, arXiv:1511.08060.
13. Esgario, J.G.M.; Krohling, R.A.; Ventura, J.A. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Comput. Electron. Agric.* **2020**, *169*, 105162. [CrossRef]
14. Mahlein, A.K.; Steiner, U.; Hillnhütter, C.; Dehne, H.W.; Oerke, E.C. Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant Methods* **2012**, *8*, 3. [CrossRef] [PubMed]
15. Thomas, S.; Kuska, M.T.; Bohnenkamp, D.; Brugger, A.; Alisaac, E.; Wahabzada, M.; Behmann, J.; Mahlein, A.K. Benefits of hyperspectral imaging for plant disease detection and plant protection: A technical perspective. *J. Plant Dis. Prot.* **2018**, *125*, 5–20. [CrossRef]
16. Abdulridha, J.; Ehsani, R.; De Castro, A. Detection and differentiation between laurel wilt disease, phytophthora disease, and salinity damage using a hyperspectral sensing technique. *Agriculture* **2016**, *6*, 56. [CrossRef]
17. Wang, D.; Vinson, R.; Holmes, M.; Seibel, G.; Bechar, A.; Nof, S.; Tao, Y. Early Detection of Tomato Spotted Wilt Virus by Hyperspectral Imaging and Outlier Removal Auxiliary Classifier Generative Adversarial Nets (OR-AC-GAN). *Sci. Rep.* **2019**, *9*, 1–14. [CrossRef] [PubMed]
18. Lowe, A.; Harrison, N.; French, A.P. Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods* **2017**, *13*, 80. [CrossRef]
19. Bebronne, R.; Michez, A.; Leemans, V.; Vermeulen, P.; Dumont, B.; Mercatoris, B. Characterisation of fungal diseases on winter wheat crop using proximal and remote multispectral imaging. In *Proceedings of the Precision Agriculture'19*; Wageningen Academic Publishers: Wageningen, The Netherlands, 2019; pp. 255–261.
20. Amara, J.; Bouaziz, B.; Algergawy, A. A deep learning-based approach for banana leaf diseases classification. *Lect. Notes Inform. (LNI), Proc. Ser. Ges. Fur Inform.* **2017**, *266*, 79–88.
21. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using deep learning for image-based plant disease detection. *Front. Plant Sci.* **2016**, *7*, 1419. [CrossRef]
22. Wan, L.; Li, H.; Li, C.; Wang, A.; Yang, Y.; Wang, P. Hyperspectral Sensing of Plant Diseases: Principle and Methods. *Agronomy* **2022**, *12*, 1451. [CrossRef]
23. Boulent, J.; Foucher, S.; Théau, J.; St-Charles, P.L. Convolutional Neural Networks for the Automatic Identification of Plant Diseases. *Front. Plant Sci.* **2019**, *10*, 941. [CrossRef]
24. Bhupathi, P.; Sevugan, P. Application of Hyperspectral Remote Sensing Technology for Plant Disease Forecasting: An Applied Review. *Ann. Rom. Soc. Cell Biol.* **2021**, *25*, 4555–4566.
25. Mahlein, A.K.; Kuska, M.T.; Thomas, S.; Bohnenkamp, D.; Alisaac, E.; Behmann, J.; Wahabzada, M.; Kersting, K. Plant disease detection by hyperspectral imaging: From the lab to the field. *Adv. Anim. Biosci.* **2017**, *8*, 238–243. [CrossRef]
26. Whetton, R.L.; Waine, T.W.; Mouazen, A.M. Hyperspectral measurements of yellow rust and fusarium head blight in cereal crops: Part 2: On-line field measurement. *Biosyst. Eng.* **2018**, *167*, 144–158. [CrossRef]
27. Odilbekov, F.; Armoniené, R.; Henriksson, T.; Chawade, A. Proximal phenotyping and machine learning methods to identify septoria tritici blotch disease symptoms in wheat. *Front. Plant Sci.* **2018**, *9*, 685. [CrossRef] [PubMed]
28. Yu, K.; Anderegg, J.; Mikaberidze, A.; Karisto, P.; Mascher, F.; McDonald, B.A.; Walter, A.; Hund, A. Hyperspectral canopy sensing of wheat septoria tritici blotch disease. *Front. Plant Sci.* **2018**, *9*, 1195. [CrossRef] [PubMed]
29. Anderegg, J.; Hund, A.; Karisto, P.; Mikaberidze, A. In-Field Detection and Quantification of Septoria Tritici Blotch in Diverse Wheat Germplasm Using Spectral–Temporal Features. *Front. Plant Sci.* **2019**, *10*, 1355. [CrossRef]
30. Cheshkova, A.F. A review of hyperspectral image analysis techniques for plant disease detection and identification. *Vavilovskii Zhurnal Genet. Sel.* **2022**, *26*, 202–213. [CrossRef]
31. Doktor, D.; Brosinsky, A.; Lausch, A.; Merbach, I. Comparison of imaging and non-imaging hyperspectral sensors observing vegetation in the laboratory. *Geophys. Res. Abstr.* **2011**, *13*, 9251.
32. Bohnenkamp, D.; Behmann, J.; Paulus, S.; Steiner, U.; Mahlein, A.K. A Hyperspectral Library of Foliar Diseases of Wheat. *Phytopathology* **2021**, *111*, 1583–1593. [CrossRef] [PubMed]

33. Paulus, S.; Mahlein, A.K. Technical workflows for hyperspectral plant image assessment and processing on the greenhouse and laboratory scale. *Gigascience* **2020**, *9*, giaa090. [CrossRef] [PubMed]
34. Dammer, K.H. Proof of concept study: A novel mobile in-canopy imaging system for detecting symptoms of fungal diseases in cereals. *J. Plant Dis. Prot.* **2022**, *129*, 769–773. [CrossRef]
35. Iori, A.; Scala, V.; Cesare, D.; Pinzari, F.; D'Egidio, M.G.; Fanelli, C.; Fabbri, A.A.; Reverberi, M.; Serranti, S. Hyperspectral and molecular analysis of Stagonospora nodorum blotch disease in durum wheat. *Eur. J. Plant Pathol.* **2015**, *141*, 689–702. [CrossRef]
36. Singh, V.; Sharma, N.; Singh, S. A review of imaging techniques for plant disease detection. *Artif. Intell. Agric.* **2020**, *4*, 229–242. [CrossRef]
37. Meier, U.; Bleiholder, H.; Buhr, L.; Feller, C.; Hack, H.; Heß, M.; Lancashire, P.; Schnock, U.; Stauß, R.; Van den Boom, T.; et al. The BBCH system to coding the phenological growth stages of plants-history and publications. *J. Für Kult.* **2009**, *61*, 41–52. [CrossRef]
38. Ocean Optics, Ocean View Installation & Operation Manual United States America, 2016. Available online: https://www.oceaninsight.com/globalassets/catalog-blocks-and-images/manuals--instruction-ocean-optics/software/oceanviewio.pdf (accessed on 6 November 2022).
39. Perceptron Neural Networks Description. Available online: https://www.mathworks.com/help/deeplearning/ug/perceptron-neural-networks.html (accessed on 24 February 2023).
40. Bohnenkamp, D.; Behmann, J.; Mahlein, A.K. In-field detection of Yellow Rust in Wheat on the Ground Canopy and UAV Scale. *Remote Sens.* **2019**, *11*, 2495. [CrossRef]
41. Golhani, K.; Balasundram, S.K.; Vadamalai, G.; Pradhan, B. Red-Edge indices to diagnose orange spotting disease of oil palm in Malaysia. In Proceedings of the 38th Asian Conference on Remote Sensing - Space Applications: Touching Human Lives, ACRS 2017, New Delhi, India, 23–27 October 2017.
42. Kristoffersen, R.; Eriksen, L.B.; Nielsen, G.C.; Jørgensen, J.R.; Jørgensen, L.N. Management of Septoria Tritici Blotch Using Cultivar Mixtures. *Plant Dis.* **2022**, *106*, 1341–1349. [CrossRef]
43. Castro, A.C.; Simón, M.R. Effect of tolerance to Septoria tritici blotch on grain yield, yield components and grain quality in Argentinean wheat cultivars. *Crop Prot.* **2016**, *90*, 66–76. [CrossRef]
44. Ficke, A.; Cowger, C.; Bergstrom, G.; Brodal, G. Understanding yield loss and pathogen biology to improve disease management: Septoria nodorum blotch—A case study in wheat. *Plant Dis.* **2018**, *102*, 696–707. [CrossRef]
45. Ghaffary, S.M.T.; Chawade, A.; Singh, P.K. Practical breeding strategies to improve resistance to Septoria tritici blotch of wheat. *Euphytica* **2018**, *214*, 122. [CrossRef]
46. Cunfer, B.M. Stagonospora and Septoria diseases of barley, oat, and rye. *Can. J. Plant Pathol.* **2000**, *22*, 332–348. [CrossRef]
47. Simko, I.; Jimenez-Berni, J.A.; Sirault, X.R.R. Phenomic approaches and tools for phytopathologists. *Phytopathology* **2017**, *107*, 6–17. [CrossRef]
48. Mahlein, A.K.; Kuska, M.T.; Thomas, S.; Wahabzada, M.; Behmann, J.; Rascher, U.; Kersting, K. Quantitative and qualitative phenotyping of disease resistance of crops by hyperspectral sensors: Seamless interlocking of phytopathology, sensors, and machine learning is needed! *Curr. Opin. Plant Biol.* **2019**, *50*, 156–162. [CrossRef]
49. Zahir, S.A.D.M.; Omar, A.F.; Jamlos, M.F.; Azmi, M.A.M.; Muncan, J. A review of visible and near-infrared (Vis-NIR) spectroscopy application in plant stress detection. *Sens. Actuators A Phys.* **2022**, *338*, 113468. [CrossRef]
50. Ryckewaert, M.; Héran, D.; Simonneau, T.; Abdelghafour, F.; Boulord, R.; Saurin, N.; Moura, D.; Mas-Garcia, S.; Bendoula, R. Physiological variable predictions using VIS–NIR spectroscopy for water stress detection on grapevine: Interest in combining climate data using multiblock method. *Comput. Electron. Agric.* **2022**, *197*, 106973. [CrossRef]
51. Kremneva, O.Y.; Danilov, R.Y.; Sereda, I.I.; Tutubalina, O.V.; Pachkin, A.A.; Zimin, M. V Spectral characteristics of winter wheat varieties depending on the development degree of Pyrenophora tritici-repentis. *Precis. Agric.* **2022**. [CrossRef]
52. Wegulo, S.N.; Breathnach, J.A.; Baenziger, P.S. Effect of growth stage on the relationship between tan spot and spot blotch severity and yield in winter wheat. *Crop Prot.* **2009**, *28*, 696–702. [CrossRef]
53. Pierre, B.; Marie-Odile, B.; François, C.; David, G. Identifying traits leading to tolerance of wheat to Septoria tritici blotch. *Field Crop Res.* **2015**, *180*, 176–185. [CrossRef]
54. Jevtić, R.; Župunski, V.; Lalošević, M.; Župunski, L. Predicting potential winter wheat yield losses caused by multiple disease systems and climatic conditions. *Crop Prot.* **2017**, *99*, 17–25. [CrossRef]
55. Budka, A.; Łacka, A.; Gaj, R.; Jajor, E.; Korbas, M. Predicting winter wheat yields by comparing regression equations. *Crop Prot.* **2015**, *78*, 84–91. [CrossRef]

agronomy

Article

Citrus Disease Image Generation and Classification Based on Improved FastGAN and EfficientNet-B5

Qiufang Dai [1,2,3], Yuanhang Guo [1,2], Zhen Li [1,2,3,*], Shuran Song [1,2,3], Shilei Lyu [1,2,3], Daozong Sun [1,2,3], Yuan Wang [1,2] and Ziwei Chen [1,2]

[1] College of Electronic Engineering (College of Artificial Intelligence), South China Agricultural University, Guangzhou 510642, China
[2] Division of Citrus Machinery, China Agriculture Research System, Guangzhou 510642, China
[3] Guangdong Engineering Research Center for Monitoring Agricultural Information, Guangzhou 510642, China
* Correspondence: lizhen@scau.edu.cn

Abstract: The rapid and accurate identification of citrus leaf diseases is crucial for the sustainable development of the citrus industry. Because citrus leaf disease samples are small, unevenly distributed, and difficult to collect, we redesigned the generator structure of FastGAN and added small batch standard deviations to the discriminator to produce an enhanced model called FastGAN2, which was used for generating citrus disease and nutritional deficiency (zinc and magnesium deficiency) images. The performance of the existing model degrades significantly when the training and test data exhibit large differences in appearance or originate from different regions. To solve this problem, we propose an EfficientNet-B5 network incorporating adaptive angular margin (Arcface) loss with the adversarial weight perturbation mechanism, and we call it EfficientNet-B5-pro. The FastGAN2 network can be trained using only 50 images. The Fréchet Inception Distance (FID) and Kernel Inception Distance (KID) are improved by 31.8% and 59.86%, respectively, compared to the original FastGAN network; 8000 images were generated using the FastGAN2 network (2000 black star disease, 2000 canker disease, 2000 healthy, 2000 deficiency). Only images generated by the FastGAN2 network were used as the training set to train the ten classification networks. Real images, which were not used to train the FastGAN2 network, were used as the test set. The average accuracy rates of the ten classification networks exceeded 93%. The accuracy, precision, recall, and F1 scores achieved by EfficientNet-B5-pro were 97.04%, 97.32%, 96.96%, and 97.09%, respectively, and they were 2.26%, 1.19%, 1.98%, and 1.86% higher than those of EfficientNet-B5, respectively. The classification network model can be successfully trained using only the images generated by FastGAN2, and EfficientNet-B5-pro has good generalization and robustness. The method used in this study can be an effective tool for citrus disease and nutritional deficiency image classification using a small number of samples.

Keywords: citrus diseases; generative adversarial network; classification network; FastGAN; EfficientNet

Citation: Dai, Q.; Guo, Y.; Li, Z.; Song, S.; Lyu, S.; Sun, D.; Wang, Y.; Chen, Z. Citrus Disease Image Generation and Classification Based on Improved FastGAN and EfficientNet-B5. *Agronomy* **2023**, *13*, 988. https://doi.org/10.3390/agronomy13040988

Academic Editors: Jian Zhang, Randy G. Goebel and Zhihai Wu

Received: 1 March 2023
Revised: 23 March 2023
Accepted: 24 March 2023
Published: 27 March 2023

1. Introduction

Citrus is one of the most popular fruits in the world. However, for a long time, diseases have seriously threatened the growth of citrus. They affect the yield and quality of citrus crops and have a significant impact on the citrus industry. In severe cases, the diseases can lead to the death of an entire citrus patch. Therefore, improving citrus disease prevention technology to effectively prevent the appearance of diseases and produce citrus fruit trees with high yield and quality is crucial. The traditional identification of citrus pests and diseases relies on the naked eye, which takes a long time and requires certain professional knowledge. An intelligent, low-cost, and highly accurate method for citrus disease identification would be more practical [1] (see Abbreviations Section).

In recent years, deep learning technology has been widely used in the field of agriculture, such as in fruit classification and grading [2–4], automatic picking, and the diagnosis

of diseases and pests [5–8], and the automatic diagnosis of plant diseases is one of the most active research areas in agriculture. Several deep learning-based techniques for automatic plant disease diagnosis have emerged, which can help farmers reduce the economic losses caused by pests and diseases in farming [9]. In crops, disease symptoms often appear on the leaves; therefore, crop diseases can be automatically detected by applying machine learning techniques to leaf images. For example, Zhang et al. [10] segmented diseased leaf images using K-means clustering and extracted shape and color features from the lesion information. They classified seven cucumber diseases using sparse representation with a presidential recognition rate of 85.7%. Liu et al. [11] proposed a WSRD-Net method for wheat stripe rust detection based on a convolutional neural network (CNN), which can obtain 60.8% average precision (AP) and 73.8% recall rate on a wheat stripe rust dataset. Zhong et al. [12] proposed a three regression, multi-label classification and focal loss function methods based on the DenseNet-121 deep convolutional network to identify apple leaf diseases with over 93% accuracy on 2464 images, including six apple leaf diseases.

Yao et al. [13] used an improved Xception network to classify brown spots and anthracnose of peach, eventually obtaining a 98.85% accuracy [14]. Janarthan et al. [15] proposed a lightweight, fast, and accurate deep metric learning-based architecture for detecting citrus diseases from sparse data to obtain 95.04% detection accuracy. Deep learning requires many datasets to support the model training. Otherwise, overfitting may occur [16]. The main obstacle to using machine learning in agriculture is the small dataset and the limited number of annotated samples. This becomes more evident when supervised machine learning algorithms that require labeled data are used. The collection of a large amount of plant-disease-related data may have the problem of an uneven distribution of samples. Some diseases may have a small number of samples, which is not enough to train a classification network. Although some public datasets are available, the size of the datasets and categories do not meet the requirements of all applications. Using simple data enhancement methods such as random inversion, deep random flip, increasing contrast, and adding noise [17] can suppress overfitting, but the sample data are still not sufficiently rich, and the image features are less differentiated from the original dataset. Goodfellow et al. [18] proposed a generative adversarial network (GAN) using generators and discriminators against each other. GAN is widely used in the field of computer vision, such as for image super-resolution reconstruction and image defogging [19,20], and can also be used as a data enhancement tool to expand datasets [21]. Using generated images introduces more variability, which can improve the training process of classification networks and increase accuracy.

Ma et al. [22] generated blood cell images using a DC-GAN network to increase data samples and eliminate data imbalance and missing data labels. Cap et al. [23] proposed a LeafGAN by improving CycleGAN using paired datasets to successfully transform healthy leaves into diseased leaves. Xiao et al. [24] successfully generated six types of citrus leaf images using TRL-GAN, an enhanced version of CycleGAN that removes the real scene background from the original images using Mask RCNN. They achieved a 97.45% accuracy on ResNeXt101 after expanding the original dataset using the generated images. However, expanding datasets with adversarial networks increases training time, mostly to several days, and generates low-quality images. The resolution of the generated images is often below 512×512, which cannot retain more details, and the expanded dataset has limited performance improvement for the classification network.

Karras et al. [25] proposed StyleGAN2 based on StyleGAN. The StyleGAN is a current high-performance, high-resolution image generation framework capable of generating very high-quality images on a wide range of datasets but still requires a large dataset as well as high computational resources and training time [26]. Liu et al. [27] proposed the FastGAN network, which can finish training a complete model in a dozen hours on a single RTX-2080 GPU, by improving StyleGAN2. However, when applied to plant disease sample generation, it produces checkerboard artifacts, loss of details, and insufficiently rich sample data. The performance of the classification model degrades significantly when the

training and test data are very different in appearance or originate from different regions, for example, the light of the target in the training data is very strong while the light of the target in the test data is very dark, the image acquisition devices are different, the geographical locations where they were taken are different, and so on.

Mohanty et al. [28] used 54,306 healthy and diseased leaf images from the PlantVillage dataset to train a neural network model for the identification of 26 leaf species and obtained an accuracy of 99.3% after cross-validation evaluation. The performance of the model decreased to approximately 31% when tested with a set of plant images taken in the field because the training set of the model was taken in a laboratory environment, and its images had a uniform background. Ferentinos [29] also noted that when the model was trained on images taken in a laboratory environment and tested on images taken in a planting environment, the accuracy of the model decreased from 99.5% to approximately 33%. Therefore, changes in background and shooting conditions can have a serious impact on the performance of the model.

Because the background of the FastrGAN2-generated images is not as rich as that of the real captured images, the FastrGAN2-generated images and captured images can be regarded as coming from different regions. Because our experiments only used the FastrGAN2-generated images as the training set to train the model and the real captured images to test the performance of the model, this poses a classification network performance challenge and requires the classification network to have a high generalization capability.

Here, we propose the FastGAN2 network, which overcomes the checkerboard artifact problem of the FastGAN network, improves the quality of generated images, and enhances the diversity of the generated images for small datasets. We used the generated images only as the training set of the classification network and tested it using images taken but not used for training with the FastGAN2 network. Finally, we tested it on Densenet121, ResNet50, ShuffleNetv2 [30], Mlp-Mixer [31], MobileNetv3 [32], Vision Transformer, Swin Transformer, EfficientNet-B3, EfficientNet-B5 [33], and EfficientNet-B5-pro and achieved an average accuracy of 93.52%. To improve the generalization of the model, this paper proposes an EfficientNet-B5-pro network based on EfficientNet-B5 that uses the adaptive angular margin (Arcface) loss with adversarial weight perturbation (AWP) mechanism. It achieved the highest performance compared to ten classification networks. The main contributions of this study are as follows:

By redesigning the FastGAN network generator structure and adding small batch standard deviations to the discriminator to eliminate checkerboard artifacts, the im-proved FastGAN is more suitable for citrus disease and nutritional deficiency (zinc and magnesium deficiency) image generation. It can generate higher-quality and more realistic disease and nutritional deficiency images with higher diversity when trained on a small number of datasets.

The datasets of citrus melanose, citrus nutritional deficiency, and citrus canker leaves were expanded, and the generated images had the phenotypic characteristics of the real data. With a small dataset, a classification network with 97.04% accuracy was trained using only the generated images, which could successfully identify the four types of citrus leaves.

EfficientNet-B5-pro is proposed, which improves EfficientNet-B5 by using the AWP mechanism and Arcface loss. It has better robustness, better generalization ability, and higher accuracy compared to the unimproved EfficientNet-B5.

2. Materials and Methods

2.1. Dataset and Test Environment Setup

Nutritionally deficient citrus leaves and healthy citrus leaves used in this study were collected from the orchard of the eastern district of South China Agricultural University. Citrus canker leaves were collected on rainy days from the citrus orchard of Dingkeng village, Aotou Town, Conghua district, Guangzhou, China, and citrus melanose leaves were obtained from published data on the Kaggle competition [34]. The collection device was an iPhone 13 Pro Max, the shooting distance was 15–30 cm, the picture resolution

was 3024 × 4032 pixels, the picture storage format was JPG, the picture dates were 18 and 25 May and 15 and 25 July 2022, and the shooting times were 10:00–12:00 and 14:00–17:00. The shooting scene was in natural light conditions. A total of 256 citrus canker leaves, 237 healthy citrus leaves, 224 nutritional deficiency citrus leaves, and 192 citrus melanose leaves were collected. Images are shown in Figure 1.

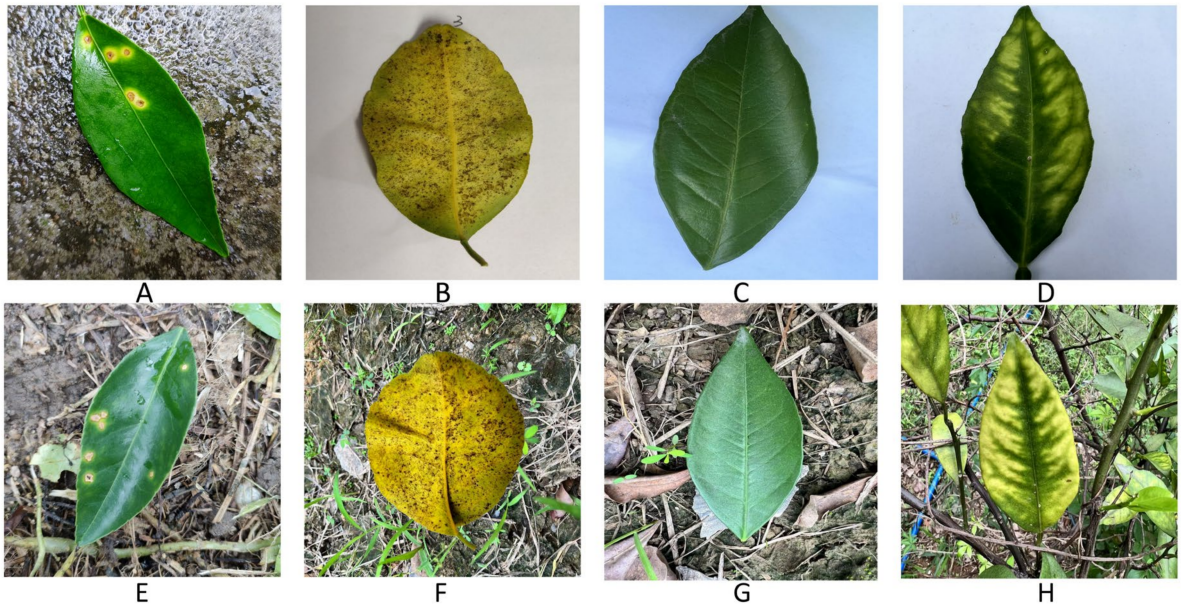


Figure 1. The four kinds of leaves used in the experiment: (**A**,**E**) canker, (**B**,**F**) melanose, (**C**,**G**) healthy, (**D**) magnesium deficiency, and (**H**) zinc deficiency; (**A**–**D**) is simple background, (**E**–**H**) is complicated background.

All collected images were used as the original dataset A. Fifty images of each of the four diseases in dataset A were randomly selected as the training set Train-GAN for the GAN network. The remaining images were used as the test set Test-CNN for the classification network. Using the trained FastGAN2 network, 2000 images were generated for each class of disease separately as the training set Train-CNN. The detailed numbers of different species of leaves can be found in Table 1.

Table 1. The number of different species of leaves in each dataset.

Leaves	Train-GAN	Train-CNN	Test-CNN
Melanose	50	2000	142
Healthy	50	2000	187
Canker	50	2000	206
Nutritional Deficiency	50	2000	174

In the training of the FastGAN2 network, the epoch number was set to 30,000, the batch size was set to 8, the input image size was 1024, the learning rate of the generator and discriminator was 0.0002, and the optimizer was Adaptive Moment Estimation (Adam). In the training of the classification network, the batch size was set to 16, the input image size was 256, the cosine annealing learning rate strategy was used [35], and AdamW was the optimizer [36]. The experiments were conducted on Ubuntu 20.04 with Python 3.8, PyTorch 1.10.0, and Cuda 11.3. The graphics card was RTX3090 with 24 GB of video memory, and the CPU was AMD EPYC 7543. The framework flowchart of generation and classification model are shown in Figure 2.

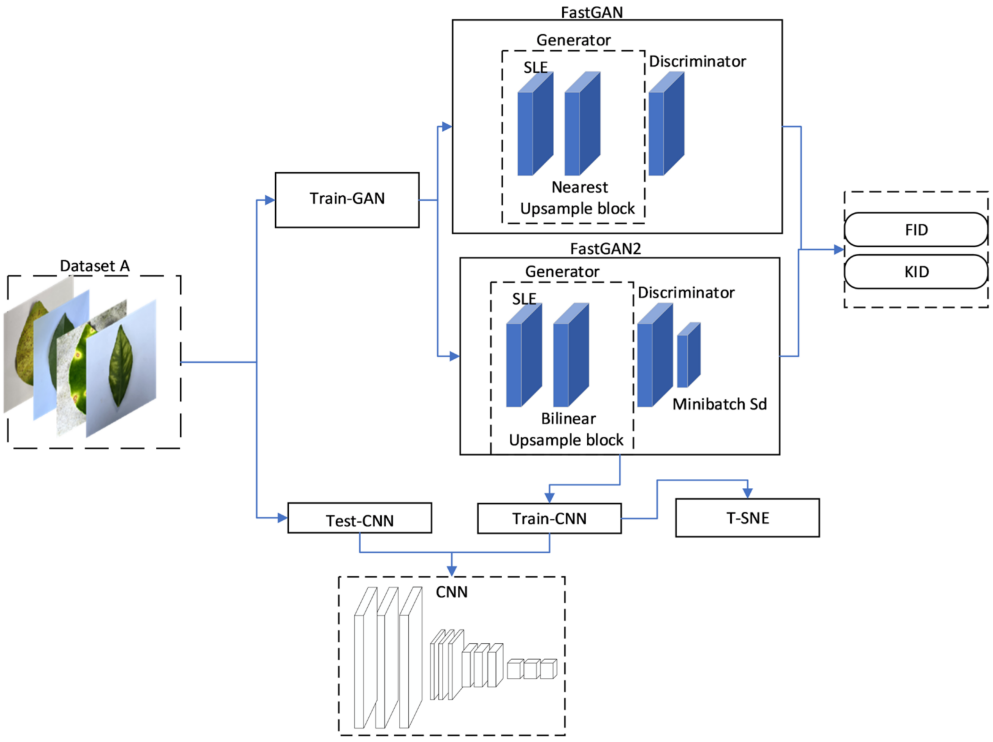


Figure 2. Framework flowchart of generation and classification. Dataset A is divided into Train-GAN and Test-CNN. Train-GAN is used to train FastGAN and FastGAN2. The trained FastGAN2 generates Train-CNN as the training set of CNN and Test-CNN as the tFest set. T-SNE is used to evaluate the quality of Train-CNN. FIID and KID are used to compare the ability of FastGAN and FastGAN2 to generate images.

2.2. FastGAN Model

The FastGAN model was proposed by Liu et al. [27]. The authors improved StyleGAN2 by designing the SLE (skip-layer excitation) module, which shortens the training time of the original model from several days to a dozen hours. They also proposed a self-supervised discriminator that can train the network more stably with a small number of training samples and limited computational resources.

When synthesizing higher-resolution images, the G (Generator) must be deeper and have more convolutional layers to meet upsampling requirements. The more convolutional the layers, the longer the training time of the model. He et al. [37] designed the residual structure ResBlock, which uses cross-layer connections to enhance the gradient signal between the layers. However, this also increases computational costs. The idea of cross-layer connectivity has been reconstructed in the SLE module. This module implements the skip connection as the summation of elements between activations from different convolutional layers, which requires the same spatial dimension of activations. The module also eliminates the heavy computation of convolution by replacing addition with multiplication, as shown in Equation (1), to perform skip connections between different resolutions:

$$y = \mathcal{F}(x_{low}, \{W_i\}) \cdot x_{high} \tag{1}$$

In the formula, x and y are the input and output feature maps of the SLE module, respectively. Function $\mathcal{F}$ contains the operation on x_{low} , and $\{W_i\}$ denotes the learned module weights, where x_{low} and x_{high} are the feature maps at resolutions of 8×8 and 128×128, respectively.

Figure 3 shows the structure of the SLE module, where the adaptive pooling layer in $\mathcal{F}$ first downsamples x_{low} to 4×4 along the spatial dimension, and a convolution layer further downsamples it to 1×1. LeakyReLU is used to model nonlinearity, and another

convolution layer projects x_{low} to have the same channel size as x_{high}. Finally, after gating by the sigmoid function, the output of $\mathcal{F}$ is multiplied by x_{high} along the channel dimension to obtain a y with the same shape as x_{high}.

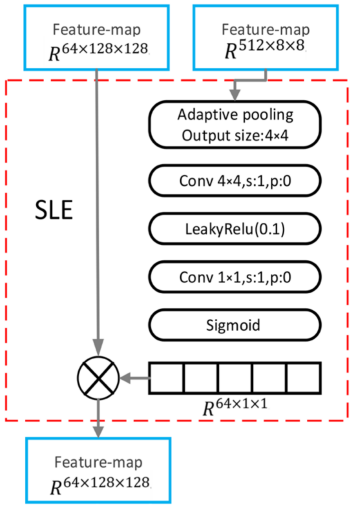


Figure 3. SLE module structure. Feature maps at 8×8 and 128×128 are, respectively, x_{low} and x_{high} in Equation (1). All the black boxes are used to change the size of x_{low} .

2.3. Improved FastGAN Model

When the FastGAN model is applied to citrus leaf disease sample generation, the generated image disease features are unclear. There is insufficient sample richness and a high repetition rate when several thousand images are generated using the trained model. To solve these problems and make the model more applicable to citrus disease sample generation, we improved FastGAN to propose the FastGAN2 model.

First, the structure of the original model generator was redesigned. The interpolation algorithm used in the original model upsampling is the Nearest Interpolation model, which has the fastest calculation speed; however, the effect is poor, and the generated image can easily produce jaggedness. We replaced it with the Bilinear model, which uses 4-pixel points in the original image to calculate the 1-pixel point in the new image to smoothen the generated image. Unlike the original model, which uses the same upsampling block for all layers, we used upsampling blocks with different strategies for different layers. The 4^2–32^2 layers representing rough features use a single convolutional layer, the 64^2–256^2 layers representing fine features use a convolutional layer and a noise layer, and the 512^2–1024^2 layers use two convolutional and two Noisejection layers. For all upsampling blocks, the PixelNorm layer replaces the BachNorm layer in the original model. The PixelNorm layer is demonstrated in Equation (2), and the improved generator structure is shown in Figure 4. The Noisejection layer was independently added to each pixel to control random changes.

$$b_{x,y} = \frac{a_{x,y}}{\sqrt{\frac{1}{N}\sum_{j=0}^{N-1}\left(a_{x,y}^{j}\right)^2 + \epsilon}}, \tag{2}$$

In the formula, N is the number of feature maps, where $a_{x,y}$ and $b_{x,y}$ are, respectively, the original and normalized feature vectors in pixels (x, y), $\epsilon = 10^{-8}$. The PixelNorm layer normalizes each pixel point to avoid extreme weights of input noise and improve stability.

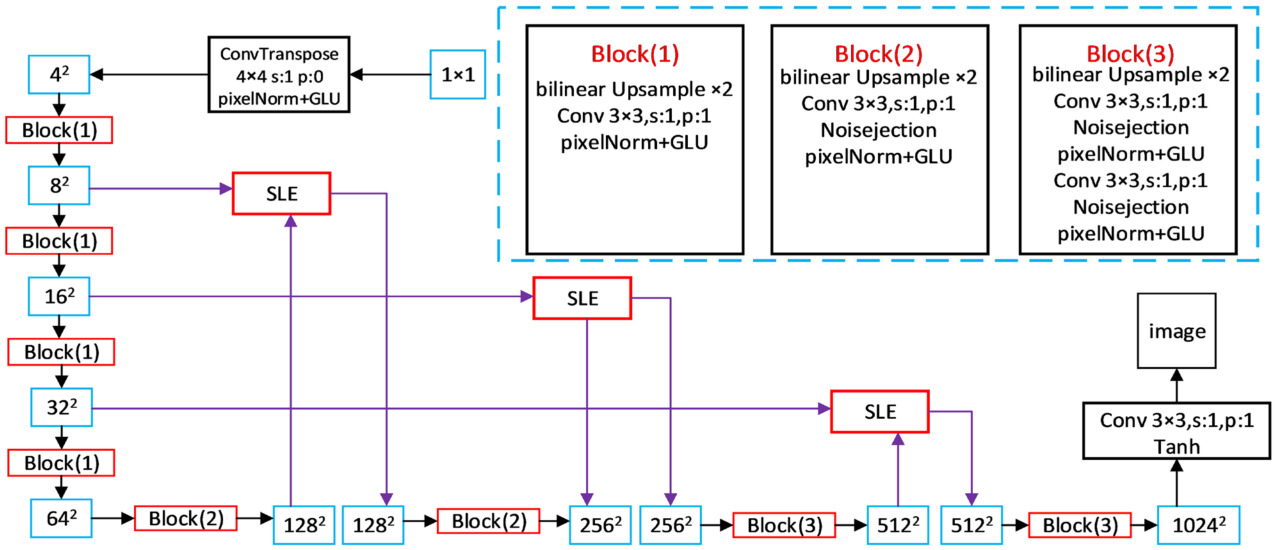


Figure 4. The redesigned generator structure. The blue dotted line frame contains three different upsample blocks. The blue solid line frame represents the size of the feature map, while the red solid line frame represents the SLE module in Figure 3.

Because the FastGAN2 model training is performed on 50 images, it is easy to cause mode collapse and generate duplicate images when generating 2000 images. Karras, Aila, Laine, and Lehtinen [38] proposed a mini-batch standard deviation to increase the diversity of generated samples. To solve this problem, we added it at the end of the discriminator of the original model, as shown in Figure 5. The method divides a batch size image into X parts evenly, where each part contains batch $\frac{\text{size}}{X}$ images, and finds the standard deviation of each image's feature map at different spatial locations; i.e., NumPy's standard function is used to find the standard along the sample dimension. This results in a new feature map, which is then averaged over the feature map to obtain a value, and this value is expanded to the size of a feature map and stitched together with the original feature map as the input to the next layer, i.e., adding a statistical channel to each image.

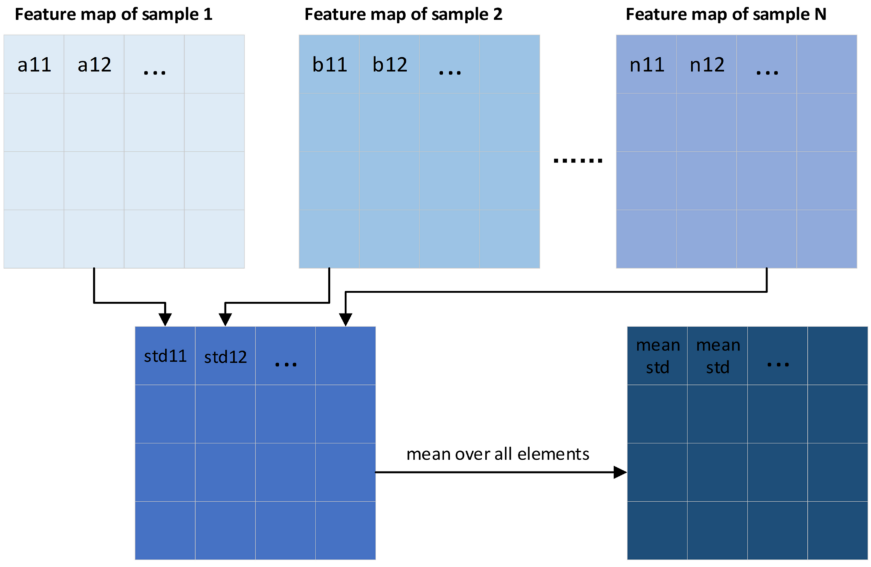


Figure 5. The minibatch standard deviation flow chart. Feature map1, 2 . . . N is divided equally into multiple parts and finds the standard deviation of each image's feature map at different spatial locations. This results in a new feature map, which is then averaged over the feature map to obtain a value, and this value is expanded to the size of a feature map.

2.4. Improved EfficientNet-B5

After comparing the classification performances of Densenet121, ResNet50, ShuffleNetv2, Mlp-Mixer, MobileNetv3, EfficientNet-B3, and EfficientNet-B5, we found that EfficientNet-B5 had the best performance (performance comparison can be found in Table 1). Therefore, we improve EfficientNet-B5 by using the AWP mechanism to add perturbations to model weights and inputs to increase the robustness of the model. We also used Arcface loss and added a cumulative corner margin between elements and target weights to enhance intra-class compactness and inter-class variability. Arcface is a loss function applied to face recognition, as proposed by Deng et al. [39], and it maximizes the classification boundaries in angular space and adds a cumulative angular margin between features and target weights to enhance intra-class compactness and inter-class differences.

Arcface was used instead of the softmax loss function because of the high similarity of various citrus leaf diseases. The softmax loss function formula is given by Equation (3):

$$L_1 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}}, \tag{3}$$

where x_i denotes the deep feature of the ith sample associated with the category y_i, j denotes the column of weight w, b_j is the bias corresponding to w_j, n is the number of categories, and N is the batch size.

The softmax loss function does not explicitly optimize the embedding function to achieve a higher similarity of samples within categories and a higher diversity of samples between categories. It is improved by first setting the bias b_j to 0, and the inner product of the weights and inputs is expressed by Equation (4):

$$W_j^T x_i = \| W_j \| \| x_i \| cos\ \theta_j \tag{4}$$

Processing w_j with L_2 regularization makes $\| W_j \| = 1$, (L_2 regularization is to divide each value in the w_j vector by the modulus of w_j separately to obtain the new w_j, and the modulus of the new w_j is 1).

Hence, Equation (5) can be obtained from Equation (3):

$$L_2 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{\|x_i\|\cos(\theta_{y_i})}}{e^{\|x_i\|\cos(\theta_{y_i})_+} + \sum_{j=1,j\neq y_i}^{n} e^{\|x_i\|\cos\theta_j}} \tag{5}$$

input x_i is also regularized with L_2 by multiplying by scale parameter s. The normalization step of the features and weights makes the prediction depend only on the angle between the features and weights. Thus, the learned embedding features were distributed on a hypersphere of radius s. Because the embedding features are distributed around the center of each feature on the hypersphere, a cumulative angular margin is added between x_i and y_i to enhance the intra-class compactness and inter-class variance [40], yielding Equation (6) for the Arcface:

$$L_3 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}} \tag{6}$$

To improve the generalization ability of the model, an adversarial training mechanism (AWP) proposed by Wu et al. [41] was used for EfficientNet-B5. Adversarial training introduces noise, which can regularize parameters and improve the robustness of the model. The weight loss landscape (WLL) is used to represent the standard generalization gap (the difference between the accuracy of the model on the training set and the accuracy of the test set) under standard training scenarios. As a result of the AWP limiting the flatness of the WLL in adversarial training, a dual perturbation mechanism is formed, that

is, the adversarial perturbation input and weights. The original adversarial loss is given by Equation (7):

$$\min_{w} \rho(w), \rho(w) = \frac{1}{n} \sum_{i=1}^{n} \max_{\|x_i' - x_i\|_p \leq \epsilon} \ell\left(f_w\left(x_i'\right), y_i\right) \tag{7}$$

where n is the number of training samples, x_i' is the adversarial sample within a ball (bounded by L_p parameterization) centered on the natural sample x_i, f_w is the deep neural network (DNN) with weight w, $\updownarrow()$ is the standard classification loss, such as cross-entropy loss, and $\rho(w)$ is the adversarial loss.

The AWP algorithm flattens the WLL by injecting worst-case weight perturbations into a DNN. To improve the test robustness, investigators must focus on training robustness and the robustness generalization gap, as shown in Equation (8):

$$\min_{w}\{\rho(w) + (\rho(w+V) - \rho(w))\} \rightarrow \min_{w} \rho(w+V) \tag{8}$$

where $\rho(w)$ is the original adversarial loss in Equation (7), $\rho(w+V) - \rho(w)$ denotes the flatness of the WLL, and V is the perturbation weight that must be carefully chosen. Unlike the commonly used random weight perturbation (random direction sampling), the adversarial loss of the AWP is significantly increased, as demonstrated in Equation (9):

$$\min_{w} \max_{V \in \mathcal{V}} \rho(w+V) \rightarrow \min_{w} \max_{V \in \mathcal{V}} \frac{1}{n} \sum_{i=1}^{n} \max_{\|x_i' - x_i\|_p \leq \epsilon} \ell\left(f_{w+V}\left(x_i'\right), y_i\right), \tag{9}$$

where $\mathcal{V}$ is the range of values of the perturbation weights V. The maximization of V depends on the entire instance to maximize the entire loss (not the loss of each instance); therefore, these two maximizations are not interchangeable.

3. Results and Discussion

3.1. Generate Image Quality Ratings

The training set Train-GAN was used to train FastGAN for approximately 2.5 h and FastGAN2 for approximately 3 h to complete the training of 30,000 epochs. The model was saved every 5000 epochs, and the images generated during the training process were saved every 1000 epochs.

T-SNE is a data dimensionality reduction and visualization method [42], which is an embedding model that can map data in a high-dimensional space to a low-dimensional space and preserve the local characteristics of the dataset. Therefore, when we want to classify a high-dimensional dataset, it is not clear whether this dataset has good separability (small intervals between similar classes and large intervals between distinct classes), and we can project the data into a 2- or 3-dimensional space for observation by T-SNE.

In this study, we reduced the dimensionality of the FastGAN2-generated images and the original images to observe their distribution in two-dimensional space. As demonstrated in Figure 6, where Figure 6A,B are the original images and generated images in the two-dimensional space distribution, respectively, by observing Figure 6A, it can be seen that the interval between the nutritional deficiency leaves and healthy leaves is a small boundary and is not clear, indicating that these two types of leaves do not exhibit a strong classification of difference; Figure 6B shows that the generated images also conform to this distribution. Figure 6C shows the fitting of the scatter plot between the generated and original images, the circle symbol representing the real image and the cross symbol representing the generated image overlap, which indicates that the four generated types of images are in the same distribution as the real image.

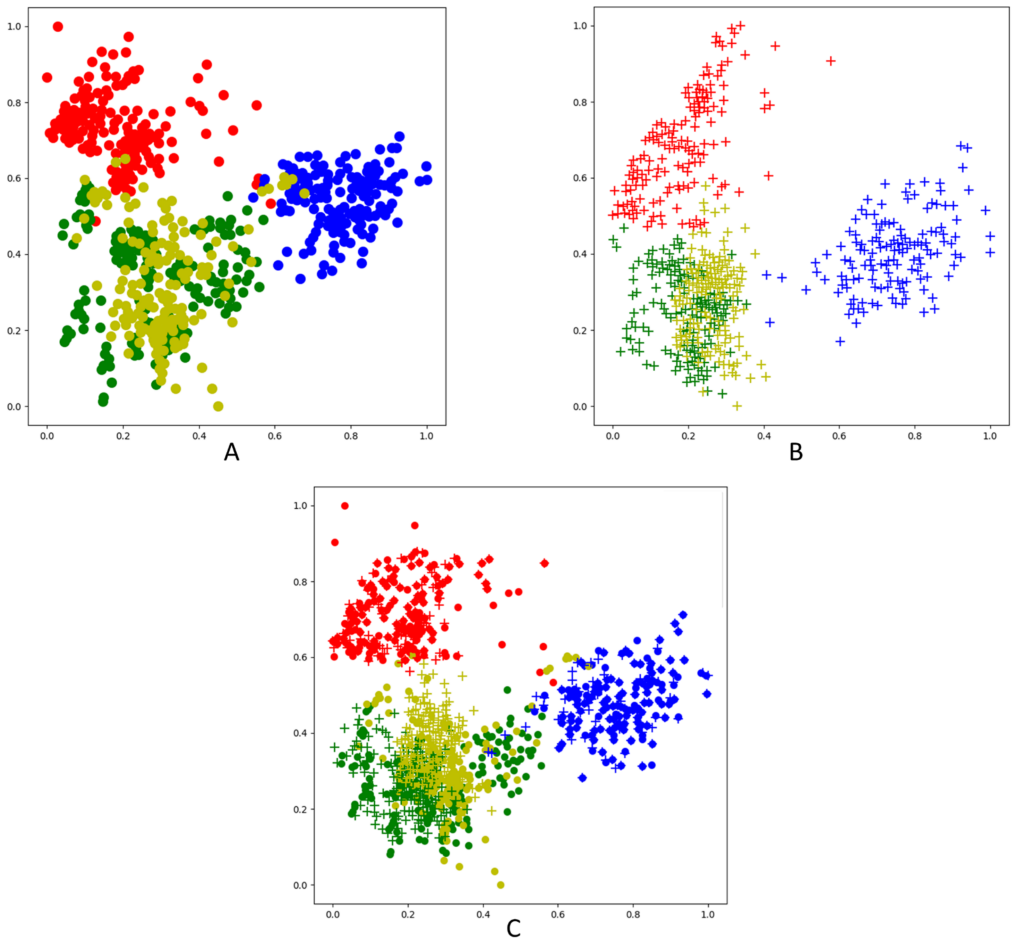


Figure 6. Scatter plot of T-SNE after dimensionality reduction: the circles are the real dataset; the crosses are the fake dataset. Red = melanose; dark-green = healthy; light green = nutritional deficiency; blue = canker. (**A**) only real dataset; (**B**) only fake dataset; (**C**) combination of real and fake dataset.

The FID is the distance between the generated and real images. A smaller distance indicates a well-generated image, that is, a sharp and diverse image. The FID is computed through the inception model, which removes the last fully connected layer of the inception model used for classification and outputs a 2048-dimensional vector from the previous layer. Here, inception no longer performs classification but feature extraction to obtain 2048-dimensional vectors, and each dimension represents some feature. N was taken for each of the generated and real images. After the transformed inception network, each obtains an N*2048-dimensional feature vector and then uses Equation (10) to calculate the distance between the two N*2048-dimensional feature vectors:

$$FID = \| \mu_r - \mu_g \|^2 + T_r(\sum r + \sum g - 2(\sum r \sum g)^{1/2}), \quad (10)$$

where μ_r denotes the feature mean of the real image, μ_g is the feature mean of the generated image, $\sum r$ is the covariance matrix of the real image, and $\sum g$ is the covariance matrix of the generated image.

KID is a GAN generation quality metric that is very similar to FID to assess the degree of GAN convergence [43], which does not require the assumption of a normal distribution of FID and is an unbiased estimation.

Table 2 summarizes the comparison of the FID and KID scores of FastGAN2 and FastGAN on the four types of leaves. The FID and KID scores of the four kinds of leaves generated by the FastGAN2 network are lower than those of the FastGAN network, indicating that the improved network generates images with better clarity and diversity.

Table 2. Comparison of FID and KID scores of FastGAN2 and FastGAN on four kinds of leaves.

Leaves	FastGAN2		FastGAN	
	FID	**KID**	**FID**	**KID**
Melanose	64.73	3.01	153.52	13.79
Healthy	73.90	2.18	90.98	4.71
Canker	55.53	2.21	64.59	3.01
Nutritional Deficiency	53.56	2.03	54.15	2.01
Average	61.93	2.36	90.81	5.88

Figure 7 compares the details of the citrus ulcer images generated by FastGAN2 and FastGAN. The images generated by the FastGAN network have unclear lesions, checkerboard artifacts on the midvein and tips, and the leaf contours are distorted (not continuous). The images generated by the FastGAN2 network are closer to the original images, and semantic information remains intact for the whole leaf and enlarged lesions, midvein, and leaf tips. Their color, texture, and contour features are more obvious.

Figure 7. "Real image", "FastGAN2", and "FastGAN" images down the sides and "Leaf", "Lesion", "Midvein", and "Leaf tip" images across the top.

Figure 8 shows the comparison between the image generated by FastGAN and FastGAN2 and the original image with the complex background. It can be seen from row B that the background of the image produced by FastGAN2 is very close to the original image, and the leaf contour and midvein transition are natural. However, it can be seen from the image generated by FastGAN in line C that the background of the image is seriously distorted, and there are problems such as distortion and unclear leaf contour. This indicates that FastGAN2 can also generate high-quality images of diseased citrus leaves under complex backgrounds.

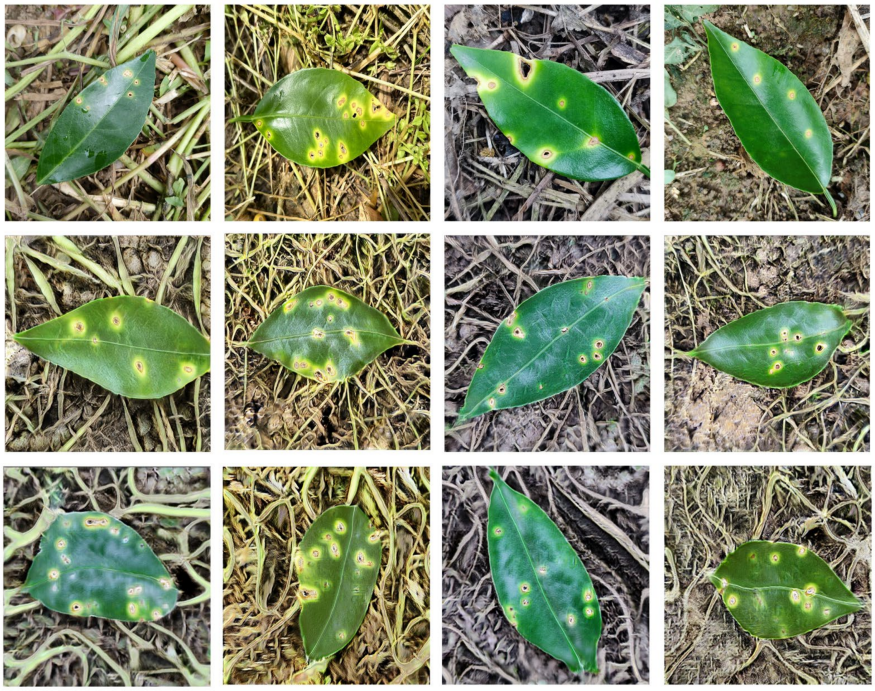


Figure 8. Comparison between the image generated by FastGAN and FastGAN2 and the original image with the complex background: "Real image", "FastGAN2" and "FastGAN" images down the sides.

Figure 9 compares the four types of citrus leaf images generated by FastGAN2 and FastGAN with the original images, and the FastGAN2 generated images are closer to the original images than those generated by FastGAN. Figure 10 shows some of the images in the dataset Train-CNN generated by the FastGAN2 network; the four generated types of citrus leaves match the detailed features of the real images, in which the traces of rain can be seen in the background of the generated images. There are reflections formed by water on the leaves as the real images of the ulcer leaves were taken on a rainy day.

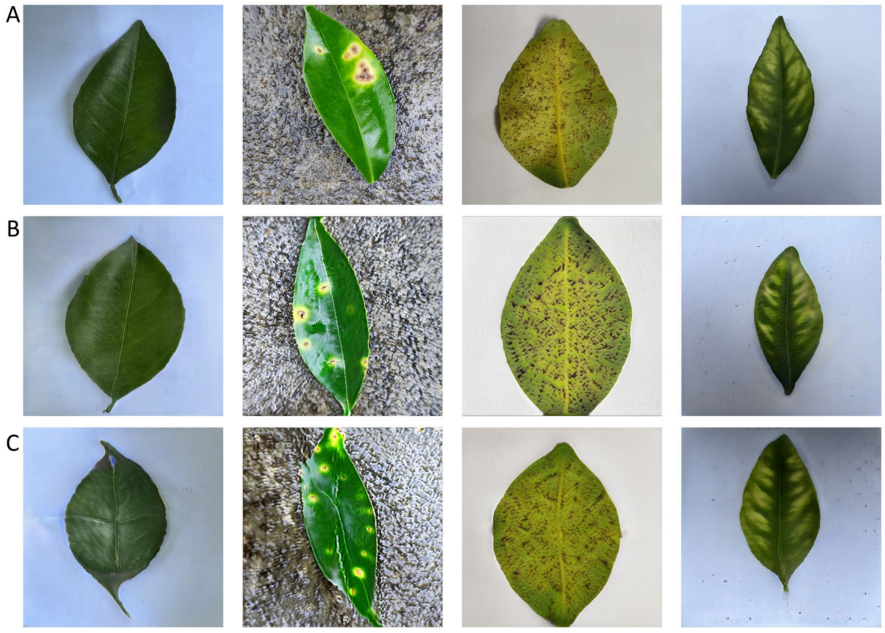


Figure 9. Comparison of four kinds of citrus leaf images generated by FastGAN and FastGAN2 with real images: (**A**) real dataset; (**B**) images generated by FastGAN2; (**C**) images generated by FastGAN.

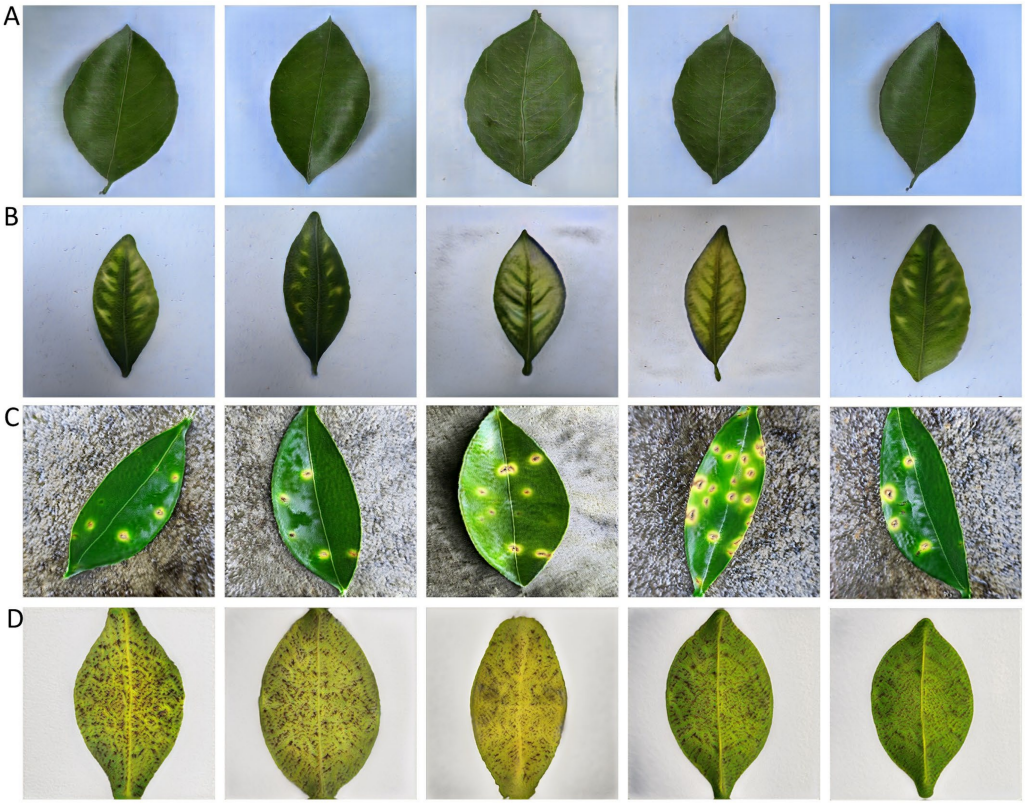


Figure 10. A portion of the dataset generated by FastGAN2: (**A**) healthy; (**B**) nutritional deficiency; (**C**) canker; (**D**) melanose.

3.2. Classified Network Performance Evaluation

To verify whether the classification network trained using the images generated by the FastGAN2 network only can successfully recognize the captured images, we trained ten classification networks using the Train-CNN, the image dataset generated by the FastGAN2 network, and evaluated the classification network performance using Test-CNN as the test set with accuracy, recall, precision, and F1-score as the evaluation metrics. This is formulated in Equations (11)–(14):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{11}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{12}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{13}$$

$$\text{F1-score} = \frac{2 * Precision * Recall}{Precision \ + \ Recall}, \tag{14}$$

where TP is the number of samples with a true positive value and positive prediction, FP is the number of samples with a true negative value but positive prediction, FN is the number of samples with a true positive value but negative prediction, and TN is the number of samples with a true negative value and negative prediction.

In this study, the images generated by FastGAN and Train-CNN are, respectively, used as training sets to train the classification network; Test-CNN was used as the test set to evaluate the performance of the classification network. It can be seen from Table 3 that the accuracy, precision, recall and F1-score achieved by using FastGAN2 to generate images for

training ten classification networks are higher than those achieved by FastGAN. Among them, EfficientNet-B5 has the highest score in each item, which is higher than the average.

Table 3. Performance comparison of ten kinds of classification networks.

Network	Accuracy%		Precision%		Recall%		F1-Score%	
	FastGan	FastGan2	FastGan	FastGan2	FastGan	FastGan2	FastGan	FastGan2
Densenet121	82.64	90.97	88.95	93.79	74.21	91.31	72.03	91.61
Shufflenetv2	80.21	94.64	83.15	94.41	70.71	94.41	70.58	94.37
Mlp Mixer	82.33	94.22	93.04	95.20	73.83	94.12	73.09	94.47
MobileNetV3	82.29	92.81	92.63	94.13	73.87	92.49	72.98	93.06
ResNet50	80.56	91.26	89.99	93.65	71.70	91.56	69.33	91.86
Vision Transformer	88.54	93.97	89.52	94.52	81.15	94.01	83.47	94.04
Swin Transformer	88.62	93.42	73.87	95.19	71.17	93.61	71.73	93.88
EfficientNet-B3	86.11	92.10	86.05	94.12	80.33	92.34	77.15	92.63
EfficientNet-B5	90.62	94.78	95.15	96.13	86.34	94.98	88.52	95.23
EfficientNet-B5-pro	94.64	97.04	95.33	97.32	94.75	96.96	94.84	97.09
Average	85.66	93.52	88.76	94.86	77.80	93.58	77.37	93.82

The confusion matrix in Figure 11 shows that EfficientNet-B5-pro has better generalization and robustness [44]. It can be seen from Figure 11A that EfficientNet-B5-pro judges that the largest number of errors is within the nutritional deficiency leaves, as it misjudges 11 nutritional deficiency leaves as healthy leaves. According to the previous T-SNE results, it can be seen that this may be due to the higher similarity between healthy and nutritional deficiency leaves. Compared with Figure 11B, it can be seen that the unimproved EfficientNet-B5 misclassifies more healthy leaves, while EfficientNet-B5-Pro has similar discrimination rates for four kinds of leaves, indicating that the improved EfficientNet-B5 has better intra-class compactness and inter-class difference.

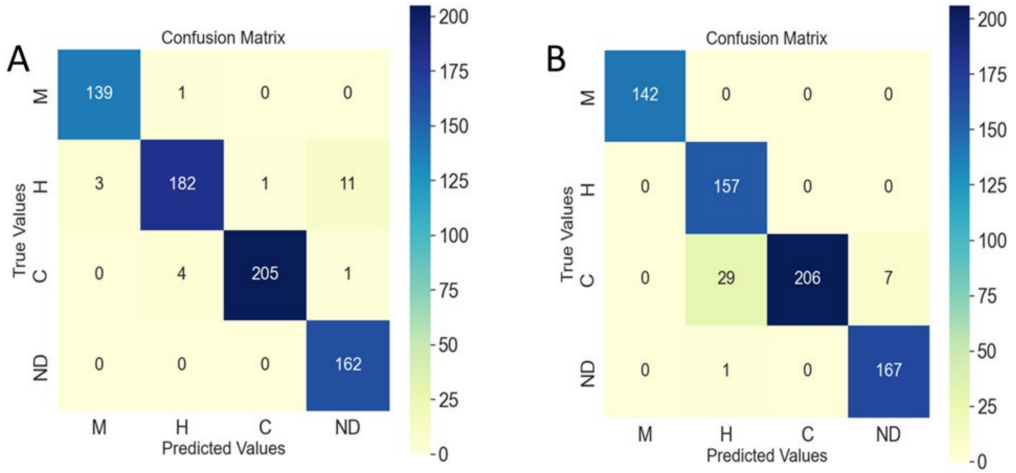


Figure 11. Confusion Matrix: (**A**) EfficientNet-B5-pro; (**B**) EfficientNet-B5. "M" = melanose; "H" = healthy; "C" = canker; "ND" = nutritional deficiency.

Table 4 depicts the classification results of EfficientNet-B5-pro for four kinds of leaves. The nutritional deficiency leaves were divided into two separate groups (zinc and magnesium deficiency). In this table, the average score for healthy leaves is the lowest and that for nutritionally deficient leaves is the second lowest. This may be due to the high similarity and small differences between healthy leaves and nutritionally deficient leaves, which makes classification difficult. It can also be seen from the previously mentioned T-SNE reduced dimensional visualization method that the boundaries between healthy and deficient leaves are not clear in the scatter plot. Figure 12 shows the training loss and test accuracy curves of the EfficientNet-B5-pro model.

Table 4. EfficientNet-B5-pro classification result.

Leaves	F1-Score%	Precision%	Recall%	Average%
Melanose	98.58	99.29	97.89	98.59
Healthy	94.79	92.39	97.33	94.84
Canker	98.56	97.62	99.51	98.56
Zinc Deficiency	95.92	97.92	94	95.95
Magnesium Deficiency	96.93	99.3	94.67	96.97

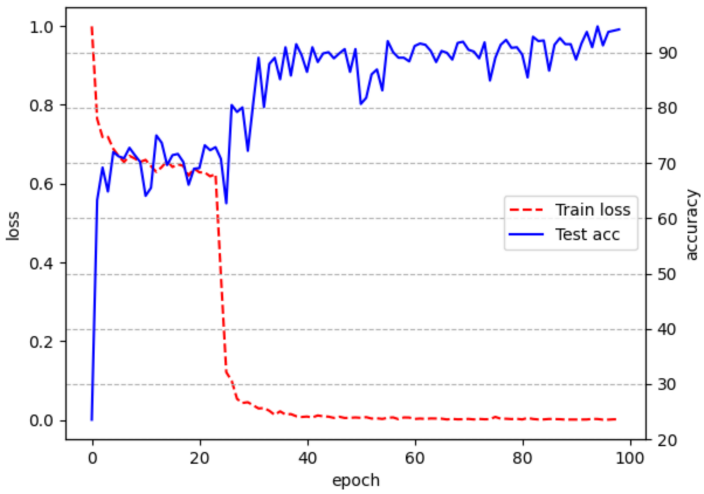


Figure 12. Training loss and test accuracy of the EfficientNet-B5-pro model.

4. Conclusions

In this study, we generated citrus disease and nutritional deficiency phenotype data using the FastGAN2 network, trained the classification network using only the generated images, and achieved high accuracy on the shot dataset. The impact on the performance of classification networks when the training and test sets are from different regions is addressed, and the generalization and robustness of classification networks in citrus disease and nutritional deficiency leaf identification are improved. Compared to most GAN methods for generating plant disease samples, this study does not require pairs of data, no background must be removed, and only a small amount of data is required to complete the training of the GAN network. Through the model proposed in this paper, high-quality citrus disease and nutritional deficiency images are successfully generated. After training the classification network solely by using the images generated in this paper, citrus disease and nutritional deficiency images in a real environment can be successfully identified with high accuracy. Although this study was conducted for citrus diseases, the method used can also provide new ideas for the classification and identification of other plant diseases where samples are not sufficient. In the future, we will develop a mobile application that will deploy the algorithms mentioned in this article to a website or software to help fruit farmers identify citrus disease.

Author Contributions: Data curation, Y.W. and Z.C.; project administration, S.L.; resources, S.S. and D.S.; supervision, Z.L.; writing—original draft, Y.G.; writing—review and editing, Q.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant No. 31971797). This study was also partly supported by the China Agriculture Research System of MOF and MARA (grant No. CARS-26) and the Guangdong Provincial Special Fund for Modern Agriculture Industry Technology Innovation Teams (grant No. 2022KJ108, No. 2023KJ108).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations

Arcface	Adaptive angular margin
FID	Fréchet inception distance
KID	Kernel inception distance
GAN	Generative adversarial network
G	Generator
SLE	Skip-layer excitation
DNN	Deep neural network
WLL	Weight loss landscape
AP	Average precision
CNN	Convolutional neural network
AWP	Adversarial weight perturbation
Adam	Adaptive moment estimation

References

1. Iqbal, Z.; Khan, M.A.; Sharif, M.; Shah, J.H.; Rehman, M.H.U.; Javed, K. An automated detection and classification of citrus plant diseases using image processing techniques. *Comput. Electron. Agric.* **2018**, *153*, 12–32. [CrossRef]
2. Gulzar, Y. Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique. *Sustainability* **2023**, *15*, 1906. [CrossRef]
3. Palei, S.; Behera, S.K.; Sethy, P.K. A Systematic Review of Citrus Disease Perceptions and Fruit Grading Using Machine Vision. *Procedia Comput. Sci.* **2023**, *218*, 2504–2519. [CrossRef]
4. Mamat, N.; Othman, M.F.; Abdulghafor, R.; Alwan, A.A.; Gulzar, Y. Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach. *Sustainability* **2023**, *15*, 901. [CrossRef]
5. Wu, F.Y.; Duan, J.L.; Ai, P.Y.; Chen, Z.Y.; Yang, Z.; Zou, X.J. Rachis detection and three-dimensional localization of cut off point for vision-based banana robot. *Comput. Electron. Agric.* **2022**, *198*, 107079. [CrossRef]
6. Wu, F.Y.; Duan, J.L.; Chen, S.Y.; Ye, Y.X.; Ai, P.Y.; Yang, Z. Multi-Target Recognition of Bananas and Automatic Positioning for the Inflorescence Axis Cutting Point. *Front. Plant. Sci.* **2021**, *12*, 705021. [CrossRef]
7. Tang, Y.C.; Zhou, H.; Wang, H.J.; Zhang, Y.Q. Fruit detection and positioning technology for a *Camellia oleifera* C. Abel orchard based on improved YOLOv4-tiny model and binocular stereo vision. *Expert Syst. Appl.* **2023**, *211*, 118573. [CrossRef]
8. Da Silva, J.C.F.; Silva, M.C.; Luz, E.J.S.; Delabrida, S.; Oliveira, R.A.R. Using Mobile Edge AI to Detect and Map Diseases in Citrus Orchards. *Sensors* **2023**, *23*, 2165. [CrossRef]
9. Kamilaris, A.; Prenafeta-Boldu, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
10. Zhang, S.W.; Wu, X.W.; You, Z.H.; Zhang, L.Q. Leaf image based cucumber disease recognition using sparse representation classification. *Comput. Electron. Agric.* **2017**, *134*, 135–141. [CrossRef]
11. Liu, H.Y.; Jiao, L.; Wang, R.J.; Xie, C.J.; Du, J.M.; Chen, H.B.; Li, R. WSRD-Net: A Convolutional Neural Network-Based Arbitrary-Oriented Wheat Stripe Rust Detection Method. *Front. Plant. Sci.* **2022**, *13*, 876069. [CrossRef] [PubMed]
12. Zhong, Y.; Zhao, M. Research on deep learning in apple leaf disease recognition. *Comput. Electron. Agric.* **2020**, *168*, 105146. [CrossRef]
13. Yao, N.; Ni, F.; Wang, Z.; Luo, J.; Sung, W.K.; Luo, C.; Li, G. L2MXception: An improved Xception network for classification of peach diseases. *Plant Methods* **2021**, *17*, 36. [CrossRef] [PubMed]
14. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions in Computer Vision and Pattern Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
15. Janarthan, S.; Thuseethan, S.; Rajasegarar, S.; Lyu, Q.; Zheng, Y.Q.; Yearwood, J. Deep Metric Learning Based Citrus Disease Classification with Sparse Data. *IEEE Access* **2020**, *8*, 162588–162600. [CrossRef]
16. Salman, S.; Liu, X. Overfitting Mechanism and Avoidance in Deep Neural Networks. *arXiv* **2019**, arXiv:1901.06566v1.
17. Li, W.; Chen, C.; Zhang, M.M.; Li, H.C.; Du, Q. Data Augmentation for Hyperspectral Image Classification with Deep CNN. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 593–597. [CrossRef]
18. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the 28th Conference on Neural Information Processing Systems, Montréal, QC, Canada, 8–13 August 2014.
19. You, C.; Li, G.; Zhang, Y.; Zhang, X.; Shan, H.; Li, M.; Ju, S.; Zhao, Z.; Zhang, Z.; Cong, W.; et al. CT Super-Resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE). *IEEE Trans. Med. Imaging* **2020**, *39*, 188–203. [CrossRef]
20. Lin, Y.; Li, Y.; Cui, H.; Feng, Z. WeaGAN: Generative Adversarial Network for Weather Translation of Image among Multi-domain. In Proceedings of the International Conference on Behavioral, Economic and Socio-Cultural Computing, Beijing, China, 28–30 October 2019.

21. Niu, S.L.; Li, B.; Wang, X.G.; Lin, H. Defect Image Sample Generation with GAN for Improving Defect Recognition. *IEEE Trans. Automat. Sci. Eng.* **2020**, *17*, 1611–1622. [CrossRef]
22. Ma, L.; Shuai, R.; Ran, X.; Liu, W.; Ye, C. Combining DC-GAN with ResNet for blood cell image classification. *Med. Biol. Eng. Comput.* **2020**, *58*, 1251–1264. [CrossRef]
23. Cap, Q.H.; Uga, H.; Kagiwada, S.; Iyatomi, H. LeafGAN: An Effective Data Augmentation Method for Practical Plant Disease Diagnosis. *IEEE Trans. Automat. Sci. Eng.* **2020**, *9*, 1258–1267. [CrossRef]
24. Xiao, D.Q.; Zeng, R.L.; Liu, Y.F.; Huang, Y.G.; Liu, J.B.; Feng, J.Z.; Zhang, X.L. Citrus greening disease recognition algorithm based on classification network using TRL-GAN. *Comput. Electron. Agric.* **2022**, *200*, 107206. [CrossRef]
25. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and Improving the Image Quality of StyleGAN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
26. Karras, T.; Laine, S.; Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 4217–4228. [CrossRef] [PubMed]
27. Liu, B.; Zhu, Y.; Song, K.; Elgammal, A. Towards Faster and Stabilized GAN Training for High-fidelity Few-shot Image Synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 19–25 June 2021.
28. Mohanty, S.P.; Hughes, D.P.; Salathe, M. Using Deep Learning for Image-Based Plant Disease Detection. *Front. Plant. Sci.* **2016**, *7*, 1419. [CrossRef] [PubMed]
29. Ferentinos, K.P. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* **2018**, *145*, 311–318. [CrossRef]
30. Ma, N.; Zhang, X.; Zheng, H.-T.; Sun, J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
31. Tolstikhin, I.; Houlsby, N.; Kolesnikov, A.; Beyer, L.; Zhai, X.; Unterthiner, T.; Yung, J.; Steiner, A.; Keysers, D.; Uszkoreit, J.; et al. MLP-Mixer: An all-MLP Architecture for Vision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 19–25 June 2021.
32. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019.
33. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019.
34. Kaggle. CCL'20 | Kaggle. Available online: https://www.kaggle.com/datasets/downloader007/ccl20 (accessed on 23 August 2022).
35. Loshchilov, I.; Hutter, F. SGDR: Stochastic Gradient Descent with Warm Restarts. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
36. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
37. 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, Boston, MA, USA, 7–12 June 2015.
38. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive Growing of GANS for Improved Quality, Stability, and Variation. *arXiv* **2018**, arXiv:1710.10196.
39. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
40. Xu, Q.; Huang, G.; Yuan, Y.; Guo, C.; Sun, Y.; Wu, F.; Weinberger, K. An empirical study on evaluation metrics of generative adversarial networks. *arXiv* **2018**, arXiv:1806.07755.
41. Wu, D.; Xia, S.-T.; Wang, Y. Adversarial Weight Perturbation Helps Robust Generalization. In Proceedings of the Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems, Online, 6–12 June 2020.
42. Arora, S.; Hu, W.; Kothari, P.K. An Analysis of the t-SNE Algorithm for Data Visualization. *PlMR* **2018**, *75*, 1455–1462.
43. Bińkowski, M.; Sutherland, D.J.; Arbel, M.; Gretton, A. Demystifying MMD GANs. *arXiv* **2018**, arXiv:1801.01401.
44. Luque, A.; Carrasco, A.; Martín, A.; de las Heras, A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognit.* **2019**, *91*, 216–231. [CrossRef]

agronomy

MDPI

Communication

Evaluation of a Real-Time Monitoring and Management System of Soybean Precision Seed Metering Devices

Jicheng Zhang [1], Yinghui Hou [1], Wenyi Ji [2], Ping Zheng [1], Shichao Yan [1], Shouyin Hou [2] and Changqing Cai [3,4,*]

[1] College of Electrical and Information, Northeast Agricultural University, Harbin 150030, China
[2] College of Engineering, Northeast Agricultural University, Harbin 150030, China
[3] College of Electrical and Information Engineering, Changchun Institute of Technology, Changchun 130012, China
[4] National and Local Joint Engineering Research Center for Smart Distribution Network Measurement, Control and Safe Operation Technology, Changchun 130012, China
* Correspondence: caicq@ccit.edu.cn; Tel.: +86-13-504-400-544

Abstract: Aiming at precise evaluation of the performance of soybean seed metering devices, a photoelectric sensor-based real-time monitoring system was designed. The proposed system mainly included a photoelectric sensor module for seeding signal collecting, Hall sensors speeding module, microcontroller unit (MCU), light and sound alarm module, human–machine interface (HMI), and other parts. The indexes of miss, multiples, flow rate, and application rate were estimated on the basis of seeder speed, seed metering disk rotation rate, photoelectric sensor signals, and clock signals. These real-time statistics of the seeding process were recorded by seeding management system. The laboratory results showed that the detection errors of seeding quantity of both big- and small-diameter soybeans were less than 2.0%. Miss and multiples index estimated by this system were 0.4% and 0.5% than that of seeding image monitoring platform (SIMP), respectively. In field tests, miss and multiples index can be used to evaluate the performance of seed metering device, and big-diameter seeds can be detected more precisely than small ones by these photoelectric sensors. This system can provide support for evaluation of working performance of seed metering devices and have a positive effect on seeding monitoring technology.

Keywords: seed metering device; monitoring system; photoelectric sensor; miss; multiples; flow rate; smart agriculture

Citation: Zhang, J.; Hou, Y.; Ji, W.; Zheng, P.; Yan, S.; Hou, S.; Cai, C. Evaluation of a Real-Time Monitoring and Management System of Soybean Precision Seed Metering Devices. *Agronomy* **2023**, *13*, 541. https://doi.org/10.3390/agronomy13020541

Academic Editor: Gniewko Niedbała

Received: 27 December 2022
Revised: 7 February 2023
Accepted: 10 February 2023
Published: 14 February 2023

1. Introduction

The precision seeding machine has been widely used in precision/smart agriculture recently. It plays an important role in crops growth, yield, and final production income [1]. At present, most precision seeders used are mechanical and pneumatic technology [2]. The phenomenon of miss seeding [3], multiples, and blockage have been reported to often occur [4,5]. However, detecting these problems promptly is difficult for users, which tends to lead to large-scale missed seeding and serious economic losses [3,6]. So, the real-time monitoring of seeding quantity and quality is necessary for modern precision seeding machines.

Especially in the Huang-Huai-Hai region of China, rotation of winter wheat and summer soybean cropping is the main mode of agricultural production. Due to the increase of straw retained after the wheat harvest without rotary tillage or plow tillage, no-tillage seeding mechanism possibly leads to the blockage of later soybean planters [7], which results in large-scale seedless and laborious replanting. Seeding monitor system is one of the important methods to solve with this issue. As we have known, there is no monitoring system for no-tillage seeding machines, in which there were more complex field conditions and difficulty to find out missed seeding. Meanwhile, statistical indexes for the seeding process, such as seeding quantity, sowing rate, application rate, and planting density, have

important monitoring value [8]. Therefore, the performance monitoring of seed metering devices under operational conditions needs to meet the requirements of crop seeding flow rate and its working efficiency [9].

Common monitoring techniques for seeding include photoelectric [10], machine vision [11], laser [1], and microwave detection [12] technology. The photoelectric monitoring method is widely used because of the characteristics of its high accuracy and low cost [3,13,14]. Karimi et al. evaluated that the infra-red (IR) detection technique was more suitable for non-contact sensing techniques for estimating of the seed flow rate than other investigated sensing methods [10]. Hao et al. set the speed of a seedling bed to 3, 4, and 5 $km \cdot h^{-1}$ based on the seed metering performance monitoring system with a precision seeder with a photoelectric sensor. They obtained 94.4% monitoring accuracy with the system [15]. Zhang et al. also evaluated a soybean precision seeder monitoring device with a photosensitive sensor. They set the moving speed of the seeder in field operation to 3.6 $km \cdot h^{-1}$ and 5.4 $km \cdot h^{-1}$ and obtained a monitoring accuracy of more than 96% [16]. The methods and technology of getting more precise is one of the main development directions in seeding monitoring field.

Currently, data is an increasingly important factor in agricultural production, and agricultural machinery is a major vehicle for the acquisition and utilization of agricultural data [17]. There are a large number of studies about agricultural environmental data collection. However, only a limited number of statistical studies had been conducted on the work process of seeders, which led to single-machine data management deficiency and wide-area big data management issues. A real-time seeding monitoring system equipped on a precision seeder, which is a fundamental data collect unit, should not only be used for the seeding process, but also for the development level of agricultural informationization and digitalization [18]. Zhao et al. performed indirect measurement methods of seed and fertilizer operation amounts to calculate the relative errors [19]. Karimi et al. also designed a seeding monitoring system based on the infrared seed sensor, which could provide a calibration method and monitor the sowing process in the field [20]. The scatter plot of application-rate information was drawn by using an application rate measuring system equipped on a mechanical wheat seed–fertilizer drill machine, which could help to evaluate the dynamic process of fertilization and seeding to precisely measure application rate, and to achieve a variable rate application [21]. Along with the development of big data application, agricultural machinery working information is necessary to establish a cross-regional and national data system for extensive investigation and to provide production decisions [17]. So, on the basis of precise seed metering monitoring, it not only accurately detected abnormal seeding conditions of miss, multiples, and blocked seeding, but also recorded real-time seeding statistics and promptly reflected the seeding quality and working efficiency for users, thereby providing important production guidelines and economic benefits [22]. This kind of monitoring system with the characteristics of low-cost and efficiently will be one of the main scopes of precise agricultural machine research in the future.

In this study, a real-time monitoring and management system based on photoelectric sensors for soybean precision seed metering devices was presented, which could monitor hermetic seed tubes without a blind area and could trigger alarms for abnormal status. A statistical system was incorporated using moving speed, planting space, and sensor monitoring signals to calculate seeding quantity, seeding rate, and planting space in real time. The objective of this study was to evaluate the performance of seeding sensors and statistics systems, improving the level of seeding operation management.

2. Materials and Methods

2.1. System Components

Seed metering monitoring system was designed as shown in Figure 1. The proposed system mainly included a photoelectric sensors module for seeding signal collecting, Hall speed monitor module, microcontroller unit (MCU), sound–light alarm module, human–

machine interface (HMI), and communication parts. Hall sensors converted quantities of ground wheel rotation rate into pulse signal [23,24]. A human–machinery interface (HMI) was used to display data and for user input. An alarm circuit with sound and light and two types of prompt was designed to better transmit alerts to the users [25]. The required power supply was provided by a tractor battery after DC 12 V–DC 5 V voltage reduction.

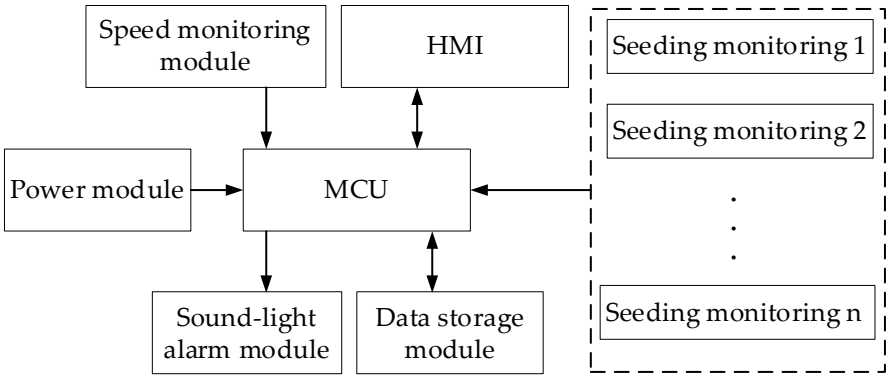


Figure 1. Structure diagram of the monitoring system of the seeder.

2.2. Working Principle

A convert circuit made the intermittent seed flow signals into electrical pulse signals. The time interval Δt (s) between two adjacent pulse signals reflected the time difference of falling seeds. Through time interval Δt and the moving speed of seeder tractor $v_{tractor}$ (m·s^{-1}), the actual grain planting space can be estimated.

In accordance with Chinese National Standard as GBT 6973–2005 (Chinese National Standard, 2005), the theoretical and actual grain spacing grain had to be compared. The formula judging the occurrence of multiples or miss seeding was as follows:

$$\begin{cases} v_{tractor}\Delta t \leq 0.5\overline{d}(multiplied\ seeding) \\ v_{tractor}\Delta t > 1.5\overline{d}(missed\ seeding) \end{cases} \quad (1)$$

where $\overline{d}$ was the theoretical planting space, cm.

To increase the accuracy of rotation measurement, some magnetic steels were installed on the ground wheel of seeder. Assume the number of magnetic steels was Z. Hall sensors were used to collect the rotation pulse signals of the ground wheel and to record the pulse number N in the speed measuring period T (s).

The slip ratio of the ground wheel was key to ground wheel speed measurement. The equation was:

$$\sigma = \frac{S - \pi Dn}{\pi Dn} \times 100 \quad (2)$$

where σ was the slip ratio of the ground wheel, %. In general, its range was 0.05–0.12; S was the actual forward distance of the ground wheel, m; D was the diameter of the ground wheel, m; n was the number of rotations of the ground wheel during S m.

The seeder speed $v_{tractor}$ (m·s^{-1}) was calculated as follows:

$$v_{tractor} = \frac{N}{Z}\frac{\pi D(1+\sigma)}{T} \quad (3)$$

The rotation rate of the seed metering disk ω (r·min^{-1}) was calculated by the equation:

$$\omega = 60i\frac{N}{Z}\frac{(1+\sigma)}{T} \quad (4)$$

where i was the transmission coefficient between the wheel and the seed metering rotation rate.

Assume that the seed metering device had m seed metering holes, the seed metering quantity in per min was Q, which was expressed as follows:

$$Q = m\omega = 60i\frac{mN}{Z}\frac{(1+\sigma)}{T} \tag{5}$$

Then, from Equations (3)–(5), the actual planting space Δt ' (s) of two adjacent seeds was approximately calculated as follows:

$$\Delta t' = \frac{60}{Q} = \frac{ZT}{i \times m \times N(1+\sigma)} \tag{6}$$

Subsequently, Δt 'derived from Equation (6) and the time interval of two falling seeds Δt monitored by the seed metering sensor can be used to cross-validate the monitoring system for multiples and miss seeding, thereby improving the sensitivity of the system in monitoring seeding status.

The seed metering number Q_i (1, 2, . . . , n) of each seed metering device in period t(s) can be accumulatively obtained through the signals of the seed metering sensors, and the total quantity of seed metering Q_{total} (grain) and the sowing rate of the seeder v_{sow_rate} (grain·s^{-1}) were listed as:

$$Q_{total} = \sum_{1}^{n} Q_i \tag{7}$$

$$v_{sow\_rate} = Q_{total}/t \tag{8}$$

The working area of the seeder was the product of the working distance and working width of the seeder. The application rate of the seeder $Q_{average}$ (grain·m^{-2}) can be deduced by the below equation:

$$Q_{average} = \frac{Q_{total}}{v_{tractor} \times t \times L} \tag{9}$$

where L was the working width of the seed metering devices, m. $Q_{average}$ reflected the working efficiency of the seed metering device and can estimate planting density [26,27].

2.3. Software Working Flow Design

The flow diagram of the monitoring system included system initialization, timing interruption, signal acquisition, sound–light alarm, and seeding statistics. The system software flow was shown in Figure 2.

The program was first initialized after the normal activation of the monitoring system, and its self-checking function was executed. The system automatically checked for the occurrence of any abnormality.

After self-checking, three types of data had to be input by HMI. The first was agronomic requirements, for example, pre-set seeding quantity and planting space. The second type was physical property of the seed tractor, such as ground wheel diameter, the parameters of slip ratio, transmission coefficient for ground wheel speed, and seed metering amount, respectively. The last one was alarm threshold. The range of threshold could be set to adjust the sensitivity of this system. If the range was smaller, more miss and multiples were detected and the more precise the monitoring system was. However, it would frequently lead to alarms to affect normal work. If the range was bigger, much abnormal status might be lost at seeding process. So, a suitable threshold range had to be input according to actual requirements. Before field working, the alarm threshold had to be tested at first. Then, the alarm worked when time interval Δt between two adjacent pulse signals of seeders were out of the range of threshold based on Formula (1). Otherwise, the buzzer did not alarm.

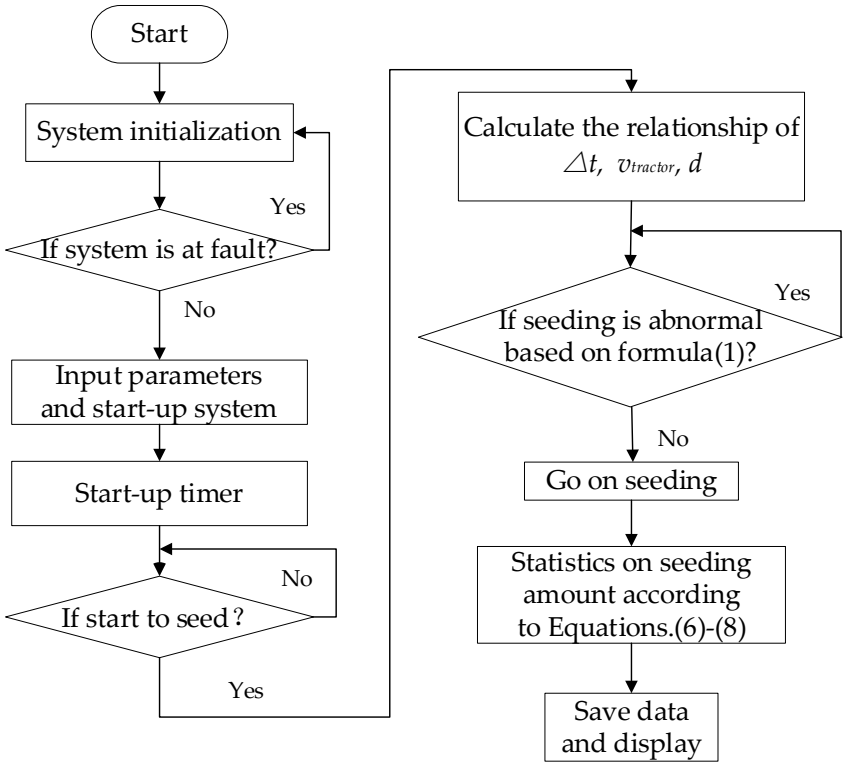


Figure 2. Monitoring system flow diagram.

After the start of the monitoring process and after receiving data from sensors, the microcontroller performs relevant calculations. It simultaneously monitored seed flowing, recorded the time interval between adjacent seeds, and detected miss seeding and multiples rate on the basis of the Equation (6). The system can also analyze whether the alarm coefficient exceeded thresholds according to Formula (1). Through the time interval Δt and the tractor moving speed $v_{tractor}$, the actual planting space was calculated. If actual planting space was less than half theoretical planting space, the system judged it was multiples. If actual planting space d was more than one and a half times of theoretical planting space, the status of miss seeding may occur and the sound–light alarm worked.

However, when the seeder should stop planting at the edge of the planting rows, the ground wheel speed was much smaller than usual speed and no seeds passed seed sensors, and the alarm was not needed.

Taking into account the input signals received from each seed tube and pre-registered settings by users, the seed metering statistics, such as total seeding quantity, seeding rate, miss index, and multiples index, were stored on the memory card and were also transferred through the serial port to HMI. They can provide performance information for users. Finally, the worker shut up the monitoring system and kept the data in the memory card for later analysis and decision.

2.4. Seed Metering Monitoring Sensor

Low-power infrared-emitting diodes and phototransistors are widely used for their characteristics of rapid response, high transmission efficiency, and high monitoring sensitivity. The light beam was concentrated, and its dust-penetrating ability is stronger than that of visible light.

Photoelectric sensors module was used by covering the seed metering tube with the emitting and receiving ends of three pairs of infrared light-emitting diodes and phototransistors. An infrared-emitting diode with a diameter of 10 mm emit infrared light with a wavelength of 850 nm and an emission angle of 60° [16]. The monitoring sensor was designed as a cuboid structure with a length, width, and height of 45 mm × 23 mm × 40 mm,

respectively, to enable the emitted light to cover the entire monitoring device without a blind area (Figure 3). In this study, after the finalization of the circuit and the design of a shell for the seed sensor, the shells were produced using a 3D printer. The material used should not let ambient light disturb the sensor inside, and these sensors should be installed where soybeans flew vertically to improve measurement accuracy.

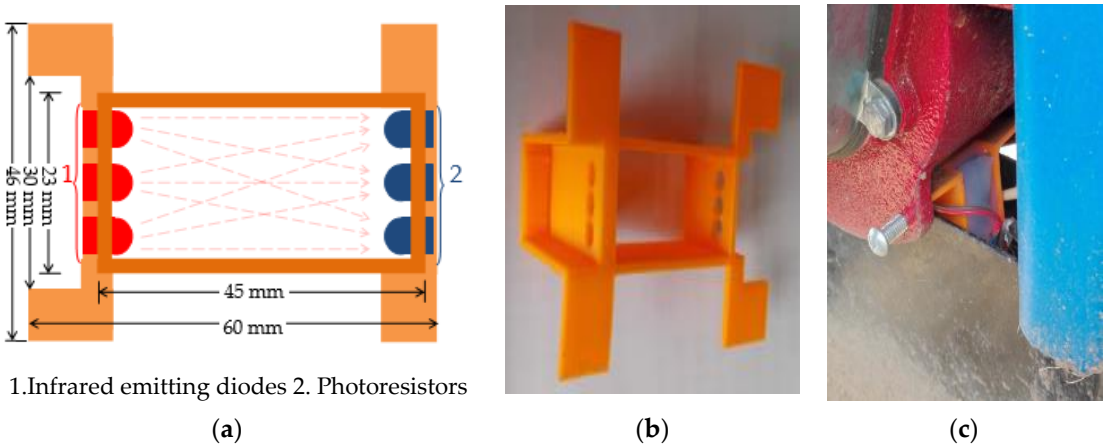


Figure 3. Photoelectric sensor design. (**a**) Schematic diagram; (**b**) a sensor shell produced by a 3D printer; (**c**) sensor installation.

During the seeding operation, seeds entered the seed metering tube and continuously passed through the monitoring sensors installed at the bottom of the seed metering devices (Figure 4).

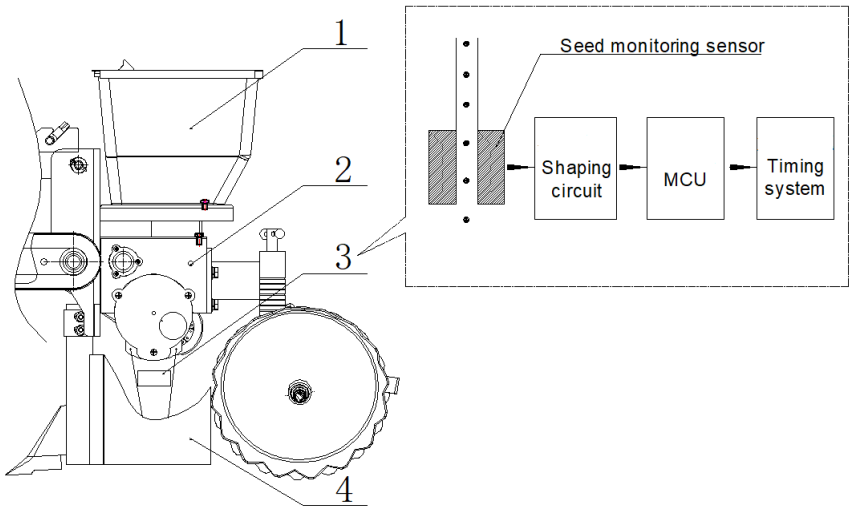


Figure 4. Schematic diagram of monitoring sensor position. (1). Seed box; (2). spoon-type seed metering device; (3). monitoring sensor; (4). ditching device side panel.

As shown in Figure 5, the seed metering monitoring circuit was primarily used to obtain information when seeds passed through the monitoring device. The monitoring circuit was primarily composed of an infrared-emitting diode, a phototransistor, a voltage comparator, and other devices. LM324 was used as a voltage comparator to consider the reliability and stability of the captured data. A sliding rheostat R4 was used to adjust the sensitivity of the sensor on the basis of the different precision required by the variety of seeds.

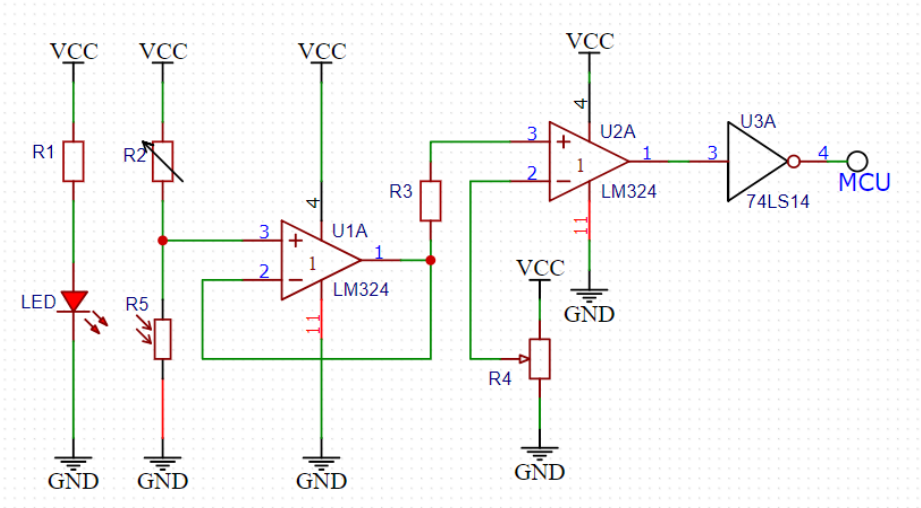


Figure 5. Sensor circuitry principal diagram.

When no seed passed through, all photodiodes were exposed to light at the same time. When there was a seed passing through, the photosensitive diode could not receive the light. At this time, the corresponding transistor was in the cut-off state, making the collector output a high level. Schmidt trigger 74LS14 was used to filter, reduce noise, and shape the signal at the collector output end to obtain square wave pulse, which was convenient for the calculation and processing of MCU.

3. Results

3.1. Monitoring Test for Sensitivity

To evaluate the monitoring sensitivity of the seed metering monitoring sensor, laboratory tests were conducted with two soybean varieties with different diameters [28]. The variety of small diameter was Dongnong 46, and its 100-grain weight was approximately 21.78 ± 0.40 g, with a diameter of 4.5 ± 0.2 mm. The other variety of large diameter was Dongnong 252, and its 100-grain weight was 28.28 ± 0.52 g, with a diameter of 5.5 ± 0.3 mm. Some researchers also focused on the sowing of various seeds with these types of physical properties [13].

The number of seeds was counted when seeding speed was stable. Seed metering devices 1 to 5 were filled with small seeds, and seed metering device 6 was filled with large seeds (Figure 6). The monitor results were recorded in Table 1.

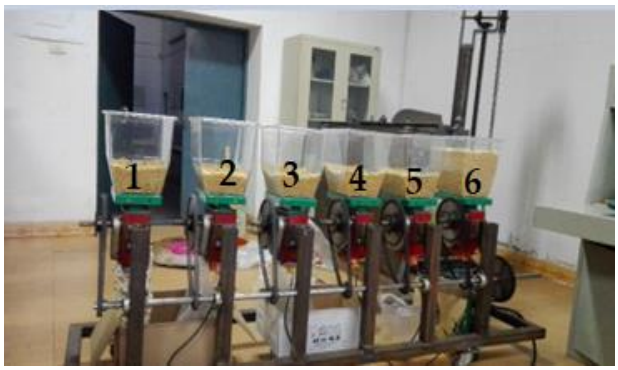


Figure 6. Seeding test platform.

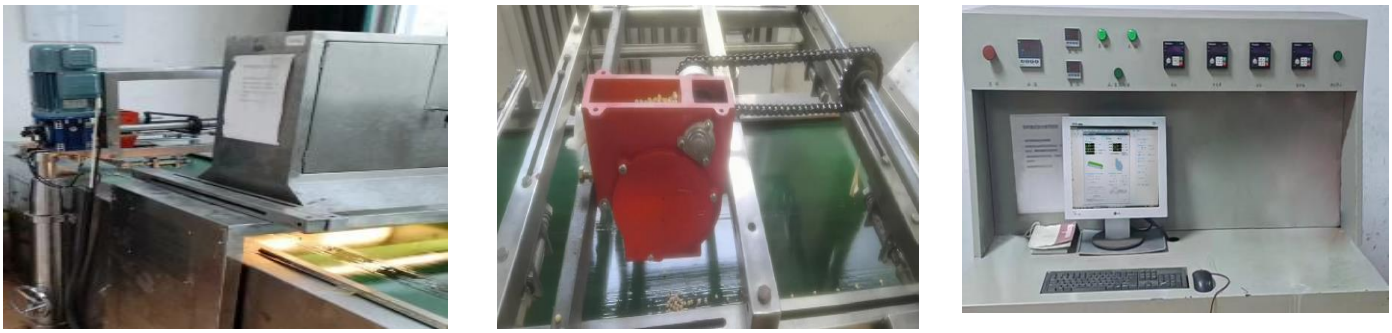
Table 1. Monitoring data for soybeans with different diameters.

Items	Small Seeds					Big Seeds
	Sensor 1	**Sensor 2**	**Sensor 3**	**Sensor 4**	**Sensor 5**	**Sensor 6**
Seeds quantity monitored/grain	11,473	11,553	11,472	11,432	11,519	11,000
Seeds weight/g	2540.00	2560.12	2539.24	2536.13	2561.63	3062.85
Seeds quantity calculated/grain	11,660	11,753	11,657	11,642	11,760	10,829
Error/%	1.6	1.7	1.6	1.8	2.0	1.6

As indicated in Table 1, the error of the monitoring sensors for the soybean variety with a small diameter was higher than that for large-diameter soybeans. This result was mostly attributed to the large-diameter soybean screening out infrared beam for a slightly longer time in the sensors. The signal was transmitted to the single-chip calculator, where it could be effectively identified and counted. It can be seen that the monitoring errors of both of them were less than 2.0% for both big and small soybeans. These errors may be caused by multiples in which two adjacent and nearly overlapping soybeans fell down at the same time. This phenomenon could not be precisely detected by this kind of sensor.

3.2. Static Monitoring Test

To evaluate the monitor performance, seed metering devices were installed on a seeding image monitor platform (SIMP) (Figure 7) to compare with the photoelectric sensor monitoring system [29,30].



(**a**) (**b**) (**c**)

Figure 7. Seeding image monitor platform (SIMP). (**a**) Camera-monitoring platform; (**b**) seed metering speed control bracket; (**c**) image-monitoring interface.

Soybean seeds (variety: Dongnong 46) were used. A "Shuang Fu" brand mechanical precision seed metering device, which had a vertical spoon wheel type with 33 holes, was selected. The inverter was used to control three rotation speeds of the seed metering wheel (9, 18, 27 ($r \cdot min^{-1}$)). The inverter speed was set to repeat five times at each rotation speed. The test data of multiples, miss of SIMP, and this monitoring system were recorded when speed was stable, and the average value of the five tests was used as the test result in Table 2. Flow rate was estimated by Equation (8), and application rate was estimated by Equation (9).

As shown in Table 2, the errors of the miss index between monitoring system and SIMP were less than 0.4% at three rotation speeds, and the errors of multiples index between them were less than 0.5%. This monitoring accuracy was nearly equal to article [1] based on laser sensors, in which the average monitoring error of the seeding qualified rate was less than 0.5% on a simulated seeder at a forward speed of 5–10 $km \cdot h^{-1}$.

Table 2. Statistics results of static experiment.

Speed/m·s^{-1}	Disk Rotation Rate/r·min^{-1}	Miss Index/%			Multiples Index/%			Flow Rate/Grain·s^{-1}	Application Rate/Grain·m^{-2}
		SIMP	Monitor System	Error	SIMP	Monitor System	Error		
0.50	9.00	1.20	1.10	0.10	2.30	2.10	0.20	28.31 ± 2.40	16.18
1.00	18.00	1.80	1.60	0.20	3.10	2.80	0.30	56.26 ± 5.52	16.07
1.50	27.00	2.50	2.10	0.40	4.40	3.90	0.50	85.17 ± 6.82	16.22

The seeding rate and planting space had a great influence on the monitoring accuracy of the sensor. After analysis, the results were found that the errors increased with an increase of rotation speed of seed metering disk. When flow rate reached a high value of 85.17 grain·s^{-1}, the error of miss and multiples index of this monitoring system were the biggest of these three speeds. Application rates were nearly unaffected, calculated by Equation (9).

3.3. Field Dynamic Monitoring Test

To test the monitoring system in the field [8,31], no-tillage machine seeding monitor tests were conducted on wheat stubble in Yongcheng City, Henan Province. Small-diameter soybean variety (Zhengdou 04024) was used in 2016, and big-diameter soybean (Hedou 33) was used in 2019. The 100-grain weight of Zhengdou 04024 was 21.3 ± 0.09 g, with the diameter 4.9 ± 0.4 mm after sampling measurement. The 100-grain weight of Hedou 33 was 24.2 ± 0.11 g, with the diameter 5.7 ± 0.5 mm.

Experimental field was without rotary tillage or plow tillage. A John Deere 954 tractor equipped with a six-row seeder had a working width of 3.80 m (Figure 8). "Shuang Fu" brand mechanical precision seed metering devices, which had vertical spoon wheel type with 33 holes, were used. Sowing rate was adjusted by mechanical transmission device of gears. Gear I, II, and III corresponded to the planting space of 5 cm, 8 cm, and 10 cm. The tractor ran at a speed about 5 km·h^{-1} (1.38 m·s^{-1}) in these two years. The higher speed of the tractor led to more axis dropped out. A straw-mulching clearing device was equipped on a no-tillage planter in 2019, which was helpful to decrease blockage of seeding tubes. However, the complex field condition especially, in a no-tillage field, most affected the performance of these optical sensors. It is impossible to completely eliminate missing inspection. This was one of main obstacles for improving the accuracy of the monitoring system [20].

Figure 8. Field test.

The monitoring sensors were installed under the seed metering devices (Figure 3). Sowing rate (estimated by Equation (8)) and application rate (estimated by Equation (9)) were recorded in Table 3. Abnormal phenomenon records were shown in Table 4.

Table 3. Dynamic experimental results of the digital monitoring system.

Gear No.	Planting Space/cm	Small Diameter (2016)		Big Diameter (2019)	
		Sowing Rate/Grain·s^{-1}	Application Rate/Grain·m^{-2}	Sowing Rate/Grain·s^{-1}	Application Rate/Grain·m^{-2}
I	5.0	160.43	32.29	172.2	32.84
II	8.0	100.58	20.24	109.18	20.82
III	10.0	80.73	16.25	85.48	16.30

Table 4. Abnormal phenomenon statistics in field tests.

Gear No.	Small Diameter (2016)						Big Diameter (2019)					
	Miss	Multi Ples	Blockage	Quantity/ Grain	Miss Index/%	Multiples Index/%	Miss	Multi Ples	Blockage	Quantity/ Grain	Miss Index/%	Multiples Index/%
I	1986	2134	5	63,650	3.12	3.36	1590	1794	4	52,360	3.04	3.42
II	1846	1952	5	63,210	2.92	3.09	1443	1726	2	53,009	2.72	3.26
III	1583	1722	6	63,300	2.50	2.72	1237	1546	3	54,132	2.29	2.86

As shown in Table 3, planting space continuously increased with an increase of seed metering gear. Meanwhile, sowing rate and application rate (planting density) decreased. Through analysis, flow rate and application rate in 2016 were a little smaller than 2019 (Table 3). Meanwhile, the percent of multiples in 2019 was a little larger than 2016 (Table 4). That meant a greater percentage of big-size soybeans was detected under the same theoretical sowing rate. Through the monitor, miss index of big-size soybean was a little lower than small-size ones. Miss index and multiples index continuously decreased with a decrease of metering devices speed, and it did not exceed 3.12% and 3.42%, respectively. It can be seen that this kind of soybean seed metering device had a good seeding performance at these working conditions.

After cleaning out the impurities, the number of blockage alarms in 2019 was obviously lower than 2016 (Table 4).

After the field test, three major phenomena had to been stated: (1) Seeds in the seed boxes contained many impurities, which affected the filling effect of the seed metering holes. Light and sound alarms occurred when miss seeding. (2) The uneven ground in the test field prevented the tractor from moving at a uniform speed. When seed metering speed was occasionally fast and slow, the statistical value of seed metering fluctuated a lot. (3) The dust amount of no-tillage seeding in the wheat stubble field was considerable, which affected the accuracy of the monitoring system. The self-clean or dust-reduction device should be employed on seeders.

4. Conclusions

(1) The seed metering monitoring sensor used three pairs of infrared light-emitting diodes and phototransistors as the emitting and receiving ends of the photoelectric sensor, which could achieve the blind area free monitoring of the soybean seed metering tube. This sensor was simple, practical, and low-cost.

(2) In laboratory tests, the monitoring of soybeans with big and small diameters was basically consistent, and the monitoring errors of seeder quantity was less than 2.0%. The errors of miss and multiples index between monitoring system and SIMP were less than 0.4% and 0.5%, respectively, at three rotation speeds. This demonstrates that this kind of sensors have a good monitoring performance.

(3) This monitoring system can evaluate the precision of the performance of seed metering device in no-tillage machine seeding tests and verify that the big-diameter seeds can be detected more precisely than small ones.

This real-time monitoring and management system of soybean precision seed metering devices was helpful to prevent large-scale miss seeding and collect and record the quantity and quality information of seeding process.

Author Contributions: Conceptualization, W.J. and S.H.; methodology, J.Z.; software, S.Y.; validation, P.Z. and Y.H.; formal analysis, J.Z.; data curation, J.Z.; writing—original draft preparation, P.Z. and C.C.; writing—review and editing, J.Z. and C.C.; project administration, J.Z.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Foundation of Heilongjiang Province, grant number LH2020E002, the Department of Science and Technology of Jilin Province, grant number 20200402116NC and Soybean Industrial Technology System of China, grant number CARS-04-PS27.

Data Availability Statement: Not applicable.

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

References

1. Xie, C.; Zhang, D.; Yang, L.; Cui, T.; He, X.; Du, Z. Precision seeding parameter monitoring system based on laser sensor and wireless serial port communication. *Comput. Electron. Agric.* **2021**, *190*, 106429. [CrossRef]
2. Zhang, J.L.; Geng, Y.; Guo, F.; Li, X.G.; Wan, S.B. Research progress on the mechanism of improving peanut yield by single-seed precision sowing. *J. Integr. Agric.* **2020**, *19*, 1919–1927. [CrossRef]
3. Nikolay, Z.; Nikolay, K.; Gao, X.; Li, Q.; Mi, G.; Huang, Y. Design and testing of novel seed miss prevention system for single seed precision metering devices. *Comput. Electron. Agric.* **2022**, *198*, 107048. [CrossRef]
4. Xie, C.; Yang, L.; Zhang, D.; Cui, T.; He, X.; Du, Z.; Xiao, T. A signal output quantity (SOQ) judgment algorithm for improving seeding quantity accuracy. *Comput. Electron. Agric.* **2022**, *201*, 107321. [CrossRef]
5. Kachman, S.D.; Smith, J.A. Alternative measures of mccuracy in plant spacing for planters using single seed metering. *Trans. ASAE* **1995**, *38*, 379–387. [CrossRef]
6. Wang, G.; Sun, W.; Zhang, H.; Liu, X.; Li, H.; Yang, X.; Zhu, L. Research on a kind of seeding-monitoring and compensating control system for potato planter without additional seed-metering channel. *Comput. Electron. Agric.* **2020**, *177*, 105681. [CrossRef]
7. Hou, S.; Zhu, Y.; Zhu, X.; Wang, Y.; Ji, W.; Chen, H. Design and experiment of a straw clearing mulching no-tillage planter. *Biosys. Eng.* **2022**, *221*, 69–80. [CrossRef]
8. Kamgar, S.; Noei-Khodabadi, F.; Shafaei, S.M. Design, development and field assessment of a controlled seed metering unit to be used in grain drills for direct seeding of wheat. *Inf. Process. Agric.* **2015**, *2*, 169–176. [CrossRef]
9. Sharipov, G.M.; Paraforos, D.S.; Pulatov, A.S.; Griepentrog, H.W. Dynamic performance of a no-till seeding assembly. *Biosys Eng.* **2017**, *158*, 64–75. [CrossRef]
10. Karimi, H.; Navid, H.; Besharati, B.; Behfar, H.; Eskandari, I. A practical approach to comparative design of non-contact sensing techniques for seed flow rate detection. *Comput. Electron. Agric.* **2017**, *142*, 165–172. [CrossRef]
11. Bai, J.; Hao, F.; Cheng, G.; Li, C. Machine vision-based supplemental seeding device for plug seedling of sweet corn. *Comput. Electron. Agric.* **2021**, *188*, 106345. [CrossRef]
12. Xie, C.; Yang, L.; Zhang, D.; Cui, T.; Zhang, K.; He, X.; Du, Z. Design of smart seed sensor based on microwave detection method and signal calculation model. *Comput. Electron. Agric.* **2022**, *199*, 107178. [CrossRef]
13. Besharati, B.; Navid, H.; Karimi, H.; Behfar, H.; Eskandari, I. Development of an infrared seed-sensing system to estimate flow rates based on physical properties of seeds. *Comput. Electron. Agric.* **2019**, *162*, 874–881. [CrossRef]
14. Bao, J.; Xie, Q. Artificial intelligence in animal farming: A systematic literature review. *J. Clean. Prod.* **2022**, *331*, 129956. [CrossRef]
15. Hao, X.Z.; He, X.P.; Zhou, Y.; Li, X.; He, R.Y. Research on the sowing performance monitoring system for precision seeders based on photoelectric senso. *J. South China Agric. Univ.* **2017**, *38*, 120–124. (In Chinese)
16. Zhang, J.C.; Chen, H.T.; Ouyang, B.L.; Ji, W.Y. Monitoring system for precision seeders based on a photosensitive sensor. *J. Tsinghua Univ. (Sci. Technol.)* **2013**, *53*, 265–268. (In Chinese)
17. Wu, C.; Li, D.; Zhang, X.; Pan, J.; Quan, L.; Yang, L.; Yang, W.; Ma, Q.; Su, C.; Zhai, W. China's agricultural machinery operation big data system. *Comput. Electron. Agric.* **2023**, *205*, 107594. [CrossRef]
18. Du, R.C.; Gong, B.C.; Liu, N.N.; Wang, C.; Yang, Z.; Ma, M. Design and experiment on intelligent fuzzy monitoring system for corn planters. *Int. J. Agric. Biol. Eng.* **2013**, *6*, 11–18.
19. Zhao, B.; Kuang, L.H.; Zhang, W. Seed and fertilizer intelligent gauging and monitoring system of suction precision seeder. *Trans. CSAE* **2010**, *26*, 147–153. (In Chinese)
20. Karimi, H.; Navid, H.; Besharati, B.; Eskandari, I. Assessing an infrared-based seed drill monitoring system under field operating conditions. *Comput. Electron. Agric.* **2019**, *162*, 543–551. [CrossRef]
21. Yu, H.; Ding, Y.; Fu, X.; Liu, H.; Jin, M.; Yang, C.; Liu, Z.; Sun, G.; Dou, X. A solid fertilizer and seed application rate measuring system for a seed-fertilizer drill machine. *Comput. Electron. Agric.* **2019**, *162*, 836–844. [CrossRef]

22. Yazgi, A.; Degirmencioglu, A. Measurement of seed spacing uniformity performance of a precision metering unit as function of the number of holes on vacuum plate. *Measurement* **2014**, *56*, 128–135. [CrossRef]
23. Mellet, D.S.; du Plessis, M.D. A novel CMOS Hall effect sensor. *Sens. Actuators A Phys.* **2014**, *211*, 60–66. [CrossRef]
24. Banjevic, M.; Furrer, B.; Blagojevic, M.; Popovic, R.S. High-speed CMOS magnetic angle sensor based on miniaturized circular vertical Hall devices. *Sens. Actuators A Phys.* **2012**, *178*, 64–75. [CrossRef]
25. Zhai, J.B.; Xia, J.F.; Zhou, Y.; Zhang, S. Design and experimental study of the control system for precision seed-metering device. *Int. J. Agric. Biol. Eng.* **2014**, *7*, 13–18.
26. Qi, J.T.; Jia, H.L.; Yang, L.; Yu, H.B.; Liu, X.H.; Lan, Y.B.; Feng, X.Z.; Yang, Y.X. Design and test of fault monitoring system for corn precision planter. *Int. J. Agric. Biol. Eng.* **2015**, *8*, 13–19.
27. Sun, X.; Li, H.; Qi, X.; Nyambura, S.M.; Yin, J.; Ma, Y.; Wang, J. Performance parameters optimization of a three-row pneumatic precision metering device for brassica chinensis. *Agronomy* **2022**, *12*, 1011. [CrossRef]
28. St. Jack, D.; Hesterman, D.C.; Guzzomi, A.L. Precision metering of Santalum spicatum (Australian Sandalwood) seeds. *Biosys Eng.* **2013**, *115*, 171–183. [CrossRef]
29. Jin, M.F.; Ding, Y.Q.; Yu, H.F.; Liu, H.T.; Jiang, Y.Z.; Fu, X.Q. Optimal structure design and performance tests of seed metering device with fluted rollers for precision wheat seeding machine. *IFAC-PapersOnLine* **2018**, *51*, 509–514.
30. Siponen, M.; Seilonen, I.; Brodie, S.; Oksanen, T. Next generation task controller for agricultural machinery using OPC unified architecture. *Comput. Electron. Agric.* **2022**, *203*, 107475. [CrossRef]
31. Abdolahzare, Z.; Mehdizadeh, S.A. Real time laboratory and field monitoring of the effect of the operational parameters on seed falling speed and trajectory of pneumatic planter. *Comput. Electron. Agric.* **2018**, *145*, 187–198. [CrossRef]

agronomy

MDPI

Article

Maize Leaf Disease Identification Based on YOLOv5n Algorithm Incorporating Attention Mechanism

Li Ma [1], Qiwen Yu [1], Helong Yu [1] and Jian Zhang [2,3,*]

[1] College of Information Technology, Jilin Agricultural University, Changchun 130118, China
[2] Faculty of Agronomy, Jilin Agricultural University, Changchun 130118, China
[3] Department of Biology, University of Columbia Okanagan, Kelowna, BC V1V 1V7, Canada
* Correspondence: jian.zhang@ubc.ca

Abstract: Maize diseases are reported to occur often, and are complicated and difficult to control, which seriously affects the yield and quality of maize. This paper proposes an improved YOLOv5n model incorporating a CA (Coordinate Attention) mechanism and STR (Swin Transformer) detection head, CTR_YOLOv5n, to identify common maize leaf spot, gray spot, and rust diseases in mobile applications. Based on the lightweight model YOLOv5n, the accuracy of the model is improved by adding a CA attention module, and the global information acquisition capability is enhanced by using TR2 as the detection head. The average recognition accuracy of the algorithm model can reach 95.2%, which is 2.8 percent higher than the original model, and the memory size is reduced to 5.1MB compared to 92.9MB of YOLOv5l, which is 94.5% smaller and meets the requirement of being light weight. Compared with SE, CBAM, and ECA, which are the mainstream attention mechanisms, the recognition effect we used is better and the accuracy is higher, achieving fast and accurate recognition of maize leaf diseases with fewer computational resources, providing new ideas and methods for real-time recognition of maize and other crop spots in mobile applications.

Keywords: deep learning; attention mechanism; maize leaf disease; digital agriculture

Citation: Ma, L.; Yu, Q.; Yu, H.; Zhang, J. Maize Leaf Disease Identification Based on YOLOv5n Algorithm Incorporating Attention Mechanism. *Agronomy* **2023**, *13*, 521. https://doi.org/10.3390/agronomy13020521

Academic Editor: Paul Kwan

Received: 5 December 2022
Revised: 27 December 2022
Accepted: 3 February 2023
Published: 11 February 2023

1. Introduction

Because it is one of the three main food crops and a significant source of revenue for many farmers cross the world, maize, which has a high nutritional value, continues to play a significant role in addressing the issue of human food supply today [1]. Data show that 60% of maize in China is used as feed for livestock and poultry industries, 30% is used for industrial purposes, such as chemical, pharmaceutical, and paper making, and the remaining 10% is used for direct consumption by people. Mazie occupies an important position in the agricultural production and economic development of China. It is evident that, together with rice and wheat, maize will be the food crop with the largest production demand in the future. Therefore, increasing maize production and maintaining high quality is important to China's agricultural industry. Among the many factors affecting maize production, the problem of maize pests and diseases has the greatest negative impact on its production and quality, and once maize pests and diseases occur, they can cause varying degrees of yield reduction and quality decline, seriously affecting the economic benefits of producers and the industry as a whole [2].

At present, the category identification of maize diseases in China is based on the empirical judgment of crop pathologist experts in the field and technicians specialized in plant protection; therefore, technicians need to have good observation skills and rich experience to accurately identify the category of diseases [3]. This traditional disease identification method, which relies on individual experience, has a large limitation. Additionally, when there are too many samples to test with many different disease types, subsequently, there is a higher chance of inaccuracy in the identification process due to human errors.

Recent years, with the high-speed development of big data analysis technology and GPUs (Graphics Processing Units), the computing power of computers has been improved, and deep learning techniques have been developed rapidly and have been used in many applications such as agricultural pests and diseases [4]. Yinglai Huang et al. [5] replaced 7×7 convolutional kernels in the first convolutional layer of the conventional ResNet-50 model with three 3×3 convolutional kernels; they used the LeakyReLU activation function instead of the ReLU activation function and changed the order of the batch normalization layer, activation function, and convolutional layer in the residual block. The improved network obtained a 98.3% correct rate in maize leaf disease image classification. Haoyu Wu [6] proposed to construct a two-channel convolutional neural network based on VGG and ResNet. By adjusting the parameters of the two-channel convolutional neural network, the accuracy of identifying maize leaf disease types in the validation set can reach 98.33%, while the VGG model can reach 93.33%. Chao Wang et al. [7] proposed a method based on ResNet (Residual Neural Network) deep learning network for maize disease recognition, using ResNet as the main model for maize disease recognition, and found that the highest classification accuracy of 92.82% was obtained with ResNet50 at a batch size of 32 and epoch number of 16.

Azlah, M.A.F. et al. [8] mainly reviewed the advantages of each classifier and compared their compatibility with different leaf features recognition process. Koklu, M. et al. [9] conducted a deep learning-based classification by using images of grapevine leaves. The most successful method was obtained by extracting features from the Logits layer and reducing the feature with the chi-squares method. The most successful SVM kernel was Cubic. The classification success of the system has been determined as 97.60%. Argüeso, D. et al. [10] introduced Few-Shot Learning (FSL) algorithms for plant leaf classification using deep learning with small datasets. The FSL method outperformed the classical fine-tuning transfer learning, which had accuracies of 18.0 (16.0–24.0)% and 72.0 (68.0–77.3)% for 1 and 80 images per class, respectively.

Although there are many recognition techniques based on deep learning technology and all of them work well, there are some problems among them, such as less small-scale target data, larger memory consumption of the model, and being unsuitable for mobile deployment.

This paper, therefore, investigates the problem of disease in maize leaves, applying the current deep learning technology to design an experimental study in the hope that farmers will be able to rely on their mobile phones in the field to identify diseases on maize in a timely and effective manner, thus alleviating the problems of reduced yields and reduced quality of maize. As the ultimate goal of our research is to help farmers to identify maize diseases in real-time in the field with a mobile device on their person, the light weight and high accuracy of the model is the focus of this paper.

Currently, the commonly used target detection networks include Faster R-CNN [11], SSD [12], YOLO series [13–16], etc. Among them, the YOLO network model belongs to a one-stage target detection algorithm with a simple structure, small computation, and fast operation speed, which is widely used in crop disease identification research. Among them, YOLOv5n is the latest model of the YOLOv5 series network, which has the advantages of high detection accuracy, fast inference speed, and small storage space, and is suitable for deployment in mobile for real-time detection. In this paper, we propose a regional detection model for maize leaf diseases based on YOLOv5n: CTR_YOLOv5n, which accelerates the model convergence speed, improves the model generalization ability, and enhances the recognition accuracy and detection speed of the model, taking three common maize leaf diseases, blotch disease, gray spot, and rust, as the research objects.

2. Materials and Methods

2.1. Construction of the Data Set

The target dataset used for the experiments in this paper is maize leaf diseases. Among the various maize leaf diseases, the most common and representative ones are blotch

disease, gray spot, and rust, and images of the three leaf diseases, as well as images of healthy maize leaves, are used as the dataset.

(a) Maize blotch disease

The disease produces spots on the leaves, such as long rhombus shape. The color is generally brown or yellow-brown, and the rhombus-shaped spots are generally 5–10 cm long and 1 cm wide, approximately. The disease will gradually expand when the disease is serious, and even lead to leaf death.

(b) Maize rust

The disease mainly occurs on maize leaves, on both sides of the leaf, and causes scattered or aggregated growth of round, yellow-brown, powdery spots and scattered rust-colored powder, that is, the summer spore mounds and summer spores of the pathogenic bacteria. Later in the season, round, black winter spore mounds and winter spores grow on the spots.

(c) Maize gray spot disease

The disease initially forms oval to rectangular gray to light brown spots on the leaf surface without obvious margins, turning brown later. The spots are mostly confined between parallel leaf veins and are 4 to 20 × 2 to 5 (mm) in size. When the humidity is high, the abaxial surface of the spot produces gray moldy material, that is, the conidiophore and conidia of the disease.

(d) Healthy maize leaves

Leaf blade flattened and broad, leaf sheath with transverse veins; ligule membranous, about 2 mm long; linear-lanceolate, base rounded auriculate, glabrous or blemished pilose, midrib stout, margin slightly scabrous.

In the process of target detection model training, dataset production and image annotation are two very important steps. It is the foundation of the dataset and can be directly related to the reliability of the experiment, while the accuracy of the image annotation directly affects the training effect and the accuracy of the test.

The sample dataset of maize leaf disease images selected and used in this paper is mainly collected from the open-source website PlantVillage (https://tensorflow.google.cn/datasets/catalog/plant_village accessed on 13 June 2022) for three common maize leaf diseases and healthy maize leaf images. The total number of datasets is 4353, in which maize maculatus, maize rust, and maize gray spot are the three common diseases of maize leaves listed in this paper. The three diseases and healthy leaves are shown in Figure 1.

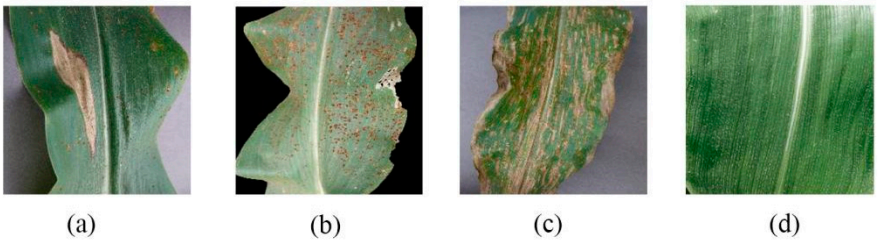


Figure 1. Examples of maize leaf diseases. (**a**) Maize blotch disease; (**b**) Maize rust; (**c**) Maize gray spot disease; (**d**) Healthy maize leaves.

The number of maize leaf data sets is shown in Table 1.

Because of the small number of large spots and gray spots, this paper additionally takes the data set of maize leaves taken from the pear test field and expands the number of data sets of gray spot and large spots by cutting out 200 sheets of maize leaves with gray spot and 200 sheets of maize leaves with the large spot from the data set taken from the pear test field and filling them into the original data set.

Table 1. Number of maize leaf datasets.

Leaf Type	Number of Images
Maize blotch disease	1000
Maize rust	1191
Maize gray spot disease	1000
Healthy maize leaves	1162

The image annotation tool used in this paper is Make Sense (https://www.makesense.ai/ accessed on 13 June 2022), an online annotation tool recommended by the authors of YOLOv5n to annotate the images, which can directly output YOLO format label files and can be directly applied to the YOLOv5n network.

The sorted data set picture files and label files were divided into the training dataset, the validation dataset, and the test dataset according to the ratio of 6:2:2, and then put into the network model for subsequent data enhancement and model training.

2.2. Data Augmentation

When we want to obtain a well-performing neural network model, we must have a large amount of data to support it, but it takes a lot of time and labor to obtain new data. If we use data augmentation [17], we can use the computer to create new data to increase the number of training samples, for example, by changing color brightness, hue saturation, scaling, rotation, panning, cropping, perspective transformation, etc., and adding some appropriate noise data to improve the model generalization.

In the YOLOv5n network model described in this paper, not only are some basic data enhancement methods included, but also the Mosaic data enhancement [18] is used, whose main idea is to select four images from the used dataset, crop and scale them randomly, and then arrange them randomly to form a new image. This has the advantage of increasing the number of datasets while augmenting the number of small sample targets, and it improves the training speed of the model. The flowchart of Mosaic data enhancement is shown in Figure 2.

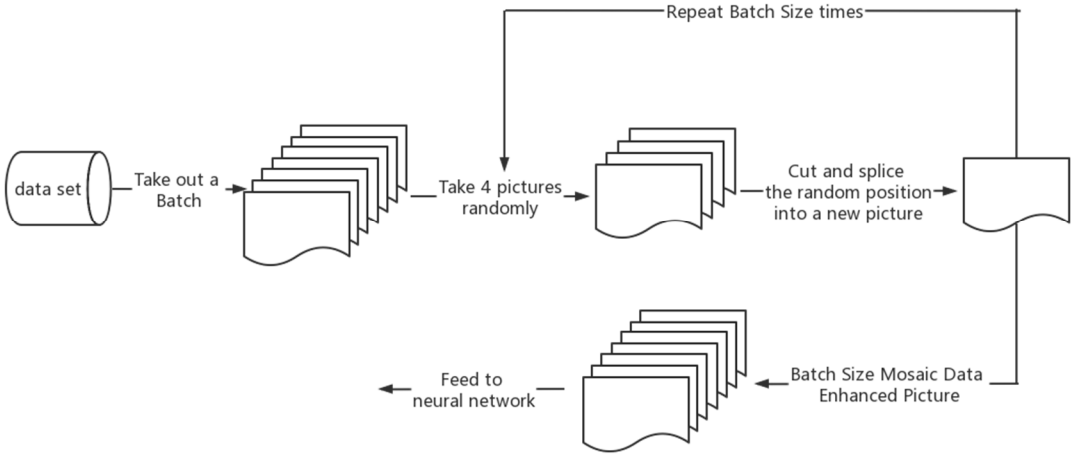


Figure 2. Mosaic data enhancement flow chart.

Mosaic data enhancement utilizes four images, which enriches the background of the detected objects and calculates the data of four images at once when BN calculates, so that the mini-batch size does not need to be large, and then a GPU can achieve better results.

In practice, Mosaic data enhancement first removes one batch of data from the total data set, takes out four images at random from it each time, crops and splices them at random positions, synthesizes new images, repeats the batch size several times, and finally gets a new batch size of one batch of images after mosaic data enhancement, then feeds to the neural network for training.

When cropping and splicing the images, the four randomly obtained images are cropped by a randomly positioned crosshair, and the corresponding parts are taken for splicing. At the same time, the target frame of each original image is limited by the crosshair crop, and will not exceed the original crop range. The implementation of Mosaic data enhancement in practice is shown in Figure 3.

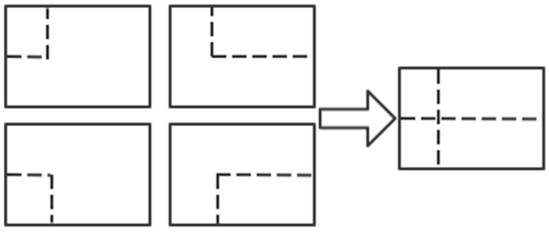


Figure 3. Implementation of Mosaic data enhancement in practice.

Mosaic has the following advantages: increases data diversity; randomly selects four images for combination; the number of images obtained from the combination is more than the number of original images; enhances model robustness; mixes four images with different semantic information; allows the model to detect targets beyond the conventional context; and enhances the effect of batch normalization. When the model is set to BN operation, the training will increase the total number of samples (BatchSize) as much as possible, because the BN principle is to calculate the mean and variance of each feature layer; if the total number of samples is larger, then the mean and variance calculated by BN will be closer to the mean and variance of the whole dataset, and the better the effect. The Mosaic data enhancement algorithm is helpful to improve the performance of small target detection. The enhanced images are stitched together from four original images, so that each image has a higher probability of containing small targets.

The operation principle of Mosaic data enhancement is equivalent to passing in four images for learning at one time during the training process, increasing the number of single training samples and target diversity, improving network training convergence speed and detection accuracy, and reducing large samples to small samples randomly, increasing the number of small-scale targets. Since the target of this paper is maize leaf disease and the disease spot is a small-scale target, Mosaic data enhancement provides important help for this study. Figure 4 shows 16 examples of data enhanced by Mosaic data. The file name in the picture is the file name of the image data involved in data enhancement. In the example, the file name is only for demonstration and will not be integrated into the picture to affect the subsequent model training. The colored boxes in the figure are identification boxes, where 0 indicates gray spot, 1 indicates rust, 2 indicates healthy maize leaves, and 3 indicates large spot disease. As shown in Figure 4.

To verify that the Mosaic data enhancement is real and effective for the experimental effect, this paper compares the parameters of the YOLOv5n network model with and without Mosaic data enhancement. The effect is shown in Figure 5.

From the Figure 5, it can be seen that the accuracy of the model is significantly and substantially improved after adding Mosaic data augmentation, and the convergence of the model is significantly improved compared to that without Mosaic data augmentation. At the same time, it can be seen that due to the Early Stopping method in the YOLOv5n model, which can resist overfitting, the model terminates early after 252 iterations in the training curve without the Mosaic data augmentation, because the accuracy no longer improves.

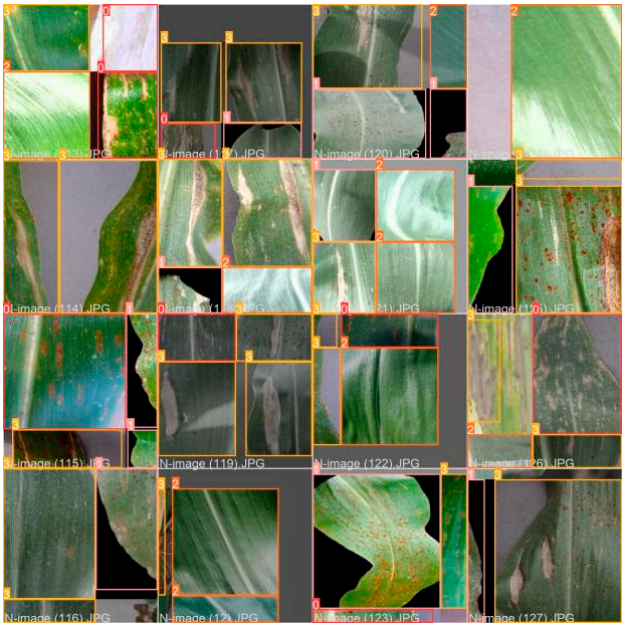
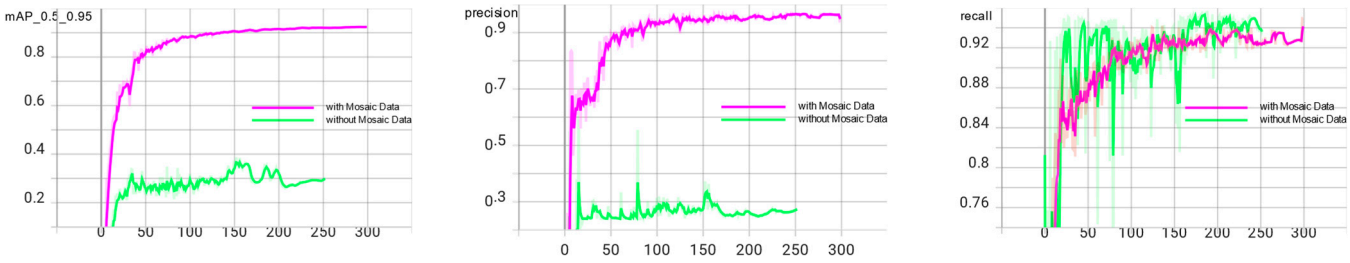


Figure 4. Sixteen examples of Mosaic data enhancement.



(**a**) (**b**) (**c**)

Figure 5. Comparison of enhancement effects with and without Mosaic data. (**a**) mAP_0.5_0.95 curve chart; (**b**) precision curve chart; (**c**) recall curve chart.

2.3. YOLOv5 Network Model

The YOLOv5 target detection algorithm is the 5th version of YOLO, whose core idea is to use the whole map as the input of the network and regress the location coordinates and category of the target directly in the output layer, which is characterized by high detection accuracy and fast detection speed to meet the demand of real-time monitoring.

The YOLOv5 network has been updated with five versions, YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x in order, with similar network structures and changes in network depth and width of feature maps based on YOLOv5s. Its accuracy and inference speed follow, of which YOLOv5n is in 2021 October after YOLOv5 update version 6.0, which has the advantage of being the fastest and the smallest model size compared to other versions. The ultimate goal is to deploy the model to mobile for real-time detection. To meet the lightweight requirement, the final study of this paper decided to use the YOLOv5n detection model with the lowest complexity to reduce the model storage footprint and increase the recognition speed.

The YOLOv5n algorithm consists of four parts: input, backbone, neck, and prediction [19]. Among them, Mosaic data enhancement is beneficial for detecting small targets and is suitable for leaf disease identification in this paper. The adaptive image scaling operation fixes images of different sizes to 640 pixels × 640 pixels as input. In the backbone network, YOLOv5n mainly uses the Conv module CSP structure and SPPF module. The feature fusion stage mainly borrows the idea from PANet [20]. The FPN (Feature Pyramid Network) and PAN (Path Aggregation Network) are borrowed to form the FPN + PAN

structure. The prediction output continues the previous idea of YOLO by outputting three sizes of prediction maps at the same time, which are suitable for detecting small, medium and large targets. The network structure of YOLOv5n is shown in Figure 6.

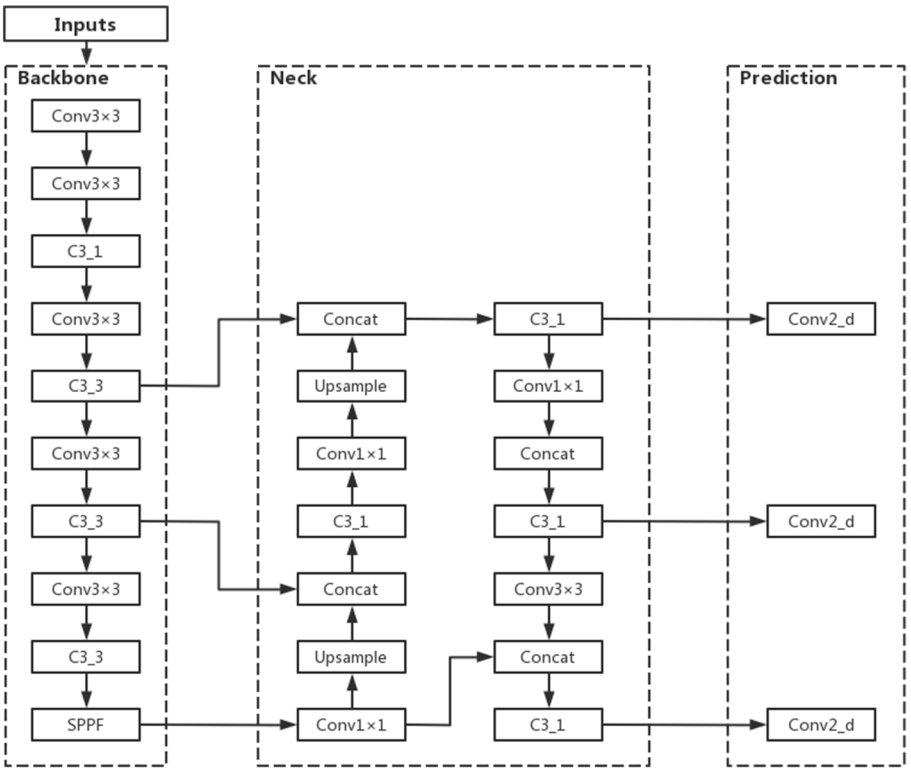


Figure 6. YOLOv5n network structure.

2.4. Improvements to the YOLOv5n Model

2.4.1. Adding CA to Improve Model Accuracy

In the task of maize leaf disease detection, since the disease spots occupy relatively few pixels of the image, their feature information is easily lost in the deep network, resulting in errors such as the wrong detection and missed detection. At this point, it would be more beneficial for the network model to recognize the images if the unsupervised network can automatically acquire the ability to focus on smaller pixel blocks. Therefore, this paper introduces the CA (Coordinate Attention) mechanism [21] in the YOLOv5n backbone network, which is used to tell the model "what" and "where," and which has been widely studied and deployed to improve the performance of neural networks. The use of lightweight attention modules can improve the network's ability to extract features from maize leaf spots while saving parameters.

For other channel attentions, they are taken to transform the input into individual feature vectors by 2D global pooling. The general idea of Coordinate Attention used in this paper is to decompose channel attention into two 1D feature encodings of aggregated features along different directions in the H-direction as well as the W-direction, that is, into $C \times H \times 1$ and $C \times 1 \times W$. CA This idea has the advantage of capturing long-range dependencies along one spatial direction while retaining accurate location information along the other spatial direction. After that, the generated feature maps are encoded separately, resulting in two direction-aware, as well as position-sensitive, feature maps, which can be complementarily applied to the input feature maps to enhance the representation of the target of interest. The two directional feature maps are then Concept spliced and then fed into a shared convolution to reduce the dimensionality to C/r, after which they are separated and allowed to Sigmoid in different directions to obtain the coefficients and then multiplied together. Finally, the feature map is obtained.

After adding CA attention to the YOLOv5n backbone network [22], keeping the parameters unchanged, the model is trained again, and the trained model has significantly improved the effect compared with the original model; the average accuracy mean value is increased from 0.924 to 0.948, and the model size is not significantly increased, which meets the requirement of being lightweight.

In this paper, after adding CA attention to the YOLOv5n backbone network, the specific structure of the backbone network is shown in Figure 7.

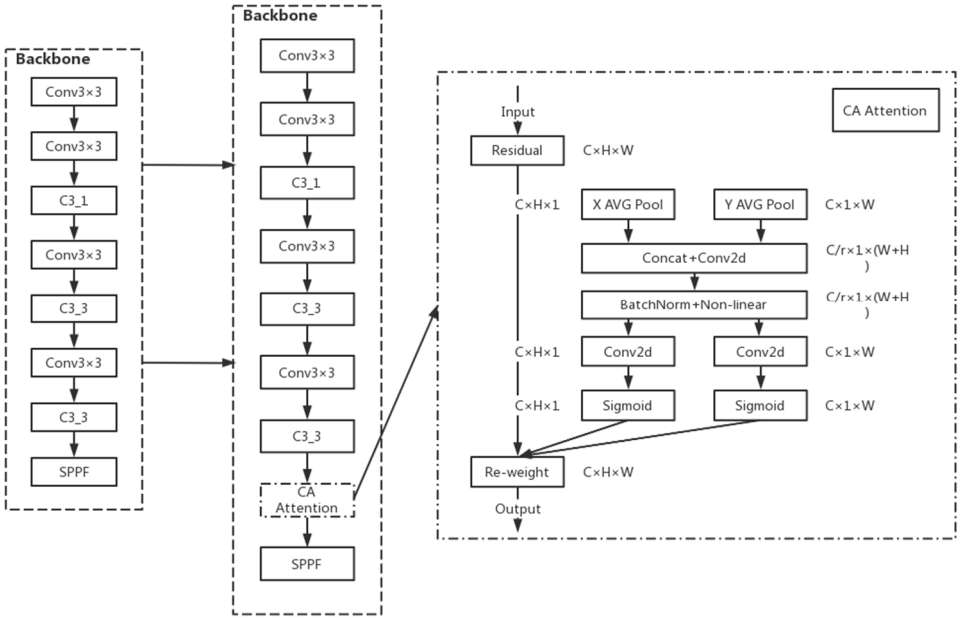


Figure 7. YOLOv5n's integration into CA attention mechanism.

2.4.2. Incorporating Swin Transformer Structure to Improve Model Generalization Performance

In the detection of maize leaf spots, the distribution of different types of spots in the leaf images differs: large spots occupy a small area of the leaf and rely more on local information of high-level features; rusts have a large distribution area and rely on global information more obviously; gray spots are moderate in size and rely on both local and global information. The performance of Convolutional Neural Networks (CNN) is more capable of capturing local information and has a certain disadvantage in global information acquisition. To alleviate the adverse effects of the non-uniform spot size, the model is improved by extracting global information using Swin Transformer [23]. In this paper, a smaller size target detection head was added to the original small, medium and large size detection heads of the YOLOv5n model to enhance its ability to identify small targets of the spots. The x-small size detection head in Figure 8, and the Swin Transformer structure, was incorporated into the large size detection head to replace the original C3 structure with the C3STR structure incorporated into the large size detection head to change the original C3 structure to C3STR structure, thus improving the model's capture of feature information.The improved network structure is shown in Figure 8.

The Swin Transformer model was proposed by Microsoft Research in 2021. Swin Transformer uses hierarchical feature maps similar to those used in convolutional neural networks, such as feature map sizes with 4×, 8×, and 16× down-sampling of images, such that the backbone helps to build on top of this for tasks such as target detection, instance segmentation, etc. The Swin Transformer network is another collision of the Transformer model in the field of vision. The Swin Transformer network is another collision of Transformer model in vision field.

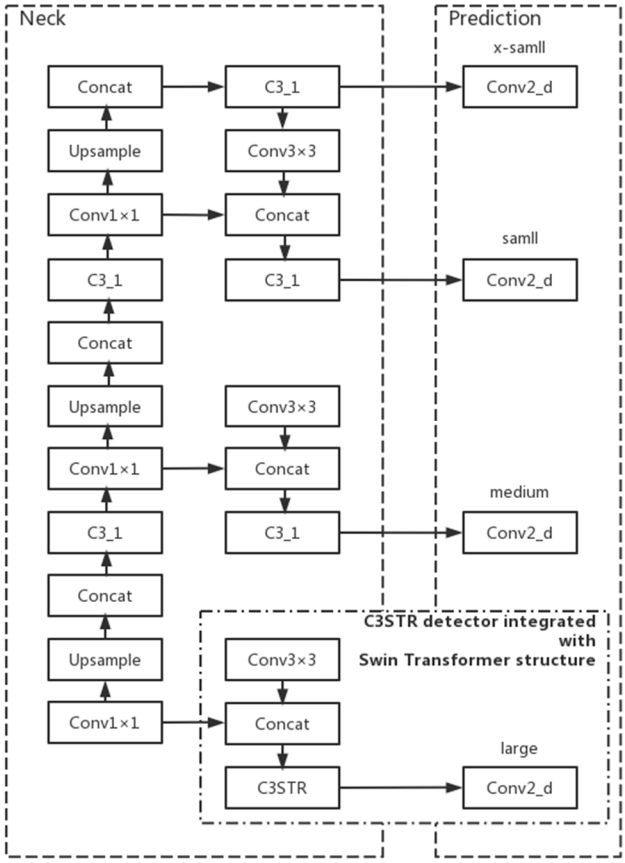


Figure 8. Add x-small detection head and Swin Transformer structure.

The concept of Windows Multi-Head Self-Attention (W-MSA) is used in Swin Transformer, for example, in the 4-fold downsampling and 8-fold downsampling in the figure below. The feature map is divided into multiple disjointed regions (Window), and Multi-Head Self-Attention is performed only within each window (Window).

The basic flow of the whole framework is as follows.

First, the image is input to the patch partition module for chunking, i.e., every 4×4 adjacent pixels is a patch, and then it is flattened in the channel direction. Assuming that the input is a three-channel RGB image, each patch has $4 \times 4 = 16$ pixels, and each pixel has three values, R, G, and B. The flattened image shape is $16 \times 3 = 48$, so the image shape changes from [H, W, 3] to [H/4, W/4, 48] after patch partition. Then the channel data of each pixel is linearly transformed by the linear embedding layer from 48 to C, i.e., the image shape is changed from [H/4, W/4, 48] to [H/4, W/4, C].

The Swin Transformer divides patches by first determining the size of each patch and then calculating the number of patches. The number of Swin Transformers decreases, and the perceptual range of each patch expands as the network depth deepens, which is designed to facilitate the construction of layers of Swin Transformer and to adapt to the multi-scale of visual tasks.The architecture of a Swin transformer (Swin-T) is shown in Figure 9.

Then the feature maps of different sizes are constructed by four stages; except for Stage 1, where a linear embedding layer is first passed, the remaining three stages are first downsampled by a patch merging layer. Note that there are two types of blocks, as shown in Figure 10, which differ only in that one uses the W-MSA structure and the other uses the SW-MSA structure. Moreover, these two structures are used in pairs, with one W-MSA structure used first and then one SW-MSA structure used.

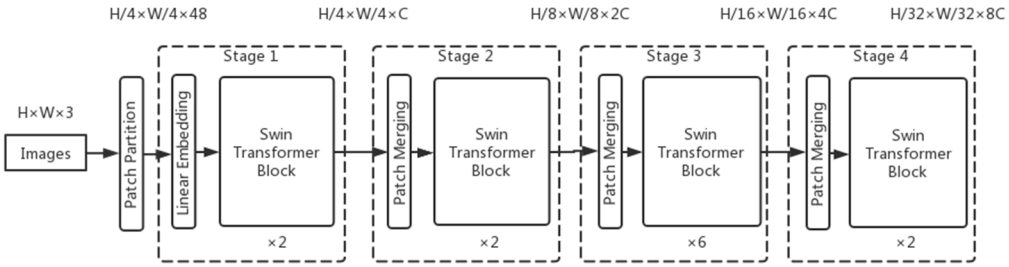


Figure 9. The architecture of a Swin Transformer (Swin-T).

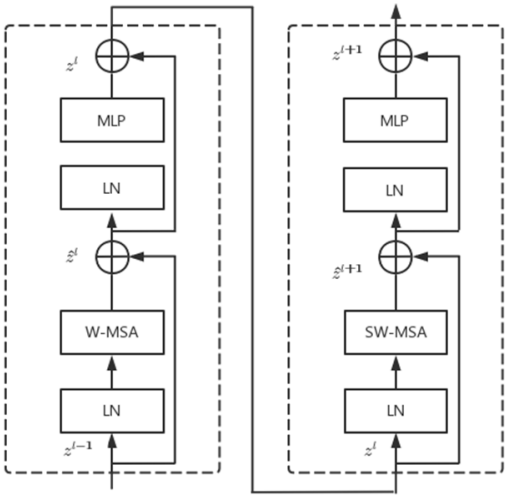


Figure 10. Two successive Swin Transformer blocks.

2.5. Test Environment

In this paper, we use the deep learning framework PyTorch to build and improve the model in the Anaconda3 environment, and train and test the model under the Windows 10 system. The computer CPU is 11th Gen Intel (R) Core (TM) i7–11700K @ 3.60 GHz and the GPU is NVIDIA GeForce GTX 1080 Ti. The GPU is used for acceleration to improve the network training speed, the Cuda version is 11.1.0, and the cudnn version is 8.1.0.

2.6. Evaluation Metrics

In this paper, the performance of the YOLOv5n network model is evaluated using several metrics from the target detection algorithm, specifically Precision (P), Recall (R), and mean Average Precision (mAP) [24]. Average Precision (AP) is the integral of the PR curve formed by taking Precision (P) as the vertical axis and Recall (R) as the horizontal axis. A recall is a metric that reflects the ability of the model to find positive sample targets, precision is used to reflect the ability of the model to classify samples, and average precision is a metric that reflects the overall performance of the model to detect and classify targets. The mean Average Precision (mAP) represents the average of the mean accuracy of all categories. Among all the metrics, mAP is the most important evaluation metric in the target detection algorithm, which can measure the accuracy of the detection algorithm. $mAP_{0.5}$ is the AP of the target detection model evaluated at an IoU threshold of 0.5. $mAP_{0.5}$ is its mean value for all categories; $AP_{0.5–0.95}$ is the mean value of the AP of the model evaluated at different IoU thresholds (0.5–0.95, step size 0.05); and $AP_{0.5–0.95}$ is the average value of AP evaluated under different IoU thresholds (0.5–0.95, step size 0.05), which is a more stringent model accuracy index. $mAP_{0.5–0.95}$ is the average value of all categories. In this paper, we choose $mAP_{0.5–0.95}$ as the evaluation index. The additional evaluation index considered in this paper is the number of parameters, and the number of parameters indicates the size of the storage space occupied by the model file in MB.

The expressions for the calculation of Precision (P), Recall (R), Average Precision (AP), and mean Average Precision (mAP) are shown in Equations (1)–(4).

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$AP = \int_0^1 P \cdot RdR \tag{3}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{4}$$

where:

- The number of true samples.
- The number of false positive samples.
- Number of pseudo-negative samples.
- The number of species in the sample.

The positive and negative samples are judged by setting the average Intersection over Union (IoU) threshold between the predicted target area and the actual target area, and if the IoU of both exceeds the threshold, the sample is positive, and if vice versa, the sample is negative.

3. Results

3.1. Model Trainsing Hyperparameter Setting

In the model training stage, the total number of training rounds is set to 300 and the iteration batch_size is set to 16. Setting too small a value for iteration batch size will lead to too slow training, and setting too large a value for iteration batch size will lead to insufficient video memory to run the experiment. In the experimental environment of this paper, the maximum YOLOv5x model runs at 83%, so setting a larger value will lead to run failure, so it is considered that the size of 16 is a more appropriate level.

In the model training process, if the learning rate is adjusted too large, the network will not converge, while if the learning rate is adjusted too small, the network convergence speed will become slow, so the appropriate learning rate is a key factor in the training process. This paper uses three different sizes of initial learning rate to compare the model experiments; the trained model parameters are shown in Table 2.

Table 2. Parameter comparison under different initial learning rates.

Model	Initial Learning Rate	Precision (P)	Recall (R)	Mean Average Precision ($mAP_{0.5–0.95}$)	Parameter Quantity/MB
YOLOv5n	0.1	0.938	0.924	0.885	3.9
	0.01	0.94	0.949	0.924	3.9
	0.001	0.963	0.922	0.889	3.9

From the table, it can be seen that, for the YOLOv5n model studied in this paper, the average precision mean is highest at an initial learning rate of 0.01. In addition, in order to see the comparison of network convergence more intuitively, this paper shows the graphs of model recall for three different sizes of initial learning rates, and the effect is shown in Figure 11.

From the figure, it can be seen that the model converges fastest when the learning rate is 0.01, the model converges best and performs more stably, and its accuracy stays around the optimal value after 100 epochs, which avoids the interference of the model fluctuation by chance on the final result; when the learning rate is 0.1, the model recall

curve fluctuates more and converges extremely poorly; when the learning rate is 0.001. Although the convergence effect is better when the learning rate is 0.001, the convergence speed is not as fast as when the learning rate is set to 0.01, and the accuracy remains around the optimal value after 150 epochs. Therefore, the initial learning rate of 0.01 is used to carry out the subsequent experiments.

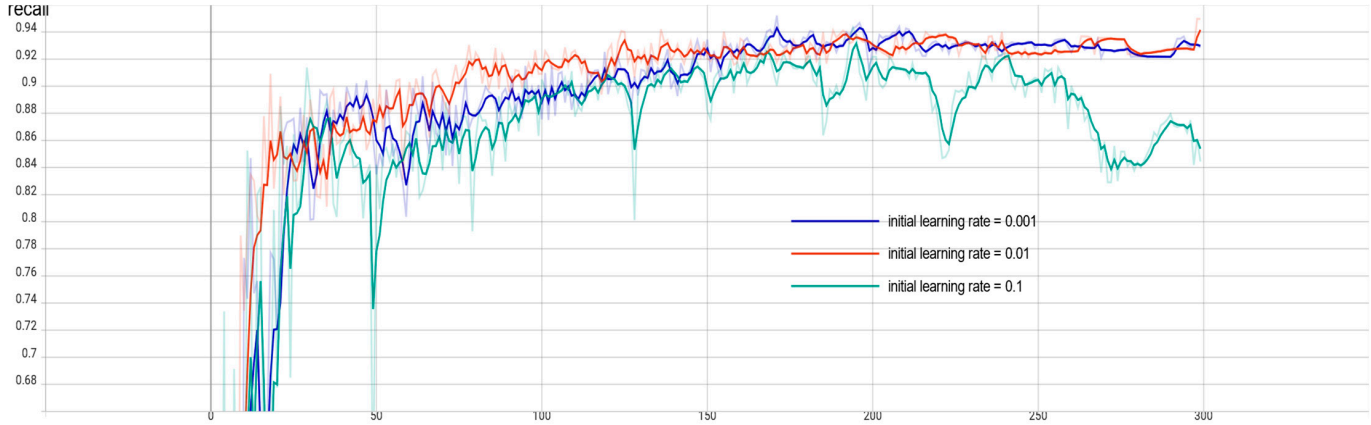


Figure 11. Comparison of recall rates under different initial learning rates.

To prevent the overfitting phenomenon, the weight decay coefficient is set to 0.0005, the confidence level is set to 0.5, and the non-maximum suppression threshold is set to 0.3.

3.2. Comparison of Different Algorithm Models

Since there are five versions of YOLOv5 so far, to ensure the authenticity, accuracy, and rigor of the experimental process, the five versions of YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x are trained on the same platform, the same framework with the same training parameters, and on the same data set in turn. The comparative data are shown in Table 3.

Table 3. Comparison of parameters of different network models.

Model	Precision (P)	Recall (R)	Mean Average Precision ($mAP_{0.5–0.95}$)	Parameter Quantity/MB
YOLOv5n	0.94	0.949	0.924	3.9
YOLOv5s	0.973	0.942	0.932	14.5
YOLOv5m	0954	0.961	0.945	40.3
YOLOv5l	0.96	0.965	0.948	92.9
YOLOv5x	0.965	0.965	0.958	173.2

From the above table, we can see that, due to the different depth and width of the models, among the five models of YOLOv5, the average accuracy of YOLOv5n is 0.924, which is the lowest accuracy among the five models, but its model occupies much less memory than the other models, only 3.9 M, and the training time of YOLOv5n is also much shorter than the other models during training. In contrast, the memory consumption of YOLOv5l and YOLOv5x models is around 100 MB after training with smaller data sets, which can be considered as suitable for deployment on large servers only.

Also, among the five models of YOLOv5, YOLOv5n has the lowest precision, but its model takes up far less memory than the other models, only 3.9 M, and the training time of YOLOv5n is also much shorter than the other models. In contrast, the memory consumption of the YOLOv5l and YOLOv5x models is around 100 MB after training with a small data set, which can be considered suitable for deployment on large servers only.

The ultimate goal of this paper is to help farmers accurately identify the diseases infecting maize leaves in real-time by using portable cell phones in the field so that timely

control can be carried out to reduce the adverse effects of low quality and yield of maize caused by the diseases.

Therefore, YOLOv5n is the most suitable lightweight model for mobile deployment to identify maize leaf spots in real-time. The model is also improved in the hope that the accuracy of the YOLOv5n model can be further enhanced to reach a similar level to other large network models without affecting the model size.

3.3. Model Enhancement by Different Attention Mechanisms

In this paper, the CTR_YOLOv5n model is built by incorporating CA attention into the baseline model and using the Swin Transformer-based C3STR block as the prediction head to further improve the generalization ability of the model and improve the model accuracy.

To verify the effectiveness of the improvements on the model, this paper adds the current mainstream attention mechanisms, such as SE (Squeeze-and-Excitation), CBAM (Convolutional Block Attention Module), and ECA (Efficient Channel Attention), to the model under the same platform and framework, using the same training parameters, and the comparison results are shown in Table 4.

Table 4. Comparison of Yolov5n effect improvement under different attention mechanisms.

Model	Precision (P)	Recall (R)	Mean Average Precision ($mAP_{0.5–0.95}$)	Parameter Quantity/MB
YOLOv5n	0.94	0.949	0.924	3.9
SE_YOLOv5n	0.946	0.959	0.931	3.9
CBAM_YOLOv5n	0.962	0.947	0.927	3.9
ECA_YOLOv5n	0.961	0.944	0.931	3.9
CA_YOLOv5n	0.948	0.961	0.948	3.9

From the above table, we can see that the addition of the attention mechanism to the YOLOv5n backbone network improves the mean accuracy of YOLOv5n, and there is no significant change in the model size, which proves that the attention mechanism has an improvement effect on YOLOv5n. Among the mainstream attention mechanisms, the SE attention mechanism improves the mean accuracy of YOLOv5n by 0.7 percentage points, CBAM by 0.3 percentage points, ECA by 0.7 percentage points, and the improved CA_YOLOv5n by 2.4 percentage points. In this study, the accuracy of the CA_YOLOv5n network model is improved to the same level as YOLOv5l while keeping the model size the same as YOLOv5n, confirming that the improvements in this paper can allow CA_YOLOv5n to be deployed on mobile devices with an accuracy similar to that of the server-deployed model.

In order to see the improvement effect of different attention mechanisms on YOLOv5n more intuitively, this paper shows the accuracy comparison of YOLOv5n with different attention mechanisms added, and the effect is shown in Figure 12.

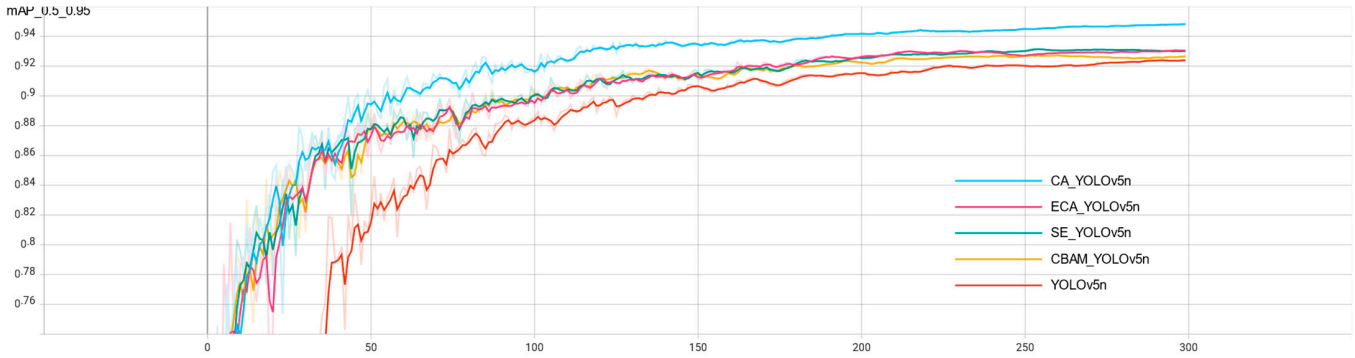


Figure 12. Precision comparison of different attention mechanisms.

From the figure, we can see that the accuracy of the YOLOv5n network without the attention mechanism starts to increase rapidly around 40 training sessions, and the increase in accuracy starts to slow down around 100 training sessions, while the accuracy of the model with the attention mechanism starts to increase rapidly around 15 training sessions, in which CA_YOLOv5n has the fastest growth rate and the largest increase in the first 50 training sessions, and the accuracy is always higher than the other models in the subsequent training sessions. The accuracy of CA_YOLOv5n is consistently higher than the other models in the subsequent training.

Comparing CA with other mainstream attention mechanisms, CA attention takes into account not only the relationship between channels but also the location information in the feature space. CA uses a more efficient method to capture location information and channel relationships. It does this by decomposing the two-dimensional global pooling operation into two directions, width and height, and then averaging the pools globally to obtain feature maps in the two different directions, width and height. The feature maps in both directions are then processed to obtain the final feature maps, which is the main reason why the CA attention is more effective than other attentions.

The experiments in this paper can prove that the model improvement approach studied in this paper is better than adding other mainstream attention mechanisms. In addition, from Figure 6, we can see that all the models reach a relatively flat accuracy after 200 training sessions, which indicates that the proposed model achieves the ideal state.

3.4. Model Improvement to Improve the Effect

Based on the CA_YOLOv5n model, the experiments in this paper make further improvements by adding a small target detection layer to the neck network neck of the CA_YOLOv5n model to enhance its ability to recognize small targets of disease spots, and by replacing the original C3 detection head into the Swin Transformer structure with the C3STR detection head to improve the model's capture of feature information.

The following Table 5 shows the integrated inference results of each model for different types of maize leaf diseases at 256 × 256 pixels. The YOLOv5n model showed high recognition ability for maize rust and healthy maize leaves, reaching 0.953 and 0.99, respectively, but low recognition rate for maize gray spot disease and maize blotch disease, reaching only 0.905 and 0.849. The overall accuracy of the model was further improved by adding the CA attention mechanism. Compared with YOLOv5n, the model CA_YOLOv5n showed a significant improvement in the recognition performance of maize gray spot disease and maize blotch disease, which was comparable to the performance of YOLOv5l model. The final proposed model, CTR_YOLOv5n with $mAP_{0.5–0.95}$, improved by 2.8% compared to the initial model YOLOv5n and by 0.4% compared to the YOLOv5l model. It is noteworthy that the $mAP_{0.5–0.95}$ of model CTR_YOLOv5n reached 0.928 and 0.925 for maize gray spot disease and maize blotch disease, which were significantly improved compared with each model, proving that the accuracy of the final model proposed in this paper is effectively improved for small target detection of disease spots. The accuracy distribution of the CTR_YOLOv5n model is more balanced, which shows the effectiveness of the improvement strategy of this paper. Although the number of parameters in the final model is slightly increased compared with YOLOv5n, it still meets the requirements of being lightweight.

Table 5. Performance of each model under different disease categories.

Model	Maize Gray Spot Disease	Maize Rust	Healthy Maize Leaves	Maize Blotch Disease	Mean Average Precision ($mAP_{0.5–0.95}$)	Parameter Quantity/MB
YOLOv5n	0.905	0.953	0.99	0.849	0.924	3.9
YOLOv5l	0.916	0.986	0.992	0.898	0.948	92.9
CA_YOLOv5n	0.924	0.985	0.992	0.891	0.948	3.9
CTR_YOLOv5n	0.928	0.964	0.99	0.925	0.952	5.1

It is confirmed that the improvements in this paper allow CTR_YOLOv5n to be deployed on mobile with an accuracy comparable to that of the server-deployed model.

4. Discussion

Crop disease identification has an important position in the field of digital agricultural information technology as well as in the construction of smart agriculture. In addition, the existing crop disease recognition has many kinds, low accuracy, and poor targeting, which cannot meet the actual needs of maize growers. Deep learning, as an emerging technology in the field of machine learning, has a wide range of applications in the field of image recognition.

In this paper, we study the improved CTR_YOLOv5n model to identify and detect maize leaf diseases, identify maize blotch disease, gray spot, and rust, provide technical reference for maize disease control, and help farmers to solve the problem of difficult maize disease identification, which has high application value. The main findings are as follows.

1. Since disease spots generally occupy a small leaf area, as a small-scale target, the number of data sets may not be sufficient. Using the mosaic data enhancement in the YOLOv5n model, the large samples were randomly reduced to small samples, increasing the number of small-scale targets. The number of datasets was effectively increased to speed up the network convergence and improve the model accuracy.

2. To address the characteristics that it is not easy to distinguish between different diseases of maize leaves, especially those with similar color and shape between maize blotch disease and maize gray spot disease, the backbone network of the YOLOv5n model is improved to incorporate the CA attention mechanism to enhance the extraction of spot features, add a minimum size detection head of YOLOv5n model, and introduce a Swin Transformer structure in the large size detection head to fuse global and local information to improve the generalization performance of the model. This improves the generalization performance of the model and, finally, improves the recognition accuracy by 2.8%.

3. Using the CTR_YOLOv5n model to identify maize leaf spots, the average recognition accuracy is 95.2%, which is 2.8% higher than the original model; compared with the YOLOv5l model, the trained model size is maintained at 5.1MB with the same accuracy, which is 94.5% smaller. The model accuracy is effectively improved with a small model footprint. It can be concluded that the improved CTR_YOLOv5n model is more suitable for real-time detection of maize leaf diseases on mobile applications and improves detection efficiency.

The results show that the improved CTR_YOLOv5n model helps to improve the overall effect of maize leaf disease recognition. Considering that the image background is often complex in actual disease recognition, whether the model can still guarantee such a high accuracy rate for complex backgrounds and occluded cases is subject to further research. In addition, the scope of this paper is limited to the identification of diseases on the leaves of maize. In addition, the scope of identification can be extended to the identification of diseases on the roots and ears of maize, such as maize stalk rot, maize silky black ears, and maize rot, which have a high incidence rate and can cause huge losses, and can threaten the life and health of humans and animals in serious cases.

5. Conclusions

In this paper, we propose a method to improve the YOLOv5n model for the recognition and detection of maize leaf spots. CA attention is incorporated into the backbone network of the YOLOv5n model to increase the weight of feature information of maize leaf spots, so that the model pays more attention to the feature information of spots during training; a smaller size target detection head is added to enhance its ability to recognize small targets of spots; and the Swin Transformer structure is incorporated into the large size detection head, so as to improve the model's ability to capture feature information. The accuracy of the model for maize leaf spot recognition is effectively improved. The experimental results show that the average recognition accuracy of the proposed method can reach 95.2%

for maize leaf spots, and the recognition accuracy of maize gray spot and maize large spot disease, which are not easy to recognize, has been effectively improved, reaching 92.8% and 92.5%, respectively. Compared with other deep learning network models, this method has the advantages of higher accuracy and smaller size. In the future, we will collect some images of maize leaf diseases in natural environments and improve the model in a more lightweight way to develop a crop disease recognition system that can be applied to mobile devices.

Author Contributions: The contributors are J.Z. and Q.Y. for conceptualization; H.Y. for methodology; J.Z. for formal analysis; L.M. and Q.Y. for investigation/writing—original draft/supervision; L.M. and J.Z. for visualization; J.Z. for writing—review/editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China–Joint Fund (u19a2061), Jilin Provincial Department of Education Project (No. JJKH20210336KJ), Jilin Province Ecological Environment Department Project (2021–07), Jilin Province Science and Technology Development Plan Project (No. 20200301047RQ), and Jilin Provincial Natural Science Foundation (No. 20200201288JC).

Data Availability Statement: Data supporting the findings of this study are available from the corresponding author.

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

References

1. Cao, T.; Zhang, X.; Chen, X.; Peng, X.; Lin, J. Maize Disease Classification Method Based on Spatial Attention Mechanism and DenseNet. *Radio Eng.* **2022**, *52*, 1710–1717.
2. Fan, X.; Zhou, J.; Xu, Y.; Peng, X. Maize Disease Recognition under Complicated Background Based on Improved Convolutional Neural Network. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 210–217.
3. Bao, W.; Huang, X.; Hu, G.; Liang, D. Identification of maize leaf diseases using improved convolutional neural network. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 160–167.
4. Wang, Y.; Wu, J.; Lan, P.; Li, F.; Ge, C.; Sun, F. Apple disease identification using improved Faster R-CNN. *J. For. Eng.* **2022**, *7*, 153–159. [CrossRef]
5. Huang, Y.; Ai, X. Research on Classification of Corn Leaf Disease Image by Improved Residual Network. *Comput. Eng. Appl.* **2021**, *57*, 178–184. [CrossRef]
6. Wu, Y. Identification of Maize Leaf Diseases Based on Convolutional Neural Network. *J. Phys. Conf. Ser.* **2021**, *1748*, 032004. [CrossRef]
7. Wang, C.; Wang, C.; Liu, J. Identification of Maize Leaf Diseases based on Deep Learning. *Mod. Agric. Res.* **2022**, *28*, 102–106. [CrossRef]
8. Azlah, M.A.F.; Chua, L.S.; Rahmad, F.R.; Abdullah, F.I.; Wan Alwi, S.R. Review on Techniques for Plant Leaf Classification and Recognition. *Computers* **2019**, *8*, 77. [CrossRef]
9. Koklu, M.; Unlersen, M.F.; Ozkan, I.A.; Aslan, M.F.; Sabanci, K. A CNN-SVM Study Based on Selected Deep Features for Grapevine Leaves Classification. *Measurement* **2022**, *188*, 110425. [CrossRef]
10. Argüeso, D.; Picon, A.; Irusta, U.; Medela, A.; San-Emeterio, M.G.; Bereciartua, A.; Alvarez-Gila, A. Few-Shot Learning Approach for Plant Disease Classification Using Images Taken in the Field. *Comput. Electron. Agric.* **2020**, *175*, 105542. [CrossRef]
11. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–15 December 2015; Curran Associates, Inc.: Red Hook, NY, USA, 2015; Volume 28.
12. 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 Computer Vision, ECCV, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Cham, Switzerland, 2016; pp. 21–37.
13. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
14. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
15. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
16. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
17. Chen, L.; Fu, D. Survey on Machine Learning Methods for Small Sample Data. *Comput. Eng.* **2022**, *48*, 1–13. [CrossRef]
18. Chen, C.; Fan, Y.; Wang, L. Logo Detection Based on Improved Mosaic Data Enhancement and Feature Fusion. *Comput. Meas. Control.* **2022**, *30*, 188–201. [CrossRef]
19. Tian, M.; Liao, Z. Research on Flower Image Classification Method Based on YOLOv5. *J. Phys. Conf. Ser.* **2021**, *2024*, 012022. [CrossRef]
20. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. *arXiv* **2018**, 8759–8768.

21. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. *arXiv* **2021**, 13713–13722.
22. Li, S.; Li, K.; Qiao, Y.; Zhang, L. A Multi-Scale Cucumber Disease Detection Method in Natural Scenes Based on YOLOv5. *Comput. Electron. Agric.* **2022**, *202*, 107363. [CrossRef]
23. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *arXiv* **2021**, arXiv:2103.14030.
24. Shang, Y.; Zhang, Q.; Song, H. Application of deep learning based on YOLOv5s to apple flower detection in natural scenes. *Trans. Chin. Soc. Agric. Eng.* **2022**, *38*, 222–229. [CrossRef]

agronomy

MDPI

Article

Based on FCN and DenseNet Framework for the Research of Rice Pest Identification Methods

He Gong [1,2,3,4], Tonghe Liu [1], Tianye Luo [1], Jie Guo [1], Ruilong Feng [1], Ji Li [1], Xiaodan Ma [1], Ye Mu [1,2,3,4], Tianli Hu [1,2,3,4], Yu Sun [1,2,3,4], Shijun Li [5,6], Qinglan Wang [7] and Ying Guo [1,2,3,4,*]

1 College of Information Technology, Jilin Agricultural University, Changchun 130118, China
2 Jilin Province Agricultural Internet of Things Technology Collaborative Innovation Center, Changchun 130118, China
3 Jilin Province Intelligent Environmental Engineering Research Center, Changchun 130118, China
4 Jilin Province Colleges and Universities and the 13th Five-Year Engineering Research Center, Changchun 130118, China
5 College of Information Technology, Wuzhou University, Wuzhou 543003, China
6 Guangxi Key Laboratory of Machine Vision and Inteligent Control, Wuzhou 543003, China
7 Jilin Academy of Agricultural Sciences, Changchun 130033, China
* Correspondence: guoying@jlau.edu.cn

Abstract: One of the most important food crops is rice. For this reason, the accurate identification of rice pests is a critical foundation for rice pest control. In this study, we propose an algorithm for automatic rice pest identification and classification based on fully convolutional networks (FCNs) and select 10 rice pests for experiments. First, we introduce a new encoder–decoder in the FCN and a series of sub-networks connected by jump paths that combine long jumps and shortcut connections for accurate and fine-grained insect boundary detection. Secondly, the network also integrates a conditional random field (CRF) module for insect contour refinement and boundary localization, and finally, a novel DenseNet framework that introduces an attention mechanism (ECA) is proposed to focus on extracting insect edge features for effective rice pest classification. The proposed model was tested on the data set collected in this paper, and the final recognition accuracy was 98.28%. Compared with the other four models in the paper, the proposed model in this paper is more accurate, faster, and has good robustness; meanwhile, it can be demonstrated from our results that effective segmentation of insect images before classification can improve the detection performance of deep-learning-based classification systems.

Keywords: pest identification; FCN; DenseNet; attention mechanism

Citation: Gong, H.; Liu, T.; Luo, T.; Guo, J.; Feng, R.; Li, J.; Ma, X.; Mu, Y.; Hu, T.; Sun, Y.; et al. Based on FCN and DenseNet Framework for the Research of Rice Pest Identification Methods. *Agronomy* **2023**, *13*, 410. https://doi.org/10.3390/agronomy13020410

Academic Editors: Jian Zhang, Randy G. Goebel and Zhihai Wu

Received: 14 December 2022
Revised: 18 January 2023
Accepted: 19 January 2023
Published: 30 January 2023

1. Introduction

Climate, ecology, natural catastrophes, and many other factors have profound impacts on the production of grains, with insect damage being one of the major factors affecting crop productivity. Numerous crops, including wheat, maize, and rice, have lower yields as a result of agricultural pests. Hence, to effectively control pests, it is necessary to predict the occurrence trend, quantity, population dynamics, and potential damage of pests, while real-time and accurate forecasting is very important. The correct identification and classification of rice pests is a prerequisite for the prevention and management of rice pests. Insect experts or insect taxonomists typically carry out traditional insect classification and identification work based on specialized expertise, research experience, or reference maps. However, a lot of time and effort are required by this method, which has generally low work efficiency and extremely unstable accuracy [1]. The development of an automatic identification and classification system of pests will remarkably reduce the labor intensity of plant protection personnel and improve the accuracy of forecasting, thereby reducing the loss of rice yield.

With the continuous development of Internet technology, the use of computer vision technology for pest identification has gradually been widely studied. Many works in the

literature have applied the traditional machine vision technique to pest identification [2,3]. However, this method not only has difficulty in meeting actual needs, but also the generalization ability and robustness of the model are relatively poor. In the past few years, with the continuous development and updating of artificial intelligence technologies, such as deep learning technology and big data technology, more possibilities have emerged for examining the pest identification problem [4]. Rice is considered one of the most important food crops, and its yield directly affects many issues, such as food security. Rice will inevitably be affected by different pests during the growth process. Thus, controlling the scale of pests and diseases at an appropriate time can reduce the amount of pesticide spraying and avoid a large reduction in rice production. Therefore, the accurate identification of pests has become an important basis for pest control, which renders the investigation of the rice pest identification technology particularly important.

Currently, the scientific community has more options to create increasingly automated systems that can accurately recognize objects of any kind thanks to the astonishing progress of picture classification [5]. Yang et al. [6] used the shape of the insect image, extracted the features with color, and then utilized the radial basis neural network to classify it with an accuracy rate of 96%. Zhang et al. [7] proposed a faster R-CNN framework that was composed of automatic recognition algorithms for two types of insects. The recognition accuracy rate of this model reached 90.7%. Chen et al. [8] reported the combination of machine learning and convolutional neural networks to identify five kinds of corn pests in the northeast cold region. Cheng et al. [9] introduced a deep convolutional neural network. After the feature extraction was performed with color, the accuracy rate for classification using a radial basis neural network was 97.6%. Based on the standard convolutional neural network, Sun et al. [10] introduced the attention mechanism and created a convolutional neural network model based on the attention mechanism to recognize soybean aphids. However, the above methods still have defects, such as insufficient sample size of data, complex data preprocessing, insufficient feature extraction, large fluctuation of model fitting degree, similar target features, etc. Moreover, the features vary greatly among different insects and there are many types of rice pests, the above methods do not have relevant parameter adjustment and optimization for rice pests, and the models do not have good generalization, so they are not applicable to the classification of rice pests.

Segmentation is the basic stage of recognition and classification. The main purpose is to filter the image noise that is generated when the image is collected, remove the redundant background information of the target image, and extract the target object in a targeted manner to extract accurate, concise, and expressive feature information in the follow-up work. Before the advent of fully convolutional networks (FCN) [11], there were also some traditional methods for semantic segmentation, such as normalized cut [12], structured random forests [13], and SVM [14]. FCN is the first network that delves semantic segmentation to the pixel level, because FCN avoids the problem of repeated storage and computational convolution due to the use of pixel blocks; therefore, FCN is more efficient than traditional networks based on conventional convolutional neural networks (CNNs) [15] for segmentation, and there is no limit on the input image size. DenseNet proposes a more radical dense connection mechanism: interconnecting all layers. Specifically, each layer accepts all the previous layers as its additional input. Due to the dense connection approach, DenseNet improves the back propagation of gradients, making the network easier to train and with smaller and more computationally efficient parameters. To better solve the pest classification problem, we proposed a simple and efficient full convolutional network based on FCN and introduced an encoder–decoder CRF network, long-hop and short-hop connection mechanisms to solve the problems of the poorly segmented and detail insensitive FCN, and an efficient channel attention (ECA) mechanism [16] of the DenseNet network to further enhance the performance of the network. While strengthening the extraction of insect features, the extraction of invalid background features is inhibited, so as to improve the identification accuracy and generalization ability of the network and ensure the efficient and accurate identification of insect pests.

2. Materials and Methods

2.1. Experimental Data Set

The source of the data stems from the intelligent insect forecasting lamp, and the location is multiple rice fields in Shuangyang District and Jiutai District, Changchun City, Jilin Province, from July to September 2021. The intelligent insect forecasting lamp was automatically turned on after sunset every day, and the trapped insects were heated and inactivated by the electric heating plate to make it easier to shoot. Finally, the white conveyor belt was used to tile the insect corpses. The rolling time can be set according to the period of a high incidence of the insect pests and the conveyor belt can be manually and remotely controlled to ensure that the data are not repeated and are not unreliable. The industrial camera was used for vertical shooting, and the shooting was synchronized with the rolling of the transmission belt. The model of the industrial camera was MV-CE120-10UM/UC 1/1.7′CMOS MV-CE120-10UC with 12 million pixels, while the manufacturer was HIKROBOT. It can also be shot manually and remotely at any time. The resolution was 4024 × 3036, the unit was px, and 4236 original pictures were obtained. The captured insect images were automatically uploaded to the cloud database for subsequent processing. Part of the original images are depicted in Figure 1.

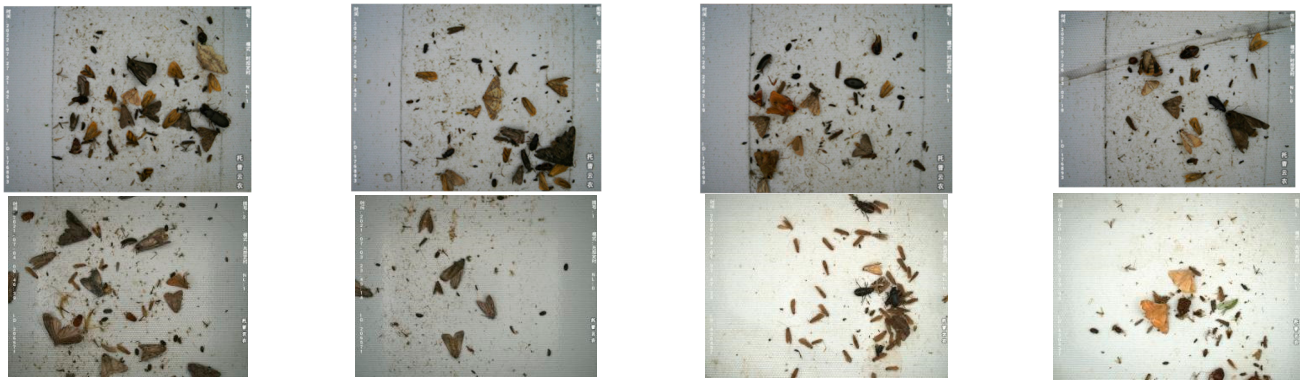


Figure 1. Part of the original images.

Due to the impact of the wild biodiversity, collection locations, and other factors, most of the insect populations are relatively small. In order to ensure the reliability of the experiment and the accuracy of the classification, according to the data collection of the actual local insect species, the classification was carried out with reference to the 2022 version of the "Catalogue of Biological Species in China". The relevant experts from the Animal Science and Technology College of Jilin Agricultural University also provided valuable help. The most harmful insect pests are *Chilo suppressalis* (*Lepidoptera*:Pyralidae), *Naranga aenescens* (*Lepidoptera*:Noctuidae), *Cnaphalocrocis medinalis* (*Lepidoptera*:Pyralidae), *Nilaparvata lugens* (*Homoptera*:Delphacidae), *Agrotis ypsilon* (*Lepidoptera*:Noctuidae), *Gryllotalpa* sp. (*Orthoptera*:Grylloidea), *Mythimna separata* (*Lepidoptera*:Noctuidae), *Helicoverpa armigera* (*Lepidoptera*:Noctuidae), *Gryllidae* (*Orthoptera*:Gryllidae), and *Holotrichia diomphalia* (*Coleoptera*:Melolonthidae) as the target insects. Since the traps are based on the phototaxis of insects, the identification studies in this work are for adults. The insects in the original image have great uncertainty, while the repeated and wrong images were removed through manual screening. In order to facilitate the display of the pest images, the target insects in the acquired data were marked and classified. The tagging work was completed by using the software Visual Object TaggingTool(VoTT)v2.2.0 developed by Microsoft Corporation, and the storage format was PascalVOC. The experimental data set had a total of 2225 pictures. Among them, there are 422 pictures of *C. suppressalis*, 356 pictures of *N. aenescens*, 335 pictures of *C. medinalis*, 256 pictures of *N. lugens*, 239 pictures of *A. ypsilon*, 123 pictures of *G.* sp, 366 pictures of *M. separata*, 189 pictures of *H. armigera*, 125 pictures of

Gryllidae, and 128 pictures of *H. diomphalia*. Partial images of some rice pests selected from the test set are shown in Figure 2.

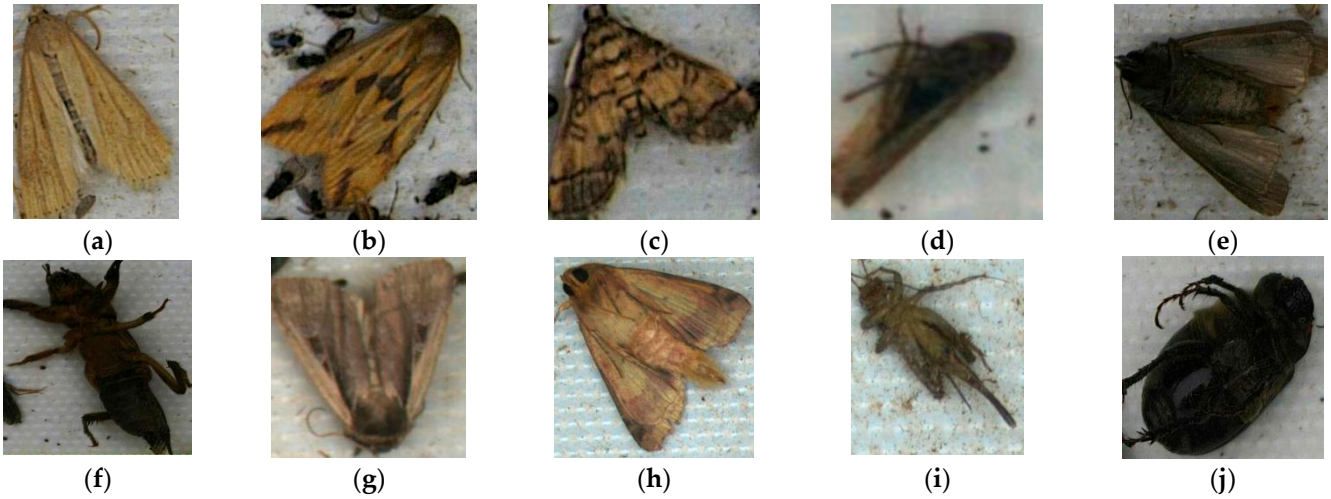


Figure 2. Examples of pest images: (**a**) *C. suppressalis*, (**b**) *N. aenescens*, (**c**) *C. medinalis*, (d) *N. lugens*, (**e**) *A. ypsilon*, (**f**) *G.* sp, (**g**) *M. separata*, (**h**) *H. armigera*, (**i**) Gryllidae, (**j**) *H. diomphalia*.

2.2. Data Augmentation

In deep learning, for the training of convolutional neural networks, a large amount of data sets is often required. Otherwise, various phenomena such as overfitting and low recognition accuracy will take place. However, under the existing conditions, due to the difficulty in collecting rice pest data sets, and the lack of existing rice pest data, some data expansion methods were used here to achieve the purpose of increasing the data set. Particularly, shift, scale, rotation, flip, noise, brightness, and other data expansion methods were utilized to expand the original data set by 10 times, which corresponds to 4220 pictures of *C. suppressalis*, 3560 pictures of *N. aenescens*, 3350 pictures of *C. medinalis*, 2560 pictures of *N. lugens*, 2390 pictures of *A. ypsilon*, 1230 pictures of *G.* sp, 3660 pictures of *M. separata*, 1890 pictures of *H. armigera*, 1250 pictures of Gryllidae, and 1280 pictures of *H. diomphalia*. According to the ratio of 6:2:2, these were divided into training data, verification data, and test data.

3. Model Refinement

By using the FCN algorithm as the basic framework, the DenseNet as the feature extraction network of the FCN algorithm, and introducing the channel attention mechanism, a rice pest recognition algorithm was proposed. We named the proposed model in this paper FCN-ECAD for ease of presentation and use. In Figure 3, the methodological framework of this study is described and illustrated and in the follow-up is thoroughly discussed.

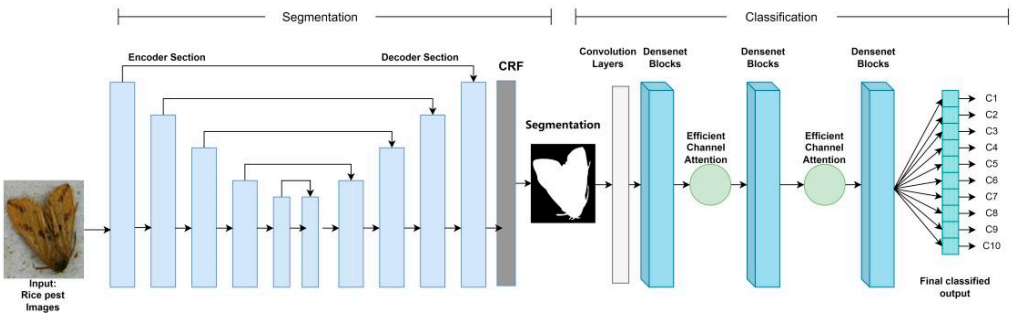


Figure 3. Schematic diagram of the proposed deep learning framework for image segmentation and classification of rice pests.

3.1. Feature Extraction Based on Encoder–Decoder Network

Figure 4 depicts the network's encoder and decoder parts, each of which has five different successive phases. Each stage consists of a cascaded layer, a ReLU activation layer, a convolutional layer with a 3 × 3 kernel size, and a series of skip routes. In the final step, there are 1024 convolutional filters instead of the initial 64. Moreover, instead of using the typical short jump connections, a sequence of jump pathways with both long and short jump connections was created. Nonlinearities were also added by using the ReLU activation module, which might hasten the network's training.

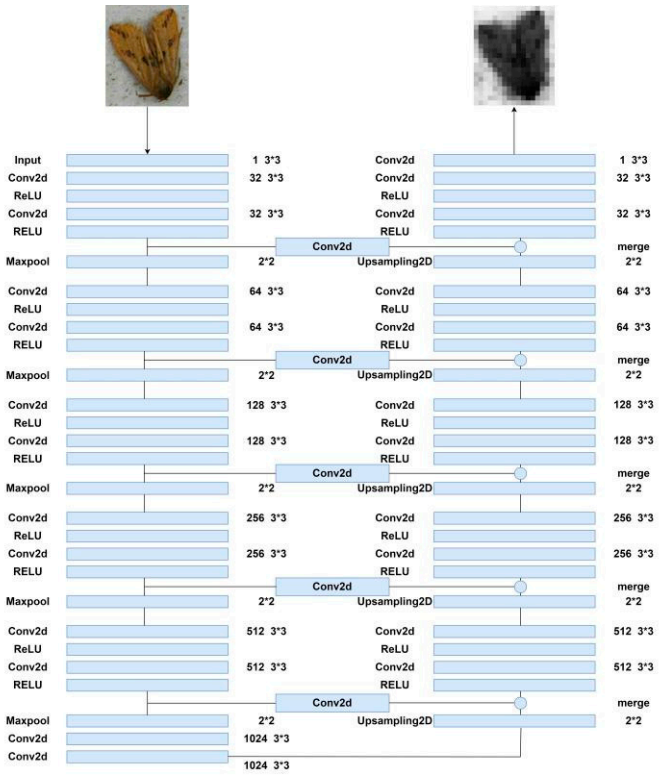


Figure 4. Deep convolutional encoder–decoder network structure diagram.

Additionally, the downsampling function was carried out by the encoder component using a max pooling module. As indicated in the equation, the decoder portion passes the pooled index to the appropriate upsampling layer after a feature vector has been retrieved from the input picture by employing a convolutional layer and downsampling by half using a max pooling module (1).

$$Y_i = U(F(I : r) : d) \tag{1}$$

where the final output, F denotes the downsampling feature map, r represents the RELU activation function, d stands for the downsampling module, and U is the upsampling module.

The decoder then applied an upsampling layer and multiplied the sample size by a factor of 2 to the feature vector from the preceding layer. In order to offer rich information, prevent gradients from dissipating, and restore the lost feature information, the matching output feature maps of the matched encoder component were next concatenated with these feature vectors. With a convolutional layer that has a 1 × 1 kernel and a softmax module, the decoder component was finished. The projected split was discovered to correspond to the class with the highest probability for each pixel by using a softmax classifier and the probability output from the N-channel image, as demonstrated by Equation (2).

$$P(\mathrm{y} = \mathrm{i}|\, x) = \frac{e^{x^T w_i}}{\sum_{n=1} e^{x^T w_n}} \tag{2}$$

where x is the feature map, w refers to the kernel operator, and n represents the number of classes.

3.2. Long-Hop and Short-Hop Connections

Figure 4 depicts the usage of the skip route of both long-hop and short-hop connections. As a reinforcement learning strategy for effective feature extraction, the system used short-hop connections to create very deep FCNs. Shortcut connections were also used to accelerate feature extraction and learning by using 2×2 convolutional layers. The method utilized a variety of skip pathways to hierarchically integrate downsampled and upsampled features and bring the encoder feature map's semantic level closer to that of the decoder. The spatial information that was lost during downsampling was replaced in the long-hop connections utilized for the upsampling step by the extracted features.

3.3. CRF

The CRF algorithm [17] is a standard algorithm that is widely used in edge detection. The CRF algorithm was introduced to ensure contour refinement and insect boundary localization to improve classification performance.

3.4. Feature Extraction Network DenseNet

In this work, DenseNet was adopted for insect feature extraction because the best performance based on the ImageNet classification task is provided. The TOP-1 of those popular pretrained models is summarized in Ref. [18] and the comparison results are displayed in Figure 5. As can be observed, DenseNet outperformed other pretrained models. Therefore, DenseNet was chosen as the feature extraction model in this work.

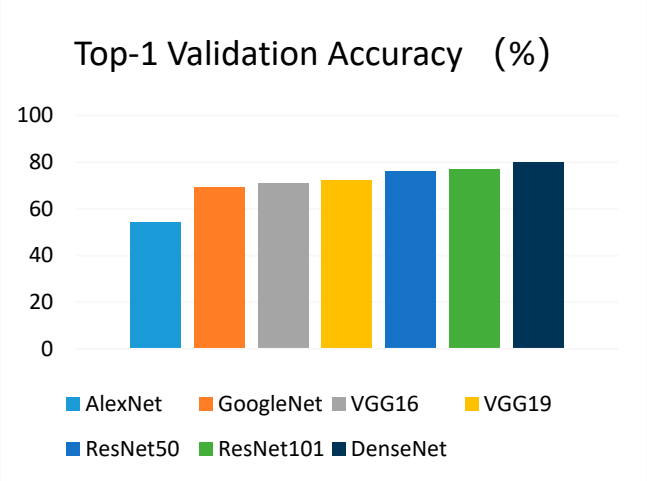


Figure 5. Comparison of the pretrained models.

A deeper network with dense connections was created via DenseNet. The dense block layer, which aimed to maximize information flow between network layers, is considered the most crucial component of the architecture [19]. Each layer in this architecture receives input from all layers before it and passes the feature mappings to all layers after it. The efficient transmission of the earlier features to the later ones for automatic feature reuse is made possible by these brief connections between the layers near the input and output. As a result, this network topology may be trained more precisely and effectively, and it can be utilized to extract more general and significant properties. Some works in the

literature have merged the features that were previously derived from each layer and then reprocessed the features [20–22]. This technique, which is just a basic concatenation of various feature maps, was not meant to promote feature reuse between layers. As a result, as illustrated in Figure 6, all prior levels were taken into consideration as input layers rather than integrating all feature maps.

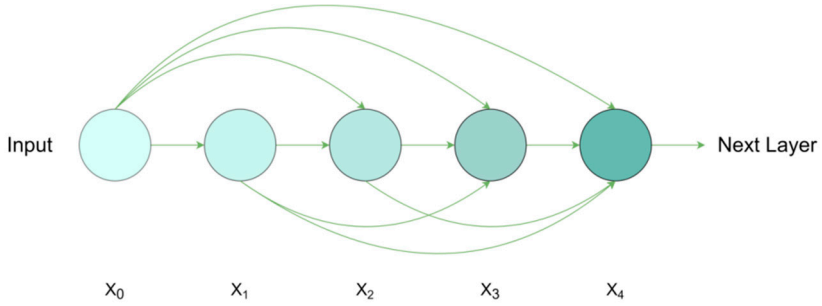


Figure 6. Operation of densely connected convolution, the dense block layer.

Therefore, unlike some traditional network structures, there are $l(l+1)/2$ connections instead of only l connections in the l layer. In this way, the input feature map of layer l can be calculated based on the previous layers.

$$x_l = T_l([x_0, x_1, ..., x_{l-1}]) \tag{3}$$

where $[x_0, x_1, \ldots x_{l-1}]$ represents the concatenation of layer 0, layer 1, and layer $l-1$ feature maps, respectively. In addition, the feature maps were concatenated dimensionally rather than using the pointwise sum as a reference. In the nonlinear transformation, (1) convolution, pooling [23], and rectified linear units [24], among other processes, can be found in $T_l(\cdot)$. For concatenated operations, each dense block consists of several sets of 1×1 and 3×3 convolutional layers with the same padding. Despite using tightly connected patterns, this structure employed fewer parameters than the conventional convolutional networks. In fact, this network architecture decreased the number of feature maps needed for network layers and did away with the need to learn redundant information. The parameter efficiency was thereby greatly increased. On the other hand, consecutive cascades of various layers necessitate access by each layer to the gradients from the initial input data, as well as the loss function. This quick access enhanced the communication across layers and lessened the disappearing gradient issue. This strategy of feature reuse aided in the development of a deeper network architecture and the extraction of deep semantic relations.

3.5. Channel Attention Mechanism

The attention mechanism was added to concentrate on extracting the target insect features in the image in the complex environment of the rice field, which was impacted by the wild biodiversity and diverse backgrounds. In order to boost the efficiency of the deep convolutional neural networks, the ECA-Net attention module was used here, which was a lightweight module. More specifically, it concentrated on the extraction of significant features and suppressed the unimportant features by combining the depth and spatial information of feature maps with an effective attention module. In challenging environments, the extraction can effectively increase the field insects' recognition accuracy. The ECA-Net diagram is shown in Figure 7. First, the input feature map underwent global average pooling, whereas each channel's feature layer was represented by a single value. To obtain the interdependence between each channel and their relationship, a one-dimensional convolution with a size of k was used in the second step. A Sigmoid activation function was then added for normalization, and finally, the weights of each channel were multiplied with the input feature map to strengthen the extraction of significant features.

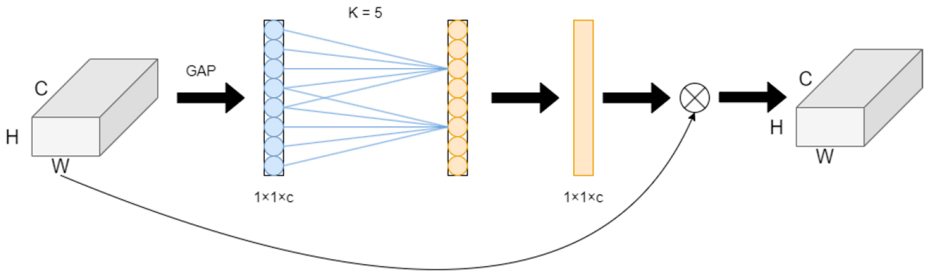


Figure 7. ECA-Net network architecture: the global pooling layer (GAP) obtains aggregated features and ECA generates channel weights by performing one-dimensional convolution of size *k*, where the value of *k* was determined by the channel dimension *C*.

ECA used a one-dimensional convolution of size *k* to interact across channels instead of a fully connected layer, which can effectively reduce the amount of computation and complexity of the fully connected layer, and then generate weights for each channel, namely

$$\omega = \delta\left(CID_{\mathrm{k}}(y)\right) \tag{4}$$

In the formula, ω denotes the channel weight, δ is the Sigmoid activation function, and *CID* represents a one-dimensional convolution. The existence of more channels of the input feature map leads to a greater value of local interaction. As a result, the value of *k* is proportional to the number of channels, *C*. The dimension-related function adaptively determines the value of *k*, namely

$$C = 2^{(\gamma \cdot \mathrm{k} - \mathrm{b})} \tag{5}$$

$$\mathrm{k} = \left| \frac{\log_2(C)}{\gamma} + \frac{\mathrm{b}}{\gamma} \right|_{\mathrm{odd}} \tag{6}$$

As shown in Figure 8, the new DenseNet network consists of dense block layers, efficient channel attention layers, averagepool, and a fully connected layer.

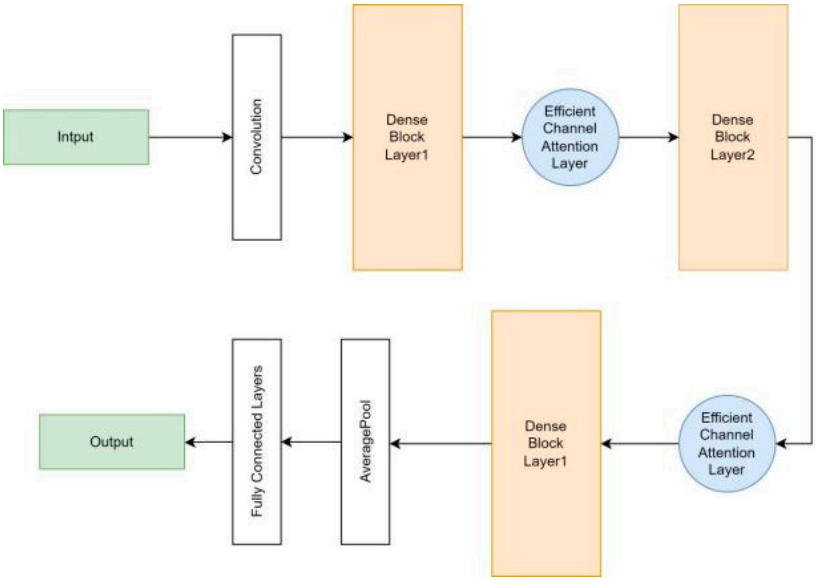


Figure 8. Schematic illustration of the overall architecture of the DenseNet (ECAD) network based on the attention mechanism.

3.6. Lab Environment

The experimental environment is presented in Table 1.

Table 1. Machine configuration information table.

Projects	Content
Central Processing Unit	Intel (R) Core (TM) i7-7700K CPU @ 4.20 GHz (Santa Clara, CA, USA)
Memory	32 G
Video card	NVIDIA GeForce GTX TITAN Xp (Santa Clara, CA, USA)
Operating System	Ubuntu 5.4.0-6ubuntu1~16.04.5
CUDA	Cuda 8.0 with cudnn
Data Processing	Python 2.7, OpenCV, and TensorFlow

3.7. Evaluation Standard

The following standard metrics were adopted here to measure the performance of the proposed system at different stages. It was defined as follows:

Dice Similarity Coefficient: This is a similarity measure between ground truth and predicted results.

$$DSC = \frac{2TP}{FP + 2TP + FN} \tag{7}$$

Accuracy: This is the proportion of the correctly predicted observations out of the total observations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

where

TP is defined as the positive sample predicted by the model as a positive class.
TN represents the negative samples predicted by the model as negative classes.
FP denotes the negative samples predicted by the model as positive.
FN refers to the positive sample predicted by the model as a negative class.

4. Results and Analysis

4.1. Parameter Settings

In experiments, a batch training approach was used to more effectively assess the discrepancy between the ground truth and the anticipated values. Other parameters include the following: loss function = cross-entropy loss; weight initialization method = Xavier; initialization bias = 0; initial learning rate of the model = 0.001; batch size = 16; momentum = 0.9; stochastic gradient descent (SGD) optimizer; and Softmax classifier. The model was reduced by 0.1 every 10 iterations. A total of 51 training epochs were completed, the input picture size was changed to 224 × 224, and the fused model was then combined with the final stored model.

4.2. Segmentation Process Method

In this part, the effectiveness of our suggested framework for automated segmentation and classification was assessed, and the findings are contrasted with the effectiveness of current approaches. The segmentation unit operates in two stages: only the multi-scale detection encoder–decoder network's performance was assessed in the first phase. The effectiveness was assessed in comparison to the techniques already in use, as indicated in Table 2. The encoder–decoder network and CRF module were merged in the second step, after which the performance was once again assessed. Only then were the results compared to the performance of the encoder–decoder system. Figure 9 compares the outcomes of the performance measures. The result was subsequently sent to a classification network for further processing.

Table 2. Segmentation performance comparison between our model and existing models.

Techniques	Accuracy (%)	Dice Score (%)
Proposed model	95.50	92.10
FrCN [25]	94.03	87.08
CNN-HRFB [26]	93.80	86.20
FCN [26]	92.70	82.30
iFCN [27]	95.30	88.64
DCL-PSI [28]	94.08	85.66

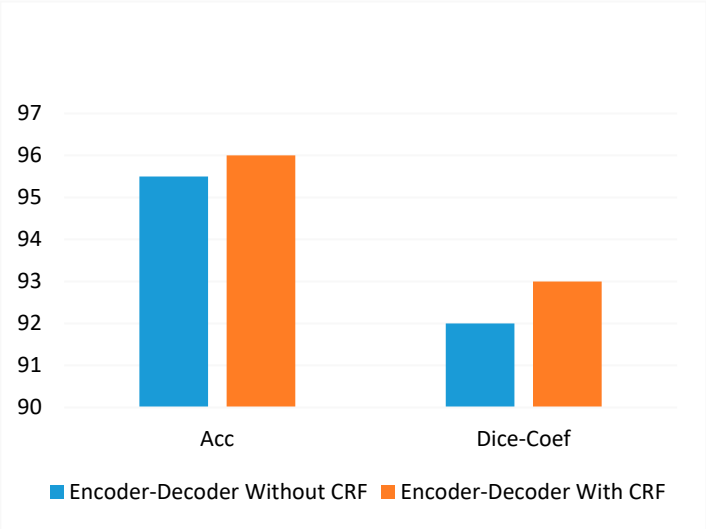


Figure 9. The figure compares the general performance of the encoder-decoder network using accuracy and Dice-coefficient when used with CRF and when used without CRF.

4.3. The Comparison of Different Classification Models

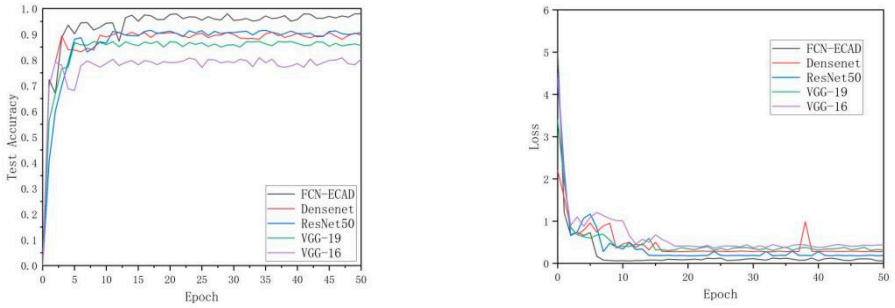
In order to verify the effectiveness of the model, different recognition models under the same experimental conditions were selected for comparative experiments. Since the standard deep convolutional neural network models VGG-16 [29], VGG-19 [29], ResNet50 [30], and DenseNet have achieved good recognition results in different fields, they were selected as the comparison model for this experiment. Table 3 shows the comparison results of each model test. The test accuracy and loss curves of the proposed model and the compared models are depicted in Figure 10.

Table 3. Classification experiment results of each model.

Techniques	Size (MB)	Loss (%)	Test Accuracy (%)	Test Time/Image (ms)
VGG-16	800.33	35.45	81.98	153.30
VGG-19	832.45	33.34	87.94	163.10
ResNet50	95.23	19.33	90.74	104.50
DenseNet	75.43	29.34	91.37	98.80
FCN-ECAD	143.50	6.55	98.28	69.20

As can be seen from Table 3, the test accuracy of the VGG16 model was 81.98%, which was obviously not suitable for this experiment. The recognition accuracy of the VGG19 model was higher than that of the VGG16 model, mainly because of the three-layer convolution added by VGG19 on the basis of VGG16. Hence, it can be inferred that the performance was improved. However, the amount of calculation and model memory also increased. Compared with VGG, the ResNet50 model based on different sparse structure designs reduced a large number of model parameters, and the performance was also significantly improved. The test accuracy rate reached 90.74%. The improved FCN-ECAD

model in this work has a test accuracy rate of 98.28%, which was obviously better than other models, the model size was also small, only 143.5 MB, and the memory usage was very small.



(**a**) (**b**)

Figure 10. Recognition accuracy and loss curves of the comparison models: (**a**) model test accuracy comparison and (**b**) model loss comparison.

Table 3 also lists the detection time of the model in this work and other deep network models for the recognition of a single insect image. Each deep neural network model carried out 10 tests on a single image, and finally, the average test time was taken as the test detection recorded the results. As can be seen from the test results, VGG-19 required the longest time to detect a single image, with an average detection time of 163.1 ms. ResNet50 and DenseNet single-sheet detection times were 104.5 ms and 98.8 ms, respectively. The FCN-ECAD model proposed in this work only took an average of 69.2 ms to detect a single image, which was more suitable for the rapid detection of rice pest images.

According to the above-mentioned table, it can be concluded that under this model, the classification of rice pest data can obtain a high accuracy rate. As can be observed from Table 4, three insects, namely HolotrichiadiomphaliaBates, Gryllidae, and mole cricket, have a higher recognition success rate. This is because compared with other insects, the characteristics of these three insects are more obvious, and the characteristics can be easily extracted for recognition. The accuracies of armyworm and cutworm were relatively low. This is because the data sets of these two insects have similar appearance characteristics in some body shapes, and it was difficult to extract effective feature points for identification. The brown rice planthopper had the lowest accuracy rate, because compared to other insects, the brown rice planthopper was too small in size, the pixels in the original data were very low, and the feature points were difficult to extract. Therefore, the accuracy of the developed model was not high. In the future, the detailed extraction of feature points should be further improved.

Table 4. Classification results of different kinds of pests.

Label	FCN-ECAD (%)	Vgg16 (%)	Vgg19 (%)	ResNet50 (%)	DenseNet (%)
C. suppressalis	98.89	81.23	86.56	88.97	90.43
N. aenescens	98.99	80.38	87.88	89.99	91.58
C. medinalis	97.28	79.99	87.56	89.69	90.43
N. lugens	96.42	75.25	80.95	85.63	87.63
A. ypsilon	97.33	81.98	86.89	90.45	91.22
G. sp	99.53	85.65	92.33	94.21	94.38
M. separata	97.35	78.65	88.15	91.59	89.92
H. armigera	98.49	82.79	87.68	90.24	90.34
Gryllidae	98.98	86.78	91.58	92.99	93.87
H. diomphalia	99.58	87.11	89.87	93.65	93.99
Average	98.28	81.98	87.94	90.74	91.37

Insects of the same family tend to have similar morphological characteristics. Hence, it was difficult to classify them precisely. Compared with other models, as shown in Table 4, when the model in this work recognized insects with large differences, the model accuracy can reach about 99%. When identifying insects with inconspicuous characteristics and small differences, the accuracy of the model can also reach about 98%. The algorithm in this work had a high recognition accuracy rate in the recognition of similar insects, which demonstrates that the proposed algorithm has good robustness. It also has high performance for object classification with similar features.

4.4. Impact of ECA on Model Performance

The improvement of the model can be divided into two cases: ECA with an attention mechanism and ECA without an attention mechanism. The three aspects of recognition accuracy, model size, and detection time of a single image were also compared. The results are shown in Table 5.

Table 5. Model experiment accuracy with or without ECA.

Techniques	Test Accuracy (%)	Size (MB)	Test Time/Image (ms)
Proposed model without ECA	94.34	142.67	68.40
Proposed model with ECA	98.28	143.50	69.20

The improved model had a higher accuracy rate on the training set and test set, and whether there is an attention mechanism ECA in the model has a greater impact on the accuracy of the model. After the attention mechanism ECA was added, the model's accuracy rate increased from 94.34% to 98.28%, which was a significant improvement. In addition, the model size only increased by less than 8 MB. This is because, during the process of feature extraction, the incorporation of the ECA attention mechanism can effectively strengthen the extraction of complex background insect features, prevent the occurrence of overfitting phenomenon, ensure that the network learns effective feature information, and greatly improve the accuracy rate.

5. Conclusions

This research offers some new approaches in segmentation and classification methods for rice insect pest images by using deep learning techniques. Firstly, we introduced a new encoder–decoder in the FCN and a series of sub-networks connected by jump paths that combine long jumps and shortcut connections for accurate and fine-grained insect boundary detection. Secondly, the network also integrates a CRF module for insect contour refinement and boundary localization, and finally, a novel DenseNet framework that introduces an ECA is proposed, focusing on extracting insect edge features for effective rice pest classification. The proposed model was tested on the data set collected in this paper with a final accuracy of 98.28%, showing a better performance than existing methods. Moreover, the model in this paper also maintains high model accuracy with good robustness in the classification of small target insects and insects with the same physical characteristics, while it can be demonstrated from our results that effective segmentation of insect images prior to classification can improve the detection performance of deep-learning-based classification systems. This paper solves the problem of the rice insect pest classification and provides a theoretical basis for the subsequent research of agricultural pest identification.

Author Contributions: Conceptualization, H.G. and T.L. (Tonghe Liu); methodology, H.G.; software, T.L. (Tonghe Liu); validation, T.L. (Tianye Luo), J.G., and R.F.; formal analysis, J.L.; investigation, X.M.; resources, Y.M.; data curation, Y.S.; writing—original draft preparation, H.G.; writing—review and editing, S.L. and Q.W.; supervision, T.H.; project administration, Y.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Changchun Science and Technology Bureau, funding number 21ZG27 (http://kjj.changchun.gov.cn) accessed on 1 July 2021, the Jilin Provincial Development and Reform Commission, funding number 2019C021 (http:/jidrc.jl.gov.cn) accessed on 1 January 2019, and the Department of Science and Technology of Jilin Province, funding number 20210302009NC (http://kjt.jl.gov.cn/) accessed on 1 November 2021.

Data Availability Statement: All new research data were presented in this contribution.

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

References

1. Zhang, B.; Zhang, M.; Chen, Y. Crop pest identification based on spatial pyramidal cisterification and deep convolutional neural network. *Chin. J. Agric. Eng.* **2019**, *35*, 209–215.
2. Zhang, Y.; Jiang, M.; Yu, P.; Yao, Q.; Yang, B.; Tang, J. Agricultural pest image recognition method based on multi feature fusion and sparse representation. *China Agric. Sci.* **2018**, *51*, 2084–2093.
3. Juan, C.; Liangyong, C.; Shengsheng, W.; Huiying, Z.; Changji, W. Landscape pest image recognition based on improved residual network. *J. Agric. Mach.* **2019**, *50*, 187–195.
4. Haitong, P.; Weiming, C.; Longhua, M.; Hongye, S. A review of pest identification techniques based on deep learning. *Agric. Eng.* **2020**, *10*, 19–24.
5. Gondal, M.D.; Khan, Y.N. Early pest detection from crop using image processing and computational intelligence. *FAST-NU Res. J.* **2015**, *1*, 59–68.
6. Hongzhen, Y.; Jianwei, Z.; Xiangtao, L.; Zuorui, S. Research on Image-Based Remote Automatic Recognition System of Insects. *J. Agric. Eng.* **2008**, *1*, 188–192.
7. Yinsong, Z.; Yindi, Z.; Muce, Y. Recognition and counting of insects on sticky board images based on the improved Faster-RCNN model. *J. China Agric. Univ.* **2019**, *24*, 115–122.
8. Feng, C.; Juntao, G.; Yulei, L.; Tianjia, H. A corn pest identification method based on machine vision and convolutional neural network in the cold region of Northeast China. *Jiangsu Agric. Sci.* **2020**, *48*, 237–244.
9. Xi, C.; Yunzhi, W.; Youhua, Z.; Yi, L. Image recognition of stored grain pests based on deep convolutional neural network. *Chin. Agric. Sci. Bull.* **2016**, *34*, 154–158.
10. Peng, S.; Guifei, C.; Liying, C. Image recognition of soybean pests based on attention convolutional neural network. *Chin. J. Agric. Mech.* **2020**, *41*, 171–176.
11. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–15 June 2015; pp. 3431–3440.
12. Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888–905.
13. Chen, M.; Wang, X.; Feng, B.; Liu, W. Structured random forest for label distribution learning. *Neurocomputing* **2018**, *320*, 171–182.
14. Boswell, D. *Introduction to Support Vector Machines*; Departement of Computer Science and Engineering University of California: San Diego, CA, USA, 2002.
15. Bao, W.; Yang, X.; Liang, D.; Hu, G.; Yang, X. Lightweight convolutional neural network model for field wheat ear disease identification. *Comput. Electron. Agric.* **2021**, *189*, 106367.
16. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. Supplementary material for 'ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 13–19.
17. Lafferty, J.; McCallum, A.; Pereira, F.C.N. *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2001.
18. Jin, X.B.; Su, T.L.; Kong, J.L.; Bai, Y.T.; Miao, B.B.; Dou, C. State-of-the-art mobile intelligence: Enabling robots to move like humans by estimating mobility with artificial intelligence. *Appl. Sci.* **2018**, *8*, 379.
19. 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.
20. Hu, F.; Xia, G.-S.; Hu, J.; Zhang, L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **2015**, *7*, 14680–14707.
21. Chaib, S.; Liu, H.; Gu, Y.; Yao, H. Deep feature fusion for VHR remote sensing scene classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4775–4784.
22. Li, E.; Xia, J.; Du, P.; Lin, C.; Samat, A. Integrating multilayer features of convolutional neural networks for remote sensing scene classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5653–5665.
23. Glorot, X.; Bordes, A.; Bengio, Y. Deep sparse rectifier neural networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings, Fort Lauderdale, FL, USA, 11–13 April 2011; pp. 315–323.
24. Zhipeng, L.; Libo, Z.; Qiongshui, W. Enhanced cervical cell image data based on generative adversarial network. *Sci. Technol. Eng.* **2020**, *20*, 11672–11677.

25. Al-Masni, M.A.; Al-Antari, M.A.; Choi, M.T.; Han, S.M.; Kim, T.S. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Comput. Methods Programs Biomed.* **2018**, *162*, 221–231.
26. Xie, F.; Yang, J.; Liu, J.; Jiang, Z.; Zheng, Y.; Wang, Y. Skin lesion segmentation using high-resolution convolutional neural network. *Comput. Methods Programs Biomed.* **2020**, *186*, 105241.
27. Öztürk, Ş.; Özkaya, U. Skin lesion segmentation with improved convolutional neural network. *J. Digit. Imaging* **2020**, *33*, 958–970.
28. Bi, L.; Kim, J.; Ahn, E.; Kumar, A.; Feng, D.; Fulham, M. Step-wise integration of deep class-specific learning for dermoscopic image segmentation. *Pattern Recognit.* **2019**, *85*, 78–89.
29. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
30. 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.

agronomy

MDPI

Article

Real-Time Measurement of Atmospheric CO_2, CH_4 and N_2O above Rice Fields Based on Laser Heterodyne Radiometers (LHR)

Jun Li [1,2,†], Zhengyue Xue [1,3,†], Yue Li [1,2], Guangyu Bo [2], Fengjiao Shen [3], Xiaoming Gao [1,2], Jian Zhang [4,5,*] and Tu Tan [2,*]

[1] School of Environmental Science and Optoelectronic Technology, University of Science and Technology of China, Hefei 230031, China
[2] Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China
[3] School of Advanced Manufacturing Engineering, Hefei University, Hefei 230601, China
[4] Faculty of Agronomy, Jilin Agricultural University, Changchun 130108, China
[5] Department of Biology, University of Columbia Okanagan, Kelowna, BC V1V 1V7, Canada
* Correspondence: jian.zhang@ubc.ca (J.Z.); tantu@aiofm.ac.cn (T.T.)
† These authors contributed equally to this work.

Abstract: High-precision observations provide an efficient way to calculate greenhouse gas emissions from agricultural fields and their spatial and temporal distributions. Two high-resolution laser heterodyne radiometers (LHRs) were deployed in the suburb of Hefei (31.9°N 117.16°E) for the remote sensing of atmospheric CO_2, CH_4 and N_2O above rice paddy fields. The atmospheric transmittance spectra of CO_2, CH_4 and N_2O were measured simultaneously in real time, and the atmospheric total column abundance was retrieved from the measured data based on the optimal estimation algorithm, with errors of 0.7 ppm, 4 ppb and 2 ppb, respectively. From July to October, the abundance of CO_2 in the atmospheric column that was influenced by emissions from rice fields increased by 0.7 ppm CH_4 by 30 ppb, and by 4 ppb N_2O. During the rice growth season, rice paddy fields play a role in carbon sequestration. CH_4 and N_2O emissions from paddy fields are negatively correlated. The method of baking rice paddy fields reduces CH_4 emissions from rice fields, but N_2O emissions from rice fields are usually subsequently increased. The measurement results showed that LHRs are highly accurate in monitoring atmospheric concentrations and have promising applications in monitoring emissions from rice paddy fields. In the observation period, rice paddy fields can sequester carbon, and CH_4 and N_2O emissions from rice fields are negatively correlated. The LHRs have strong application prospects for monitoring emissions from agricultural fields.

Keywords: laser heterodyne radiometer; carbon dioxide; methane; nitrous oxide; field measurement

Citation: Li, J.; Xue, Z.; Li, Y.; Bo, G.; Shen, F.; Gao, X.; Zhang, J.; Tan, T. Real-Time Measurement of Atmospheric CO_2, CH_4 and N_2O above Rice Fields Based on Laser Heterodyne Radiometers (LHR). *Agronomy* **2023**, *13*, 373. https://doi.org/10.3390/agronomy13020373

Academic Editor: Wei Zhang

Received: 14 December 2022
Revised: 9 January 2023
Accepted: 26 January 2023
Published: 27 January 2023

1. Introduction

Carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) are the three most important anthropogenic greenhouse gases (GHGs) in the atmosphere and play a role in global warming [1–3]. CO_2 has a lifetime in the atmosphere for several decades to a century [4]. The concentration of CO_2 in the atmosphere has increased from 178 parts per million (ppm) to 420 ppm since the Industrial Revolution, and has increased by 3 ppm per year in recent years. CH_4 and N_2O are classified as important greenhouse gases other than CO_2. The changes in CH_4 and N_2O concentrations are considered as critical factors that influence the atmospheric greenhouse effect. The lifetime of CH_4 in the atmosphere is about 12 years, and the global average total column abundance of CH_4 in 2020 was 1920 parts per billion (ppb). N_2O has a lifetime of 114 years, and the global average total column abundance of N_2O was about 340 ppb in 2020. Within 100 years, the warming potentials of CH_4 and N_2O will be 25 and 300 times higher than those of CO_2, respectively [5,6]. Moreover, nitric oxide (NO), the product of N_2O photolysis which occurs in the stratosphere,

depletes stratospheric ozone and leads to the destruction of the structure of the ozone layer. From 1990 to 2019, global agricultural GHG emissions increased by 17%. Agroforestry sources account for 22% of total global greenhouse gas emissions, whereas agricultural sources account for 14% of total global greenhouse gas emissions. Rice fields are considered one of the major sources of greenhouse gas emissions in agroecosystems [7]. Under climate, vegetation, soil and anthropogenic disturbance, the organic matter of farmland soil is decomposed by microorganisms into inorganic carbon and nitrogen, and the inorganic carbon is mostly released into the atmosphere as CO_2 under aerobic conditions and as CH_4 under anaerobic conditions. Nitrogen is converted to nitrate nitrogen by nitrifying bacteria, nitrate nitrogen is converted to various states of nitrogen oxides by denitrifying bacteria and N_2O is produced during both nitrification and denitrification. More than half of global N_2O is the product of nitrification and denitrification in soils [8]. China is a large rice growing country, with the second largest rice growing area in the world and great potential to achieve high yields and low emissions. Reducing GHG emissions from rice fields requires not only innovative rice management practices, but also accurate estimates of GHG emissions from rice fields and research on the spatial and temporal characteristics of greenhouse gas emissions. Monitoring the production, emission and transport of GHGs during rice production can help to assess the effectiveness of interventions to reduce emissions in rice production, and achieve a balance between abundant rice production and GHG emission reduction. It would provide assistance to promoting the green transformation of agricultural production as well [7].

The observation of greenhouse gas emissions from agricultural fields is mainly based on static chamber gas chromatography, which has long measurement intervals and makes it difficult to achieve real-time continuous measurements. Moreover, the eddy covariance technique obtains emission fluxes from agricultural fields by measuring the covariance of temperature, target gas concentration and vertical wind speed, which has the advantages of wide applicability and high reliability [9]. Gu et al. reported on a sensor based on the eddy covariance technique, which was designed to measure CO_2 and H_2O flux emissions at Jiangdu Agricultural Monitoring Station in Jiangsu Province [10]. In situ field observations of greenhouse gas emission fluxes from rice fields are scarce, and the simultaneous monitoring of CO_2, CH_4 and N_2O emission fluxes from rice fields is still lagging. Emission flux data measured in the field in some localized areas cannot be reliably extrapolated to regional and global scales and are not available for studying the spatial and temporal characteristics of GHG emissions from agricultural sources. High-precision observations of regional atmospheric GHG concentrations provide an effective way to calculate greenhouse gas emissions in large-scale rice paddy fields and their spatial and temporal distribution [11]. The successive measurement of atmospheric GHG column abundances over rice fields and the calculation of daily average column abundance changes can effectively monitor emissions of any given size rice field. LHR is a passive atmospheric gas remote sensing sensor that is portable, compact and has low power consumption and high sensitivity, which means that LHR could be easily deployed for long-term accurate greenhouse gas emission monitoring of rice field areas [12]. Since it was first applied to the detection of O_3 column concentration in the 1970s, it has been widely used in the measurement of vertical profiles and column concentrations of atmospheric greenhouse gases [13,14]. In recent years, LHR system integration has greatly improved with the development of low-cost lasers, waveguides and fiber technology and its application products. Wilson and coworkers [15] and Wang et al. [16] developed near infrared (NIR) LHR for the measurements of atmospheric vertical profiles of CO_2 and CH_4. Wang used the developed NIR LHR to measure CO_2 and CH_4 column concentrations in Hefei over a 6-month period, where the column abundance variation was mainly due to vehicle emissions. Weidmann et al. [17,18] and Xue [19] developed mid infrared (MIR) LHR using an inter-band cascade laser (ICL) as the local oscillator (LO) for the measurements of atmospheric column abundance of CH_4, N_2O and H_2O.

In this paper, high-precision and portable greenhouse gas measurement sensors NIR LHR and MIR LHR were developed in the laboratory and deployed in the suburbs of Hefei (31.9°N 117.16°E) to effectively monitor greenhouse gas emissions during the growth of rice. In the NIR LHR, two distributed feedback (DFB) lasers centered at 1.603 μm and 1.65 μm were used as LOs for CO_2 and CH_4 observations. In the two MIR LHR, inter-band cascade lasers centered at 3.53 μm and 3.93 μm were used as LOs for N_2O and CH_4 observations. The high-resolution atmospheric transmittance spectra of CO_2, CH_4 and N_2O were measured simultaneously, and the real-time column abundance was retrieved. The variation trend of atmospheric column concentration of three gases from July to October of 2022 was analyzed in combination with the planting change in the surrounding farmland. Based on the long-term real-time observation data, the effectiveness of rice production emission reduction interventions is evaluated and reported here. The specific experimental device, measurement and inversion results and data analysis will be presented later.

2. Materials and Methods

LHR, whether near-infrared (NIR) or mid-infrared (MIR) LHR, which extracts atmospheric molecular absorption information from sunlight via beating with a local oscillator (LO) on a photodetector, has been successfully applied to the remote sensing of the atmosphere in recent years [20]. The basic principle of LHR has been explained in detail in reference [1]. Here, the schematic diagram of LHR is shown in Figure 1a. Sunlight containing molecular absorption information is captured by a high-precision sun tracker. Then, almost perfectly aligned chop-modulated sunlight and LO are incident on the photodetector (PD). Molecular absorption information is transferred to the radio frequency (RF) component of the response current of the photodetector. The mean-square detector (M-S PD) converts the RF current into low-frequency voltage, and the atmospheric molecular absorption spectrum is obtained by demodulating the low-frequency voltage signal using a lock-in amplifier (LIA). It is worth noting that the resolution and the instrument line shape (ILS) of the LHR are determined by the electrical bandwidth of the filter, which allows for high spectral resolution. Atmospheric transmission spectra with a high resolution and signal-to-noise ratio will help with obtaining accurate atmospheric column abundances. A personal computer (PC) and high sampling rate data acquisition (DAQ) card serve for instrument control, data acquisition and processing.

The observation equipment used were laboratory-made highly integrated portable NIR and MIR LHR systems as shown in Figure 1b. The NIR LHR system was mainly composed of an all-fiber NIR LHR and an optical fiber sun tracker. In the all-fiber NIR LHR, two distributed feedback (DFB) lasers centered at 1.603 μm and 1.65 μm were used for CO_2 detection with a spectral resolution of 0.013 cm^{-1}, and for CH_4 detection one with a spectral resolution of 0.0067 cm^{-1} was used, respectively. The tracking precision of the fiber-optic solar tracker was 4×10^{-3} mrad, which is far smaller than the divergence angle of sunlight (9.6×10^{-3} rad). The alignment between the sunlight and LO was achieved by using the single-mode fiber coupler. The MIR LHR system was mainly composed of a free space solar tracker and a micro-electro-mechanical system (MEMS) modulator-based dual-channel MIR LHR. In this LHR, two inter-band cascade lasers (ICLs) centered at 3.53 μm and 3.93 μm were used as the LOs to probe the absorption lines of CH_4 with a resolution of 0.0047 cm^{-1} and of N_2O with a resolution of 0.0053 cm^{-1}, respectively. The two LHRs worked in the ground-based solar occultation mode. For the purpose of atmospheric molecular transmission spectrum detection, the laser wavenumber covered the molecular absorption range of interest by tuning the laser injection current at a fixed operating temperature. Both LHRs had CH_4 detection channels, which allowed for the validation of the accuracy of inversion results. Moreover, the reliability of the instrument and inversion algorithm have been effectively evaluated in previous work [16,20].

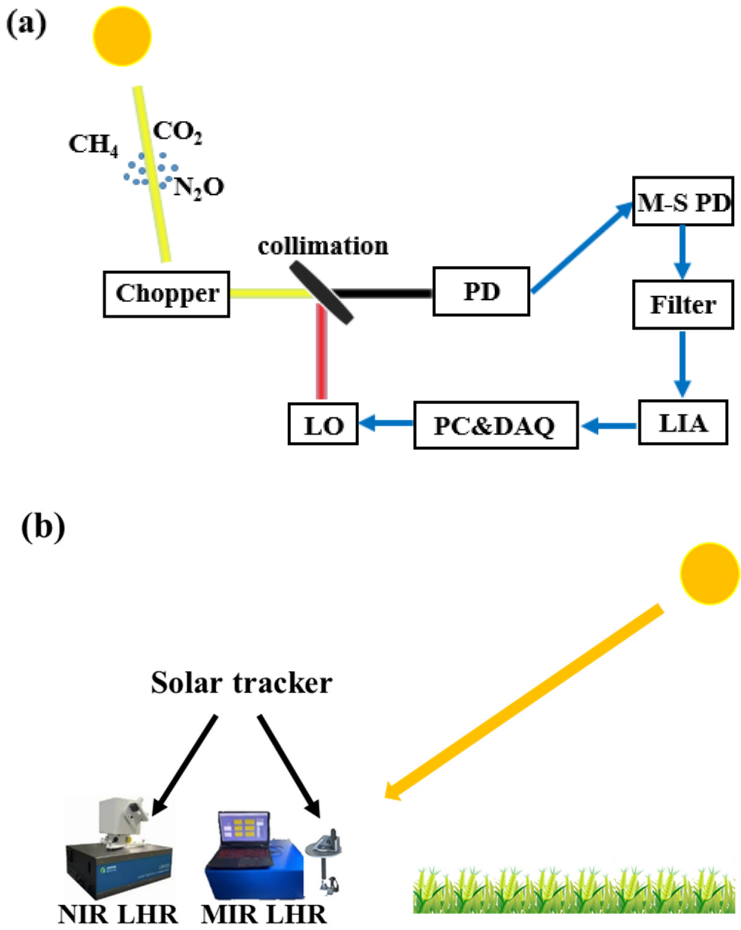


Figure 1. (**a**) Schematic diagram of LHR. LO: local oscillator, PD: photodetector, M-S PD: mean-square photodetector, LIA: lock-in amplifier, DAQ: data acquisition, PC: personal computer; (**b**) photo of the near-infrared laser heterodyne radiometer (NIR LHR) and mid-infrared laser heterodyne radiometer (MIR LHR).

The two LHRs were deployed in a suburb of Hefei (31.9°N 117.16°E) for atmospheric total column measurement of GHGs. In this work, the LO center frequency was scanned to probe the absorption line of (6238.77 cm^{-1}), CH_4 (at 6036.65 cm^{-1} and at 2831.92 cm^{-1}) and N_2O (at 2538.34 cm^{-1}). In order to obtain a high signal-to-noise (SNR) ratio atmospheric transmittance spectrum, a typical scan time was about 8 min. The observation period was from 1 July to 15 October 2022, which is basically consistent with the planting time of late season rice.

3. Results

The examples of atmospheric CO_2 (Figure 2a and CH_4: Figure 2b) transmittance spectra measured by NIR LHR are presented, and the SNR of CO_2 signal and CH_4 signal are 180 and 200, respectively.

The MIR-LHR measurement results of 3.53 μm (CH_4) and 3.93 μm (N_2O) detection channels are shown in Figure 3a (red curve) and Figure 3b (blue curve). The SNR of 3.53 μm and 3.93 μm detection channels are 70 and 90, respectively. The atmospheric transmittance spectrum of CH_4 is deformed, mainly affected by the water isotope (HDO) absorption. The measured atmospheric transmittance spectrum of HDO offers an efficient path to research the atmospheric total column abundance of water vapor, which will be the emphasis of future work. Compared with NIR LHR, the SNR of the CH_4 detection channel signal of the MIR LHR is significantly lower, which leads to large measurement errors [21–23].

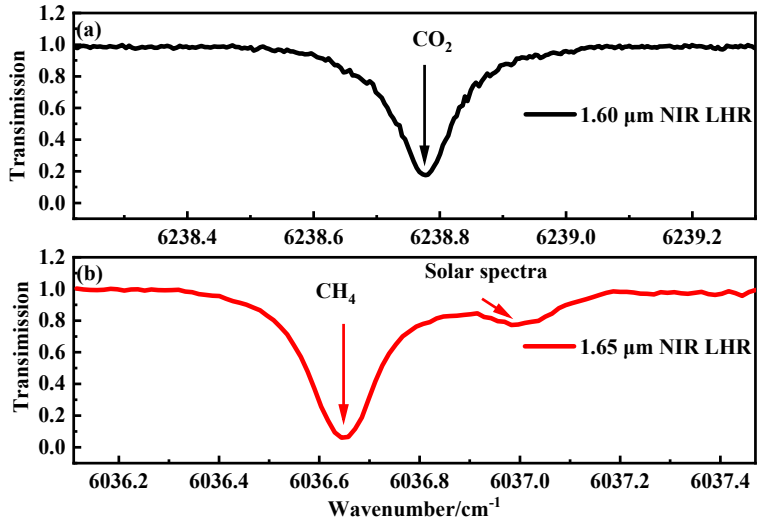


Figure 2. Atmospheric transmittance spectra measured by NIR LHR. (**a**) CO_2 at 6238.77 cm^{-1}; (**b**) CH_4 at 6036.65 cm^{-1}.

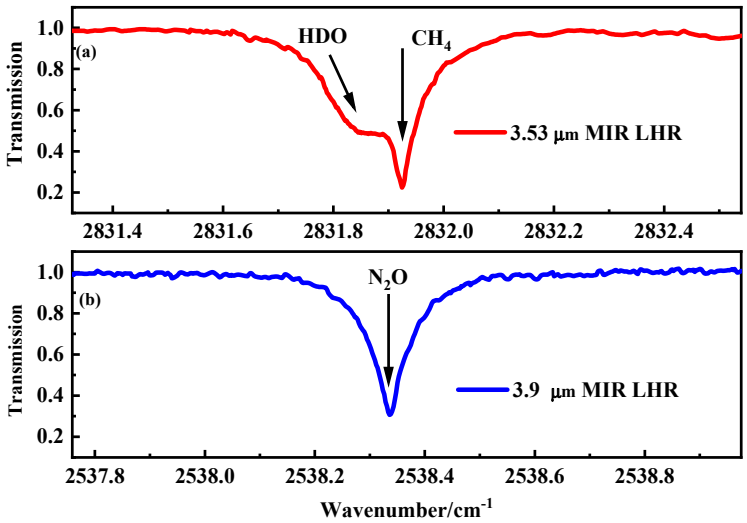


Figure 3. Atmospheric transmittance spectra measured by MIR LHR. (**a**) CH_4 at 2831.92 cm^{-1}; (**b**) N_2O at 2538.34 cm^{-1}.

LHR data retrieval was performed using the optimal estimation method (OEM), which was proposed by Rodgers for atmospheric remote sensing retrieval. Weidmann et al. gave a detailed introduction to the application of OEM in LHR data retrieval [24]. Herein, a brief recall is presented. The radiative transfer forward model (F) based on the reference forward model (RFM) was used to calculate the atmospheric transmittance spectrum of the solar radiance, which represents the absorption of solar radiation by greenhouse gases emitted from rice fields and by greenhouse gases in the atmosphere [25]. The relationship between the LHR data and the atmospheric state vector is described as:

$$y_m = F(x_n) + \varepsilon \tag{1}$$

where y_m is the measurement vector, x_n is the state vector including the temperature (T), pressure (P) and volume mixing ratio (VMR) and ε is the error vector. The OEM-based data retrieval is an iterative process to minimize the cost function ():

$$\chi^2 = (y_m - F(x_n))S_\varepsilon^{-1}(y_m - F(x_n))^T + (x_a - x_n)S_a^{-1}(x_a - x_n)^T \tag{2}$$

where S_ε is the measurement covariance matrix, S_a is the prior covariance matrix and x_a is the prior state. The Levenberg–Marquardt (LM) method was adopted in the iterative process. The iterative state vector x_{i+1} was calculated using Equation (3).

$$x_{i+1} = x_i + \left((1+\gamma)S_a^{-1} + K_i^T S_\varepsilon^{-1} K_i\right)^{-1} \times \left[K_i^T S_\varepsilon^{-1}(y_m - F_i) - S_a^{-1}(x_i - x_a)\right] \quad (3)$$

where K is the weighting function and γ is the LM parameter.

The flow-chart of LHR retrieval is displayed in Figure 4, which consists of a forward model and an inversion program. In the forward calculation, the atmospheric transmission spectrum ($F(x_n)$) combined with the corresponding weighting function (K) are calculated with the input atmospheric parameters (temperature (T), pressure (P) and volume mixing ratios ($VMRs$)), ILS, a priori state vector and solar zenith angle. In the inverse program, the forward model is iteratively called to minimize the cost function (Eq. (2)) following the LM algorithm (Eq. (3)); finally, the profile of the target gas is obtained. According to the retrieved profile, the atmospheric total column abundance of CO_2 is calculated to be ~428 ppm, the atmospheric total column abundance of CH_4 is found to be ~1940 ppb (NIR LHR) and ~1939 ppb (MIR LHR), and the atmospheric total column abundance of N_2O is found to be ~ 335 ppm, respectively. The detection accuracy of the LHRs is mainly affected by the error caused by the system noise and the retrieval error. Improving the signal-to-noise ratio of LHRs' signal and employing more accurate a priori data will improve measurement accuracy. In this study, considering the measurement error and retrieval error, the total errors are calculated to be ~0.2% (CO_2), ~0.23% (NIR LHR, CH_4), ~1% (MIR LHR, CH_4) and ~0.8% (N_2O), respectively.

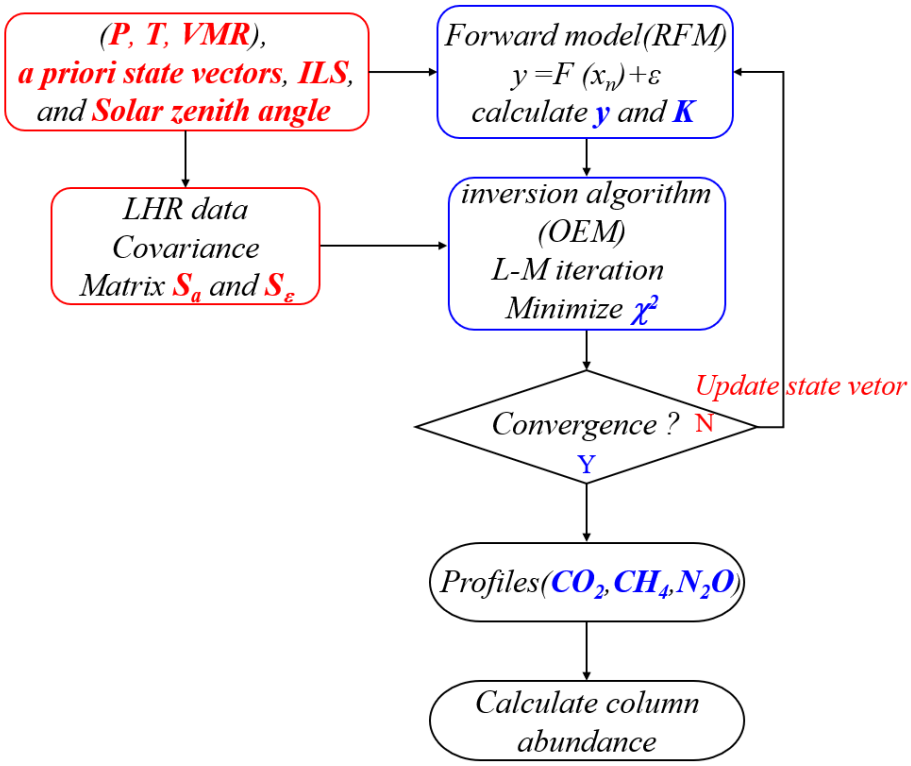


Figure 4. Flow-chart of the LHR retrieval.

4. Discussion

The farmland in the suburb of Hefei (31.9°N 117.16°E) is mainly planted with late season rice, and the planting period of late season rice is from mid-early June to mid-early October. Therefore, our observation period was from 1 July to 15 October 2022. It should be noted that LHRs extract information from solar radiation; therefore, only the results

measured on a clear, cloudless day were analyzed in this study. In addition, compared to the conventional indoor gas chromatography and eddy covariance techniques, the LHRs measure the total column abundance in the atmospheric column; they cannot measure the flux in the rice field. Since there are no other influential emission sources around the rice fields, the column abundance variation during the measurement period represents the total emission of a rice paddy area. A comprehensive understanding of the emission patterns of CO_2, CH_4 and N_2O and their interrelationships is a prerequisite and plays a key role in achieving greenhouse gas emission reduction in rice fields. In order to focus on the relationship between the emission in rice paddy fields and atmospheric column abundance of CO_2, CH_4 and N_2O, data measured from 1 July to 15 October 2022 are used to analyze the variation trend of atmospheric CO_2, CH_4 and N_2O column concentrations above paddy fields.

4.1. Atmospheric CO_2

To understand the trend of CO_2 emission from rice fields during daytime, we measured CO_2 concentration over rice fields from 10:00 a.m. to 15:00 p.m. on 1 July, 1 September, and 15 October 2022, and the measurement results are shown in Figure 5a. The high fluctuations in the total column abundance of CO_2 measured at 10:00–11:00 are probably due to the instability of atmospheric parameters (e.g., temperature and pressure) during this time period. Therefore, the daily average total column abundance mentioned in this manuscript was calculated from data measured after 11:00 on the day. During each day of measurement, the total column abundance of atmospheric CO_2 tends to decrease, which is mainly influenced by the consumption of atmospheric CO_2 via the photosynthesis of plants, such as rice.

The daily average total column abundance of CO_2 in the atmosphere above the rice paddy fields in the suburbs of Hefei from July to mid-October is shown in Figure 5b, where the gray shaded part indicates that the weather condition was cloudy or rainy on that day. The average monthly column abundance of CO_2 in July, August, September and October was 427.2, 427.4, 427.8 and 428.1 ppm, respectively. The standard deviations of total column abundance of CO_2 are calculated to be in the range of 0.6–1.4 ppm, which corresponds to an uncertainty region of 0.21–0.46%. The atmospheric total column abundance of CO_2 increased, but the growth rate of CO_2 concentration was less than the expectation reported by IPCC [8]. It can be inferred that the rice paddy field plays a certain role in carbon sequestration during the rice growing season. An increase in CO_2 emission from rice fields during the baking period and a decrease in soil CO_2 emission rate with decreasing temperature were not observed. The possible reasons for this are the lack of measurement accuracy of LHRs, the inability of the LHRs to make accurate measurements of the ground surface and atmospheric circulation.

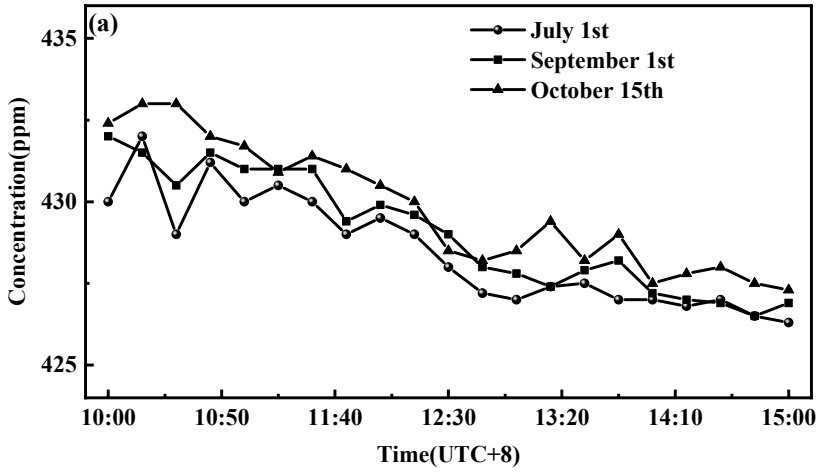


Figure 5. *Cont.*

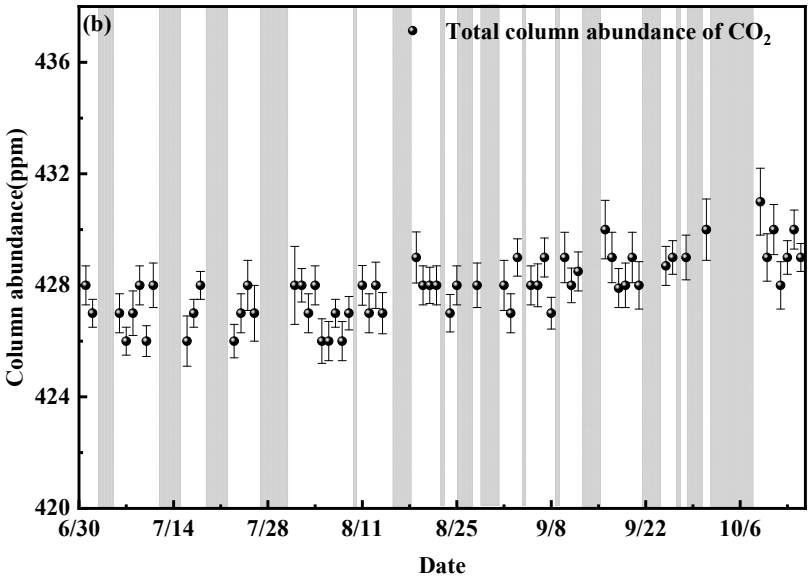


Figure 5. Total column abundance of CO_2. (**a**) Retrieval results of CO_2 from 10:00 am to 15:00 pm on 1 July, 1 September and 15 October 2022; (**b**) time series of the CO_2 retrieval results. Estimations of 95% confidence ranges are indicated.

4.2. Atmospheric CH_4

The average CH_4 emission flux of late season rice was much higher than that of early season rice. Emission fluxes from rice fields with late rice growth were greater than those with early rice growth. CH_4 emissions from rice fields were mainly concentrated in the tillering stage of rice growth (about late August to September), and decreased in the middle and late stages of rice growth, especially after baking, and slightly increased after rice fields were re-watered [26]. After that, the paddy field maintained low emissions until the rice harvest. The retrieval results of CH_4 above rice paddy fields in the suburbs of Hefei are shown in Figure 6, where the atmospheric column abundance of CH_4 increased from July and its high value appeared at the end of August. The standard deviations are found to be in the range of 5–12 ppb for the total column abundance of CH_4, which corresponds to an uncertainty region of 0.18–0.5%. During the rice growing season, the total atmospheric column abundance of CH_4 above the rice fields increased by 30 ppb. From early July to the end of August when rainfall is abundant and rice fields are in a flooded state, inorganic carbon is released into the atmosphere as CH_4 in the absence of oxygen. In addition, the increase in soil temperature promotes the production and emission of CH_4. Despite the presence of more rainy days in September and October, the methane concentration decreased significantly due to baked fields and lower temperatures. The CH_4 column concentration decreased from mid-September due to the gradual diffusive mixing of atmospheric CH_4 from the lower to middle layers and the horizontal atmospheric motion. CH_4 emissions from rice fields decline after October as rice has been harvested. During the non-rice growing season, CH_4 column concentrations over rice fields will slowly increase due to other anthropogenic sources such as fossil fuel and landfill emissions.

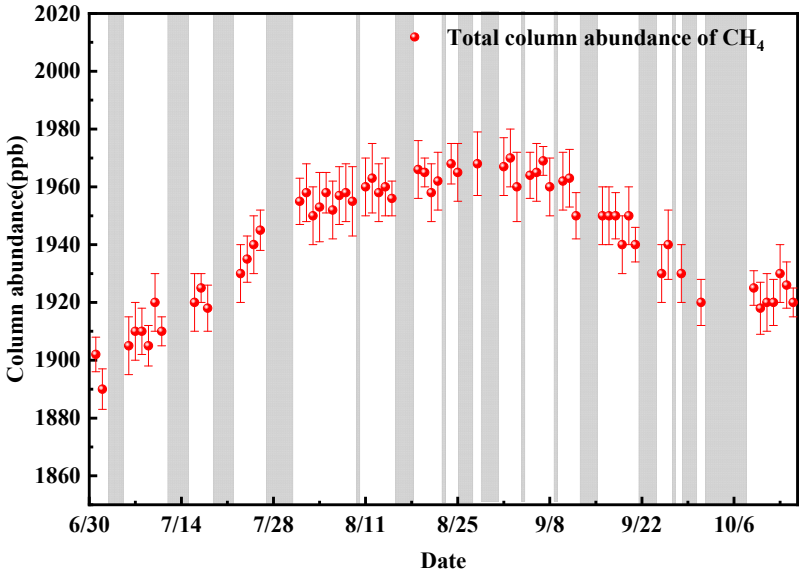


Figure 6. Time series of the CH_4 retrieval results. Estimations of 95% confidence ranges are indicated.

4.3. Atmospheric N_2O

The production and emission of N_2O in rice paddy fields are affected by soil temperature, humidity and PH. The relevant research shows that rice plants are the main channel for N_2O transmission. The atmospheric total column abundance of N_2O with a wave-like trend from July to mid-October above rice paddy fields around Hefei is displayed in Figure 7. The standard deviation of the N_20 total column is between 2 and 4 ppb with a corresponding uncertainty range of 0.3–0.8%. The purple shaded part indicates that there was a thunderstorm shower on that day. The dates were 4, 5, 28, 30 July and 14 September. Affected by paddy irrigation, N_2O emissions from paddy fields decreased. The total column abundance of atmospheric N_2O gradually decreased from July. The atmospheric column abundance of N_2O increased from the end of July to early August and in September. During this period, the paddy field changed from flooding to drying, and the soil redox potential increased, which was conducive to the conversion of ammonia to N_2O and greatly promoted the emission of N_2O. In early October, nitrification was inhibited by strong cold air and precipitation, which led to a decrease in N_2O emissions from paddy fields. As rice is harvested, although bare farmland is a weak N_2O source, it can be predicted that atmospheric N_2O total column abundance will decrease further in late October; after that, the variation in atmospheric N_2O concentration will no longer be dominated by farmland emissions. It is worth noting that when lightning occurs, nitrogen and oxygen in the air will form nitric acid through an ionization reaction and chemical reaction, reducing the PH of the soil, which will improve the efficiency of nitrification and promote the formation of N_2O. However, significant variation in atmospheric N_2O column abundance is not observed in Figure 3. On the one hand, several consecutive days' precipitation before and after the thunderstorm (28 July, 30 July) inhibits the nitrification in the soil. On the other hand, the retrieval results, which are limited by the measurement accuracy of the MIR LHR, cannot reflect the weak variation in atmospheric N_2O column abundance.

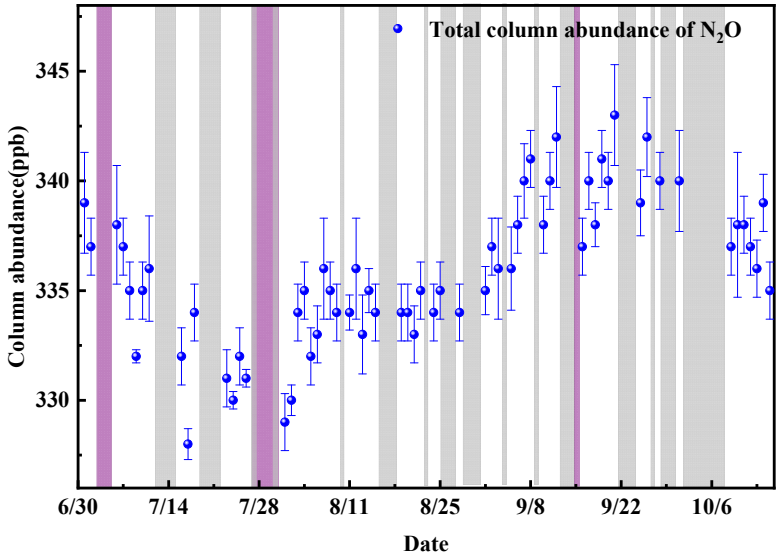


Figure 7. Time series of the N_2O retrieval results. Estimations of 95% confidence ranges are indicated.

As shown in Figures 5–7, there was no significant correlation between CO_2 and CH_4 emissions throughout the rice growing season. The emission patterns of CH_4 and N_2O in rice fields, especially the peak emission period, were mainly influenced by the moisture and soil temperature of rice fields, and the emissions of CH_4 and N_2O in rice fields were negatively correlated. The relationship between CH_4 and N_2O emissions in paddy fields is mainly shown by irrigation management and the amount of nitrogen fertilizer applied. Field drying and water-saving irrigation are key tools used to reduce CH_4 emissions from rice fields. The mid-term baking of paddy fields decreases CH_4 emissions while inevitably increasing N_2O emissions. Therefore, the appropriate timing of field baking is critical and the reduction effect should be assessed in terms of combined CH_4 and N_2O warming potential. The optimization of fertilizer management and the adjustment of the ratio of fertilizer nitrogen to potassium to phosphorus can help to control N_2O emissions from rice fields. During the rice growing period, a combination of soil properties, tillage practices, temperature and humidity are considered to maximize the reduction effect. We suggest that rice varieties with less greenhouse gas emissions should be promoted in rice seed selection, and rice fields should be kept as dry as possible to reduce CH_4 emissions during the early rice growth period, and shallow irrigation with chemical fertilizers should be used to suppress N_2O emissions during the late rice growth period. Considering the year-round CH_4 and N_2O emissions from rice fields, water management of rice fields and emission monitoring of paddy fields should not be neglected during the non-rice growing period.

5. Conclusions

In this study, NIR LHR and MIR LHR were deployed in the suburb of Hefei (31.9°N 117.16°E) for the remote sensing of GHGs above rice paddies from 1 July to 15 October 2022. The atmospheric total column abundance of CO_2, CH_4 and N_2O above rice paddies was obtained with errors of 0.7 ppm, 4 ppb and 2 ppb, respectively. Based on long-term measurement results, we found that atmospheric CO_2 concentrations were increasing, and inferred that rice paddy fields can play a role in carbon sequestration and that CH_4 and N_2O emissions from rice paddy fields are the main contributors to increasing atmospheric greenhouse gas concentrations during the rice growing season. With the advantages of portability and high accuracy, the LHRs can be used to simultaneously measure multiple GHG emissions above a large area of farmland. However, when using LHRs to detect GHG concentrations, disadvantages such as the inability to use LHRs at night and the inability to measure farmland emission fluxes limit the further application of LHRs in farmland GHG emission monitoring. The inability to monitor GHGs above rice paddy fields in

real time during thunderstorms and cloudy and rainy days greatly limited our ability to research the influence of weather on rice field emissions. In future work, we will focus on the combined use of multiple technologies, including chamber gas chromatography, the eddy covariance technique and LHRs. We would like to obtain technology to be used in high-precision flux monitoring of GHGs in small areas of farmland throughout the day under a variety of weather conditions using chamber gas chromatography and the eddy covariance technique, while the LHRs will be used for GHG concentration monitoring over large areas of farmland. Real-time monitoring of rice paddy fields' GHG emissions during the rice growing season and non-rice growing season is achieved by establishing a three-dimensional observation network. The effectiveness of various emission reduction measures will be verified. Additionally, the relationship between farmland gas emissions and farmland yield and quality will be further determined.

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

Funding: This research was funded by the Key Project of the National Natural Science Foundation of China (NO. 41730103), Natural Science Foundation of Anhui Province (NO. 2208085QF218) and University Natural Science Research Program of Anhui Province (NO. 2022AH051795).

Data Availability Statement: Not applicable.

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

References

1. Fan, K.; Yan, Y.; Xu, D.; Li, S.; Zhao, Y.; Wang, X.; Xin, X. Methane and Nitrous Oxide Fluxes with Different Land Uses in the Temperate Meadow Steppe of Inner Mongolia, China. *Agronomy* **2022**, *12*, 2810. [CrossRef]
2. McGill, B.M.; Hamilton, S.K.; Millar, N.; Robertson, G.P. The Greenhouse Gas Cost of Agricultural Intensification with Groundwater Irrigation in a Midwest U.S. Row Cropping System. *Glob. Chang. Biol.* **2018**, *24*, 5948–5960. [CrossRef] [PubMed]
3. Duval, B.D.; Martin, J.; Marsalis, M.A. The Effect of Variable Fertilizer and Irrigation Treatments on Greenhouse Gas Fluxes from Aridland Sorghum. *Agronomy* **2022**, *12*, 3109. [CrossRef]
4. Knorr, W. Is the airborne fraction of anthropogenic CO_2 emissions increasing? *Geophys. Res. Lett.* **2009**, *36*, L21710. [CrossRef]
5. Wang, G.; Liang, Y.; Zhang, Q.; Jha, S.K.; Gao, Y.; Shen, X.; Sun, J.; Duan, A. Mitigated CH_4 and N_2O Emissions and Improved Irrigation Water Use Efficiency in Winter Wheat Field with Surface Drip Irrigation in the North China Plain. *Agric. Water Manag.* **2016**, *163*, 403–407. [CrossRef]
6. Pereira, J.L.S.; Perdigão, A.; Marques, F.; Wessel, D.F.; Trindade, H.; Fangueiro, D. Mitigating Ammonia and Greenhouse Gas Emissions from Stored Pig Slurry Using Chemical Additives and Biochars. *Agronomy* **2022**, *12*, 2744. [CrossRef]
7. IPCC. Available online: https://www.ipcc.ch/report/sixth-assessment-report-cycle/ (accessed on 10 November 2022).
8. Zhu, X.; Burger, M.; Doane, T.A.; Horwath, W.P. Ammonia oxidation pathways and nitrifier denitrification are significant sources of N_2O and NO under low oxygen availability. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 6328–6333. [CrossRef]
9. Levy, P.; Drewer, J.; Jammet, M.; Leeson, S.; Friborg, T.; Skiba, U.; Oijen, M. Inference of spatial heterogeneity in surface fluxes from eddy covariance data: A case study from a subarctic mire ecosystem. *Agric. For. Meteorol.* **2020**, *280*, 107783. [CrossRef]
10. Gu, M.; Chen, J.; Mei, J.; Tan, T.; Wang, G.; Liu, K.; Liu, G.; Liu, X. Open-path anti-pollution multi-pass cell-based TDLAS sensor for the online measurement of atmospheric H2O and CO2 fluxes. *Opt. Express* **2022**, *30*, 43961–43972. [CrossRef]
11. Feng, M.Y.; Zhang, G.; Xia, L.J.; Xiong, X.; Li, B.Z.; Kong, P.; Zhan, M.J.; Zhang, Y.X. Spatial and temporal distribution of atmospheric methane in middle-low reaches of Yangtze River based on satellite observations. *Geochimica* **2021**, *50*, 121–132.
12. Wilson, E.L.; DiGregorio, A.J.; Riot, V.J.; Ammons, M.S.; Bruner, W.W.; Carter, D.; Mao, J.; Ramanathan, A.; Strahan, S.E.; Oman, L.D.; et al. A 4 U laser heterodyne radiometer for methane (CH_4) and carbon dioxide (CO_2) measurements from an occultation-viewing CubeSat. *Meas. Sci. Technol.* **2017**, *28*, 035902. [CrossRef]
13. Zenevich, S.; Gazizov, I.; Churbanov, D.; Plyashkov, Y.; Spiridonov, M.; Talipov, R.; Rodin, A. A Concept of 2U Spaceborne Multichannel Heterodyne Spectroradiometer for Greenhouse Gases Remote Sensing. *Remote Sens.* **2021**, *13*, 2235. [CrossRef]
14. Shen, F.; Wang, G.; Wang, J.; Tan, T.; Wang, G.; Jeseck, P.; Te, Y.V.; Gao, X.; Chen, W. Transportable mid-infrared laser heterodyne radiometer operating in the shot-noise dominated regime. *Opt. Lett.* **2021**, *46*, 3171. [CrossRef] [PubMed]
15. Wilson, E.L.; DiGregorio, A.J.; Villanueva, G.; Grunberg, C.E.; Souders, Z.; Miletti, K.M.; Menendez, A.; Grunberg, M.H.; Floyd, M.A.M.; Bleacher, J.E.; et al. A portable miniaturized laser heterodyne radiometer (mini-LHR) for remote measurements of column CH_4 and CO_2. *Appl. Phys. B-Lasers Opt.* **2019**, *125*, 211. [CrossRef]

16. Wang, J.; Sun, C.; Wang, G.; Zou, M.; Tan, T.; Liu, K.; Chen, W.; Gao, X. A fibered near-infrared laser heterodyne radiometer for simultaneous remote sensing of atmospheric CO_2 and CH_4. *Opt. Lasers Eng.* **2020**, *129*, 106083. [CrossRef]
17. Robinson, I.; Butcher, H.L.; Macleod, N.A.; Weidmann, D. Hollow waveguide-miniaturized quantum cascade laser heterodyne spectro-radiometer. *Opt. Express* **2021**, *29*, 2299–2308. [CrossRef] [PubMed]
18. Weidmann, D.; Perrett, B.J.; Macleod, N.A.; Jenkins, R.M. Hollow waveguide photomixing for quantum cascade laser heterodyne spectro-radiometry. *Opt. Express* **2011**, *19*, 9074–9085. [CrossRef] [PubMed]
19. Xue, Z.Y.; Shen, F.J.; Li, J.; Liu, X.H.; Wang, G.S.; Liu, K.; Gao, X.M.; Chen, W.D.; Tan, T. MEMS Modulator-Based Mid-Infrared Laser Heterodyne Radiometer for Atmospheric Remote Sensing. *Front. Phys.* **2022**, *10*, 945995. [CrossRef]
20. Xue, Z.; Shen, F.; Li, J.; Liu, X.; Wang, J.; Wang, G.; Liu, K.; Chen, W.; Gao, X.; Tan, T. A MEMS modulator-based dual-channel mid-infrared laser heterodyne radiometer for simultaneous remote sensing of atmospheric CH_4, H_2O and N_2O. *Opt. Express* **2022**, *30*, 31828–31838. [CrossRef]
21. Sun, C.; Wang, G.; Zhu, G.; Tan, T.; Liu, K.; Gao, X. Atmospheric CO_2 column concentration retrieval based on high resolution laser heterodyne spectra and evaluation method of system measuring error. *Acta Phys. Sin.* **2020**, *69*, 144201. [CrossRef]
22. Clarke, G.B.; Wilson, E.L.; Miller, J.H.; Melroy, H.R. Uncertainty analysis for the miniaturized laser heterodyne radiometer (mini-LHR) for the measurement of carbon dioxide in the atmospheric column. *Meas. Sci. Technol.* **2014**, *25*, 055204. [CrossRef]
23. Shen, F.; Wang, G.; Xue, Z.; Tan, T.; Cao, Z.; Gao, X.; Chen, W. Impact of Lock-In Time Constant on Remote Monitoring of Trace Gas in the Atmospheric Column Using Laser Heterodyne Radiometer (LHR). *Remote Sens.* **2022**, *14*, 2923. [CrossRef]
24. Weidmann, D.; Reburn, W.J.; Smith, K.M. Retrieval of atmospheric ozone profiles from an infrared quantum cascade laser heterodyne radiometer: Results and analysis. *Appl. Opt.* **2007**, *46*, 7162–7171. [CrossRef] [PubMed]
25. Dudhia, A. The reference forward model (RFM). *J. Quant. Spectrosc. Radiat. Transf.* **2017**, *186*, 243–253. [CrossRef]
26. Zou, J.; Liu, S.; Qin, Y.; Pan, G.; Zhu, D. Sewage irrigation increased methane and nitrous oxide emissions from rice paddies in southeast China. *Agric. Ecosyst. Environ.* **2009**, *129*, 516–522. [CrossRef]

agronomy

MDPI

Article

Rapid and Accurate Prediction of Soil Texture Using an Image-Based Deep Learning Autoencoder Convolutional Neural Network Random Forest (DLAC-CNN-RF) Algorithm

Zhuan Zhao [1,2], Wenkang Feng [3,4], Jinrui Xiao [3,4], Xiaochu Liu [3,4], Shusheng Pan [1,2] and Zhongwei Liang [5,*]

1 School of Physics and Materials Science, Guangzhou University, Guangzhou 510006, China
2 Research Center for Advanced Information Materials (CAIM), Huangpu Research and Graduate School of Guangzhou University, Guangzhou 510006, China
3 School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou 510006, China
4 Guangzhou Key Laboratory of Nontraditional Machining and Equipment, Guangzhou 510006, China
5 Guangdong Engineering Research Centre for Highly Efficient Utility of Water/Fertilizers and Solar-Energy Intelligent Irrigation, Guangzhou University, Guangzhou 510006, China
* Correspondence: liangzhongwei@gzhu.edu.cn

Abstract: Soil determines the degree of water infiltration, crop nutrient absorption, and germination, which in turn affects crop yield and quality. For the efficient planting of agricultural products, the accurate identification of soil texture is necessary. This study proposed a flexible smartphone-based machine vision system using a deep learning autoencoder convolutional neural network random forest (DLAC-CNN-RF) model for soil texture identification. Different image features (color, particle, and texture) were extracted and randomly combined to predict sand, clay, and silt content via RF and DLAC-CNN-RF algorithms. The results show that the proposed DLAC-CNN-RF model has good performance. When the full features were extracted, a very high prediction accuracy for sand (R^2 = 0.99), clay (R^2 = 0.98), and silt (R^2 = 0.98) was realized, which was higher than those frequently obtained by the KNN and VGG16-RF models. The possible mechanism was further discussed. Finally, a graphical user interface was designed and used to accurately predict soil types. This investigation showed that the proposed DLAC-CNN-RF model could be a promising solution to costly and time-consuming laboratory methods.

Keywords: soil texture; identification; DLAC-CNN-RF model; accuracy

Citation: Zhao, Z.; Feng, W.; Xiao, J.; Liu, X.; Pan, S.; Liang, Z. Rapid and Accurate Prediction of Soil Texture Using an Image-Based Deep Learning Autoencoder Convolutional Neural Network Random Forest (DLAC-CNN-RF) Algorithm. *Agronomy* **2022**, *12*, 3063. https://doi.org/10.3390/agronomy12123063

Academic Editors: Jian Zhang, Randy G. Goebel and Zhihai Wu

Received: 21 October 2022
Accepted: 2 December 2022
Published: 3 December 2022

1. Introduction

In recent decades, new techniques have resulted in significant progress in agriculture. Precision agriculture has become increasingly important in agricultural cultivation and management. Soil is the fundamental productive resource that provides a suitable environment for seed germination and root growth. Soil texture is one of the most important and fundamental parameters to consider when it comes to soil because it greatly influences its physical, chemical, and biological properties [1,2], such as the degree of water and air penetration, nutrient absorption, susceptibility to erosion, and germination. Thus, exact soil texture identification is required for precise agriculture and soil management. A thorough grasp of soil textural heterogeneity can benefit sound agricultural practice in various growing situations.

Conventional mechanical methods used for soil texture analysis entail numerous complex steps, such as drying, crushing, and sieving [3]. Hydrometers and pipettes are extensively used mechanical methods [4]. Although these techniques provide accurate soil textural analysis results, they are time-consuming. Moreover, these methods use H_2O_2 as a corrosive reagent, which is harmful to the environment. The soil textural report can be produced with a large dynamic range and flexibility using an advanced

laser diffraction particle size analyzer, but it induces high costs and sampling errors [5,6]. Therefore, new techniques are urgently needed for rapid, cheap, and sophisticated soil texture measurement [7].

Currently, the approaches used for the fast and low-cost measurement of the soil textural parameters rely on extracted texture and color features for further prediction and classification. Recent evidence suggests that the particle size also can be used to identify soil texture using X-ray [8], laser [9], infrared spectroscopy [10], and gamma radiation [11]. For instance, Vohland et al. [12] reported an environmentally friendly method of predicting sand and clay contents using diffuse reflectance infrared Fourier transform spectroscopy (DRIFS) without any chemical agents. Nevertheless, the above-mentioned approaches have not been recognized as standard methods because precise and complicated equipment is required. Furthermore, the particle size distribution should be predetermined before the measurement [13].

With the development of digital technology, cameras and smartphones are gaining popularity in predicting soil parameters, particularly in developing countries with limited budgets. Smartphones have many advantages, including portability, low-cost, and good image acquisition capability. Furthermore, modern digital technology promotes the development of computer vision and deep learning, which are increasingly used for soil image prediction and classification [14]. However, the complex preprocessing steps for preparing soil thin sections remain a problem. A method that combines microscopic image capturing with a continuous wavelet transform (CWT)-based computer vision algorithm has been proposed by Sudarshan et al. [15] to predict soil texture both in situ and ex situ. In those smartphone-based soil image identification studies, soil profile [16], digital RGB photography (combined with neural network modelling) [17], and digital image processing and multivariate image analysis [18] have been explored. Nevertheless, high accuracy is still difficult to realize. There is a lack of comprehensive methods in which soil images can be effectively exploited for soil texture prediction.

As a powerful ensemble learning, random forest (RF) is attracting considerable interest for classification and regression by constructing a multitude of decision trees. Scientists have already proven that RF ensemble exhibits superior prediction than a single tree [19]. For instance, Dornik et al. [20] performed a successful classification of soil types using geographic object-based image analysis. Convolutional neural network (CNN) is also a powerful algorithm widely used in image sensing and object detection [21,22]. A large amount of data from soil images can be successfully processed by CNN. For example, Swetha et al. realized a high prediction accuracy for clay (98%) and sand (98%) and moderate prediction accuracy (75%) for silt using CNN algorithms [23]. Rahim Azadnia et al. [24] predicted soil images at a distance of 20, 40, and 60 cm with an accuracy of 99.89%, 99.81%, and 99.58%, respectively, via a CNN model. However, only 11 soil textures were considered in this work, and the prediction accuracy of the proposed CNN model for the remaining soil texture is unknown.

In this study, a low-cost image acquisition system was constructed, and a method combining the DLAC, CNN, and RF algorithms was established to predict all the soil textures. Image features (color, particle, and texture) with random combinations were extracted. An environmentally friendly graphical user interface has been designed to generate the results. Surprisingly, the proposed DLAC-CNN-RF model obtained a very high average prediction accuracy (99.67% for all the soil textures). This approach provides a promising solution for the accurate identification of soil texture.

2. Materials and Methods

2.1. Sample Preparation

Soil samples were extracted from different areas in Guangzhou City, Guangdong Province (22.26° N~23.56° N, 112°57 E~114.3° E), including the Panyu, Huadu, Conghua, Haizhu, Nansha, Huangpu, and Zengcheng districts (See Figure 1). These regions have the required conditions for growing agricultural products due to their suitable climate

and abundant water resources. A total of 1000 samples were taken from the 50 locations distributed in the different city districts. The locations include paddy fields, vegetable gardens, dry land, cane land, and plantain land. Various depths of the soils were collected by a standard shovel, and the depth varied within a range of 0–15 cm from the ground soil surface. To guarantee the purity of the samples, after each sampling, the shovel was cleaned before taking the next sample. The locations of the collected samples were input into a global positioning system receiver (Garmin eTrex 20×) for geolocation analysis. All the samples were collected in a sealed pocket with a label and then shipped back to the Solar Energy Intelligent Irrigation Equipment Technology Innovation Center Laboratory of Guangzhou University for processing.

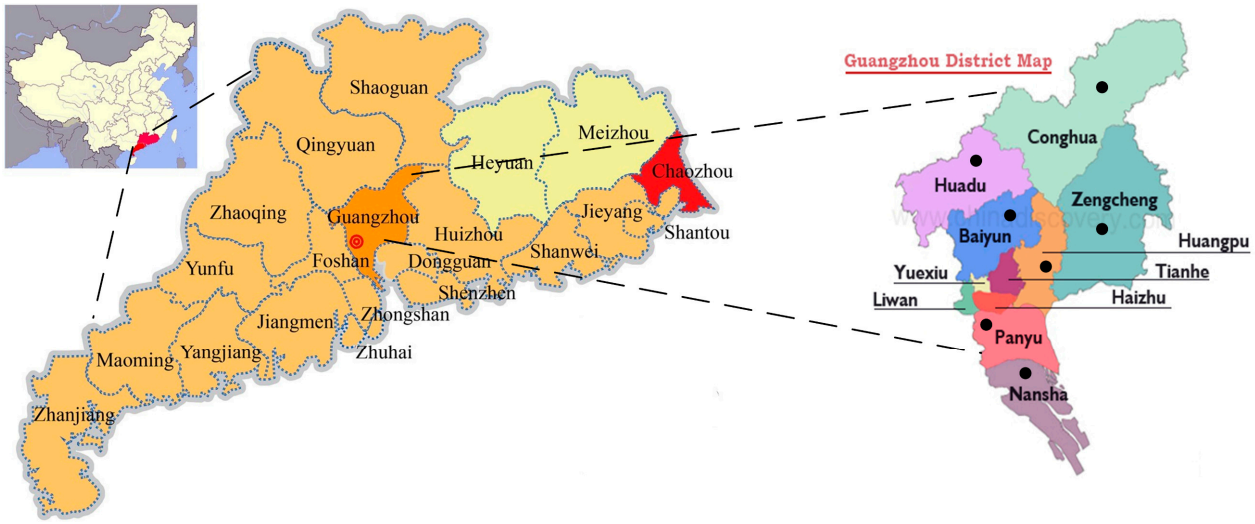


Figure 1. The location where the sample was collected.

All the samples were initially air-dried for 24 h with a light intensity of 1.5×10^4 Lux and a wind speed of 16 m/s. Then, the samples were run through a 2 mm sieve to eliminate plant debris and tiny particles. A hydrometer (ASTM mode 152H) method was utilized to measure the average percentage of the sand, silt, and clay particles in each soil texture, mainly because this method gives the benefits of simplicity, reasonable price, and rapid detection. Finally, 12 soil texture types, consisting of clay, silty clay, silty clay loam, sandy clay, sandy clay loam, clay loam, silt, silty loam, loam, sandy, loamy sand, and sand loam, were prepared for imaging.

2.2. Image Acquisition

To alleviate ambient light interference, an image acquisition device was designed and fabricated for imaging the soil samples (See Figure 2). A smartphone holder was constructed at the top of the box, and a circular window was opened to hold the camera for taking images. A black chamber (12 cm × 8 cm × 5 cm) was designed to prevent reflection light from entering the box. A rectangular-shaped holder was installed on the bottom of the box to hold the soil samples. Two LED strips with a total lumen of 80 were mounted on both sides of the box to illuminate the surrounding environment and prevent shadows. The light intensity can be adjusted by the outside button of the box. A Huawei Mate 40 Pro smartphone with a DXOMARK camera was employed to capture the images (8192 × 6144 pixels, f/1.9,f/1.8,f/3.4 in). The images were taken using the high dynamic range (HDR) mode and a landscape scene mode, which helps to capture more details in the shadows. Additionally, natural light and a LEICA standard filter were used to increase the contrast. The distance between the camera and the sample was set at 4.5 cm. All the captured images were saved as PNG files and cropped by a coded python programming language for further processing.

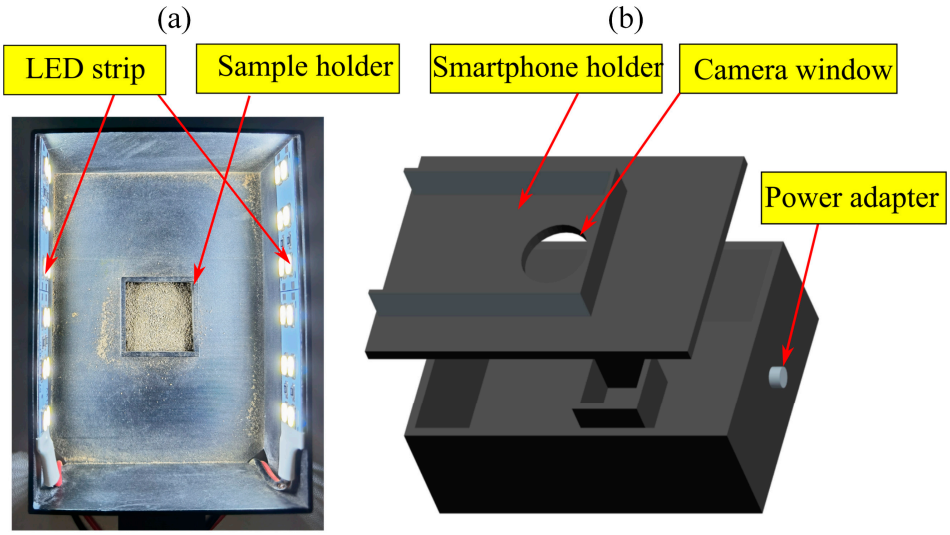


Figure 2. The designed image acquisition device; (**a**) the constructed dark chamber from top view; (**b**) 3D view of the complete assembly.

2.3. Image Feature Extraction

The physical properties of soil are typically characterized by colors and textures in the images. However, the particle features can also reflect the physical properties that need to be better investigated. In this work, a random combination of particle (Threshold binarization), color (HSV (hue, saturation, and value) and Hu moments), and textures features (Local binary patterns (LBP) and Haralick features) was extracted for soil image identification and classification, including particle, color, texture, particle + color, particle + texture, color + texture, and color + particle+ texture. Figure 3 reports the extracted particle, color, and texture features for silt, clay, and sand, respectively.

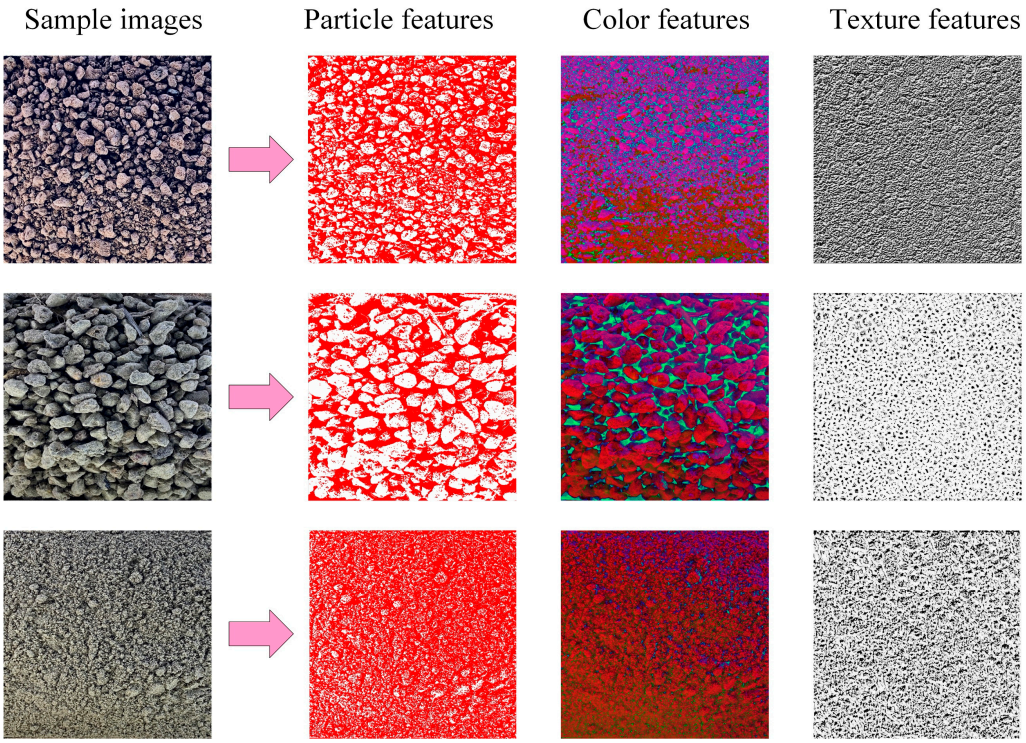


Figure 3. The extracted image features from the soil samples for further processing.

Figure 4 reports the flowchart of the processing of the extracted features for image identification. For threshold binarization, the OTSU threshold segmentation method was adopted; a threshold value of 120 and 2 channels were adjusted to transfer original soil particles to red in the images, which helps to calculate the soil particles area by reading the number of pixels. The RGB color of the sample images was converted into HSV components and reflected by a 512-dimension feature vector in HSV components. In addition, Hu moments are a set of 7 numbers calculated utilizing central moments that are constant under picture modifications. The LBP algorithm was used to characterize the texture feature of the soil images because it has been found to be an efficient approach to the traditional structural models of texture analysis. A 'skimage' package was used to generate the LBP histogram values through the texture, with a fixed number of 24 for circular symmetric neighbor set points and a radius of 8 pixels for each circle, producing 26 texture features. The Haralick feature algorithm was used to quantify an image according to its texture, and a total of 7 textural features were calculated for the following global features. Before particle, color, and texture features extraction, all the cropped soil images were processed into a grayscale image. A total of 554 global features were produced by particle (2 features), color (519 features), and texture (33 features), which were then imported into the proposed DLAC-CNN-RF model for further analysis.

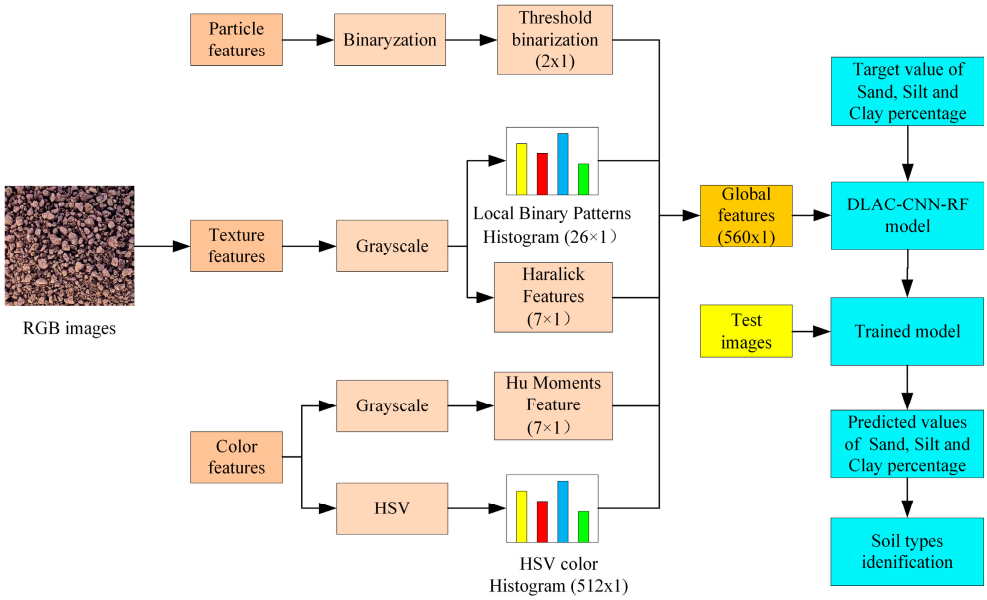


Figure 4. Flowchart showing the image feature representation process applied to the captured images, including threshold binarization, local binary pattern histogram, Haralick features, Hu moments, and HSV color histogram.

2.4. Developing a DLAC-CNN-RF Model

The DLAC is an excellent autoencoder that can be used for determining image properties since it incorporates probabilistic index extraction and process quantification in a general-purpose prediction system. A traditional autoencoder (AE) is trained to encode the input vector of image parameters $x \in [0, 1]$, into a hidden representation $y \in [0, 1]$. The image characteristics are denoted as unlabeled data.

$$\left\{ I_d \left[\left({N_{fx}}^k, {N_{fy}}^k \right); \mu, \sigma^2 \right] \right\}_{fx=1, fy=1}^{M_k, N_k} \quad (1)$$

so that the input can be reconstructed from that representation.

$$y = f_\theta\,(x) = \sigma_1(Wx + b) \quad (2)$$

where $f_\theta(x)$ is called the encoder, $\theta = [W_x, b]$, W_x is a weight matrix, and b is a bias vector; σ_1 is a nonlinear activation function for the encoder. The resulted hidden representation y is then mapped back to a reconstructed vector $z \in [0, 1]$:

$$z = g_{\theta'} \qquad (y) = \sigma_2(W'y + b') \tag{3}$$

where, $g_{\theta'}(y)$ is called the decoder of the image parameters, $\theta' = [W', b']$ with appropriately sized parameters, σ_2 is a nonlinear activation function for the decoder. Here, DLAC takes the parametric input and encodes it to a linear representation of the soil characteristics. On the other side, the decoder takes hidden representations, passes them into nonlinearity, and generates the output of the probabilistic indexes.

$$L(x, z) = -\sum_{k=1}^{D} (x_k \log z_k - (1 - x_k) \log(1 - x_k)) \tag{4}$$

DLAC tries to encode the parametric input stochastically applied to the input of the traditional autoencoder. It first uses a stochastic mapping $\widetilde{x} \cong q_D(\widetilde{x}|x)$ to encode the parametric input into a hidden representation $y = f_\theta(\widetilde{x}) = \sigma_1(W\widetilde{x} + b)$ from which we reconstruct $z = g_{\theta'}(\widetilde{y}) = \sigma_2(W\widetilde{y} + b')$. Similar to a traditional autoencoder, the network weights are trained to minimize the average computation error, but the key difference is that z is now a deterministic function rather than x. Each layer of DLAC captures the complicated, higher-order correlations between the activities of the hidden features so that the input initialization can be utilized for the initial training of the input layer of the DLAC network.

Besides, the monitoring variables are divided into N labeled datasets: $(x(1), y^*(1))$, $(x(2), y^*(2)), \ldots, (x(N), y^*(N))$ and M unlabeled datasets: $x(N + 1), x(N + 2), \ldots, x(N + M)$, where $M >> N$, y is identified by the initial grinding or optimal soil value determinations. The correlation function of the intelligent prediction is defined as follows:

$$\begin{array}{l} \widetilde{x}, \widetilde{z}^{(1)}, \ldots, \widetilde{z}^{(L)}, \widetilde{y} = g_{\theta'}(\widetilde{x}) \\ x, z^{(1)}, \ldots, z^{(L)}, y = g_\theta(\widetilde{x}) \\ \widehat{x}, \widehat{z}^{(1)}, \ldots, \widehat{z}^{(L)}, \widehat{y} = W_{\theta'}(\widehat{z}^{(1)}, \ldots, \widehat{z}^{(L)}) \end{array} \tag{5}$$

In the forward path, individual layers of DLAC are formalized into linear transformations, then the nonlinear activation function is applied as

$$\widetilde{h}^{(l)} = Activation(\gamma^{(l)}(z^{(l)} + \beta^{(l)})) \tag{6}$$

Here, $h(l)$ is the postactivation function and $W(l)$ is the weight matrix. The γ and β are the shifting and scaling parameters used before applying the nonlinearity function. The batch normalization is used to accelerate deep-learning network training. Finally, the prediction cost of DLAC is defined as

$$C(z^{(l)}, \hat{z}^{(l)}) = \left\| \frac{\hat{z}^{(l)} - \mu^{(l)}}{\sigma^{(l)}} - \hat{z}^{(l)} \right\|^2 \tag{7}$$

where $\mu(l)$ and $\sigma(l)$ are the mean and standard deviation of the encoder samples. This encoder is optimized by the objective correlation function to improve the accuracy of the intelligent prediction. In order to model each computation neutron coming out of the prediction network, generative adversarial network is used with competing behaviors for minimizing the training error and improving computational reliability. A successful DLAC training is one that gets to the same predictive cost and encoding function $f_{\theta'}(y)$. There is

an error signal checking the similarity rate $y = f_\theta(\widetilde{x}) = \sigma_1(W\widetilde{x} + b)$. The training function for this network can be formulated as follows:

$$C(z, \rho^{(z)}) = -E_{x\sim P_R(x)}[\log \gamma(l)] - E_{s\sim P_g}[\log(1 - \sigma(z(l))] \quad (8)$$

Equally, the training function for E with setting parameters $\rho(G)$ would be as follows:

$$C(\omega, \rho^{(G)}) = E_{S\sim P_R(S)}[\log(1 - \rho(\lambda_l C(\widetilde{z}^{(l)} n)))] \quad (9)$$

Combining these functions into a single frame, both λ_l and $z^{(l)}n$ would be trained and converge to a stable state of a Nash equilibrium. This means that the most optimal networks (generator and discriminator networks) could be reached as follows [25]:

$$Cost = -\sum_{n=1}^{N} \log P(\widetilde{y}(n) = y*(n)|x(n)) + \omega \sum_{n=N+1}^{M} \sum_{l=1}^{L} \lambda_l C(z^{(l)} n, \hat{z}^{(l)} n) \quad (10)$$

Here $\tilde{y}$ is the error output, y^* is the true target, λ_l is the cost multiplier, which represents the weight of DLAC loss function for each decoding layer, and ω is a weight to balance different losses. This network illustrates different operational levels and data transmission in the architecture of DLAC. It could be learned that DLAC ranges across the measurement, calculation, and prediction levels. After training the first level of DLAC, the learned encoding function $f_\theta(y)$ is used on image parametric input x. Furthermore, a logistic regression layer can be added on the top of encoders to achieve a set of supervised network learning.

The established DLAC-CNN-RF model is shown in Figure 5. Soil images (256 × 256) are given as the inputs to the network and reconstructed by the autoencoders. Two convolution layers and three Maxpool layers were constructed to extract the image features. A flattened layer was used to transfer the 2D outputs of the max-pooling layer into 1D outputs. The convolution layers were connected to the Maxpool layers and attached to the full connected layers. Convolution 1 and 2 have a filter length of 48 and 128, respectively. A kernel function (11 × 11) was used for the layer of Convolution 1, while the remaining layers adopted the kernel function of 3 × 3. In particular, two max-pooling layers and two full connected layers were employed before and after the flattened layer to classify the extracted features. Finally, an RF model was employed to predict the soil types according to the random combination of image features. Among the 1000 soil image samples, 700 were used for training, and 300 samples were used for testing the network.

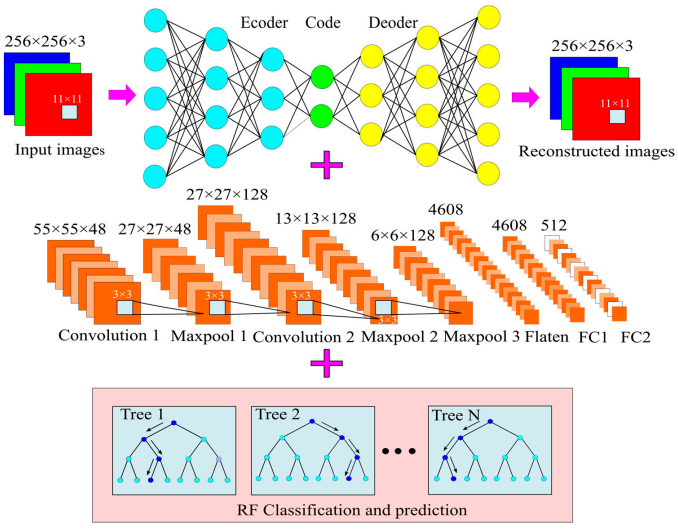


Figure 5. The proposed DLAC-CNN-RF network architecture.

2.5. Developing a Graphical User Interface

An image recognition system named "Soil type identification" was developed based on basic Java programming language (JDK 1.8), where the pretrained DLAC-CNN-RF and RF model was deployed. A user-friendly interface was designed using Visual Studio Code software, as depicted in Figure 6. First, an account should be registered before logging into the system. After logging in, Baidu map was connected through an application programming interface, which helps to identify the geographical location of the input images. The input images can be selected from the camera or album. Once the image has been confirmed, the proposed DLAC-CNN-RF model responds quickly, as shown in Figure 5. The percentage of sand, clay, and silt was clearly presented in the interface, and the soil identification records were saved.

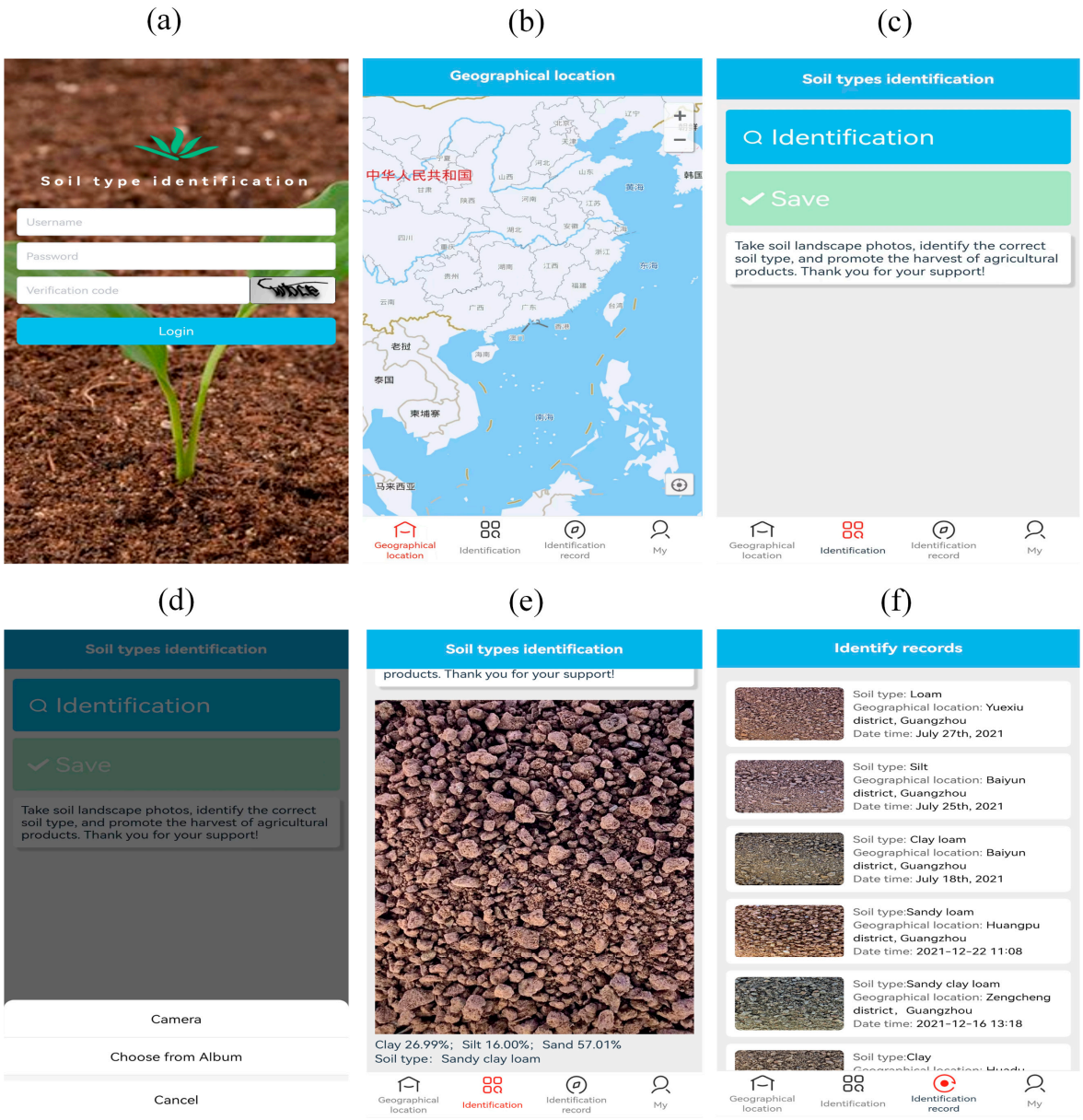


Figure 6. The running screen of the image recognition system for predicting soil texture: (**a**) the main menu; (**b**) the map of geographic location. Notably, the Chinese characters on the map represent different provinces in China, and the used Baidu map in this research only supports Chinese; (**c**,**d**) image selection; (**e**) DLAC-CNN-RF model predicted soil textural values; (**f**) identify records.

3. Results and Discussion

Figure 7 reports the average percentage of sand, silt, and clay in the prepared samples, which was measured via a hydrometer. The relationship of soil texture and the percentage of sand, silt, and clay was determined according to Stoke's law [26]. The soil classification was based on USDA soil taxonomy [27]. It can be seen that the sand samples have the lowest silt and clay values, as well as the highest sand values with 3.5%, 6.1%, and 91.4%, respectively. The lowest sand values (10.3%) were obtained in the silt samples. Also, the sand, clay, and silt percentage dominated in the sand, clay, and silt samples, respectively.

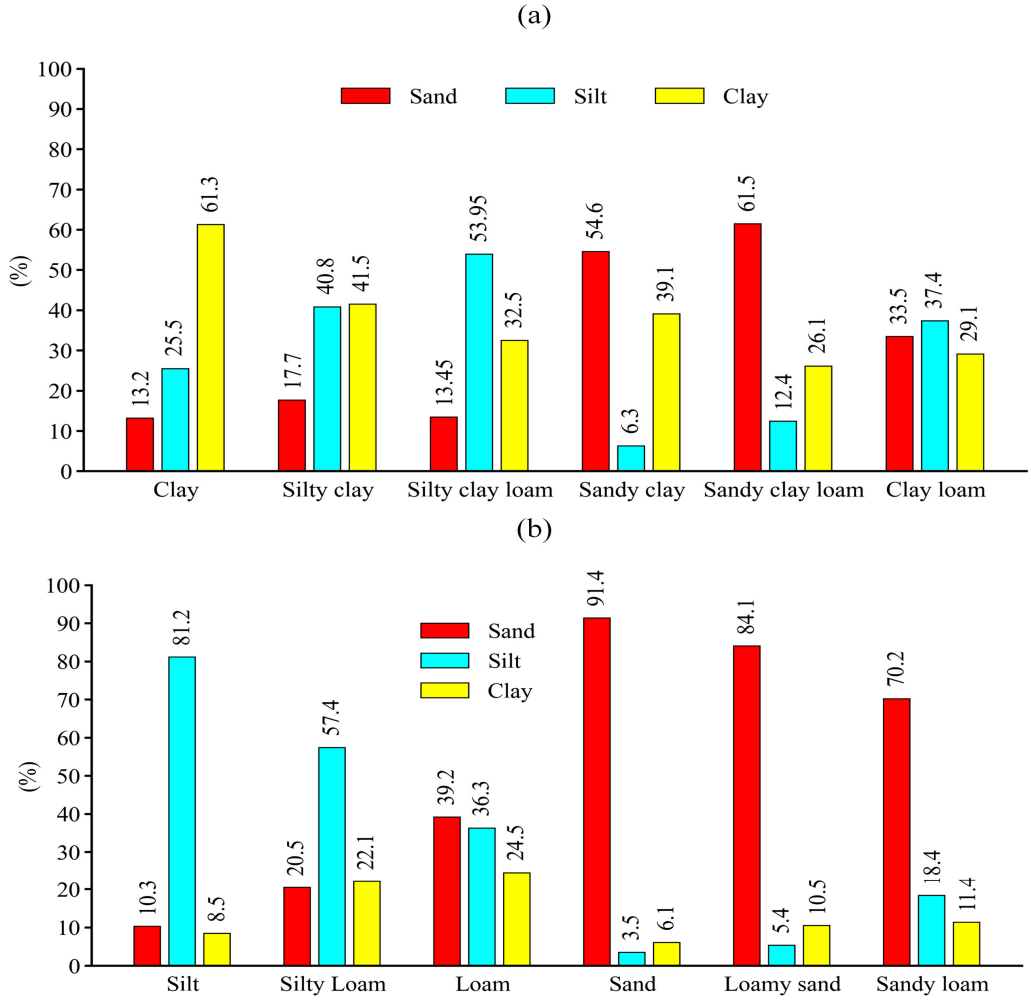


Figure 7. (**a**) The measured average percentage of clay, silt, and sand particles for clay, silty clay, silty clay loam, sandy clay, sandy clay loam, and clay loam, respectively; (**b**) the measured average percentage of clay, silt, and sand particles for silt, silty loam, loam, sand, loamy sand, and sandy loam, respectively; the number of samples counted for clay, silty clay, silty clay loam, sandy clay, sandy clay loam, clay loam, silt, silty loam, loam, sand, loamy sand, and sandy loam are 84, 79, 74, 75, 84, 102, 74, 97, 79, 87, 89, and 76, respectively.

Table 1 reports the RF and DLAC-CNN-RF model validation statistics for predicting clay, silt, and sand using different extracted features. Obviously, the DLAC-CNN-RF model shows better performance in predicting soil textures due to a higher value of R^2. When the color feature is extracted, the RF and DLAC-CNN-RF model almost reach an agreement in predicting sand (i.e., R^2 0.95 and 0.96) and clay (i.e., R^2 0.93 and 0.94), while when predicting silt, a significant improvement was obtained in R^2 using the CCN-RF model. The value of R^2 improves from 0.79 to 0.96. Similarly, the DLAC-CNN-RF model gives obvious benefits to silt prediction when the texture and particle features were extracted; an R^2 value of 0.94 was obtained, while RF only realized an R^2 of 0.73. It's noted that both the RF and

DLAC-CNN-RF models show better performance in predicting sand and clay when a single image feature is extracted. This can be attributed to the fact that the average percentage of sand and clay in these samples is much more than that of silt. When multiple features were extracted, both the RF and DLAC-CNN-RF models showed progress in predicting all the soil textures. For sand and clay prediction, both the RF and DLAC-CNN-RF model show good performance when two features were extracted, although the DLAC-CNN-RF model is a little bit superior, while for silt prediction, the DLAC-CNN-RF model results in a higher R^2 value, with a 12~15% improvement in comparison to the RF model. Surprisingly, when the full features were extracted, the DLAC-CNN-RF mode exhibited an R^2 value of approximately 99% for all the soil textures. The better performance of the DLAC-CNN-RF model can be interpreted as a deeper study of the various edges, lines, and corners of the image. Notably, the validation RMSEs for predicting clay ranged from 3.71–3.86% among all the tested DLAC-CNN-RF models, which is a greater performance since clay typically shows a higher uncertainty in traditional measurements. This is mainly due to the use of extremely higher resolution cameras in smartphones. In addition, the established optimal networks (eq 10) offer a more quick and more efficient way than traditional autoencoders, which significantly reduce the noise of input data. Thus, the RMSEs values become lower.

Table 1. The RF and DLAC-CNN-RF model validation statistics for predicting clay, silt, and sand by using different combinations of image features. The number of the clay, silty clay, silty clay loam, sandy clay, sandy clay loam, clay loam, silt, silty loam, loam, sand, loamy sand, and sandy loam samples used for training were 58, 56, 53, 51, 56, 72, 53, 58, 57, 61, 61, and 54, respectively, and a total number of 700 was used.

	Extracted Feature	Model	MAE	RMSE	R^2
Sand	Color	RF	3.67	4.44	0.95
		DLAC-CNN-RF	3.45	3.81	0.96
	Texture	RF	3.69	4.45	0.95
		DLAC-CNN-RF	3.48	3.85	0.96
	Particle	RF	3.74	4.53	0.94
		DLAC-CNN-RF	3.49	3.86	0.96
	Color + Texture	RF	3.58	4.35	0.96
		DLAC-CNN-RF	3.39	3.73	0.98
	Color + Particle	RF	3.62	4.37	0.96
		DLAC-CNN-RF	3.42	3.78	0.97
	Particles + Texture	RF	3.64	4.39	0.95
		DLAC-CNN-RF	3.44	3.80	0.97
	Color + Particle + Texture	RF	3.55	4.24	0.97
		DLAC-CNN-RF	3.37	3.71	0.99
Silt	Color	RF	3.81	4.46	0.79
		DLAC-CNN-RF	3.58	3.89	0.96
	Texture	RF	3.83	4.49	0.78
		DLAC-CNN-RF	3.59	3.94	0.94
	Particle	RF	3.89	4.57	0.73
		DLAC-CNN-RF	3.61	3.96	0.94
	Color + Texture	RF	3.73	4.40	0.85
		DLAC-CNN-RF	3.51	3.81	0.97
	Color + Particle	RF	3.74	4.43	0.82
		DLAC-CNN-RF	3.52	3.85	0.97
	Particles + Texture	RF	3.77	4.44	0.81
		DLAC-CNN-RF	3.55	3.88	0.96
	Color + Particle + Texture	RF	3.70	4.37	0.88
		DLAC-CNN-RF	3.48	3.79	0.98

Table 1. *Cont.*

	Extracted Feature	Model	MAE	RMSE	R^2
Clay	Color	RF	3.68	4.68	0.93
		DLAC-CNN-RF	3.55	3.93	0.94
	Texture	RF	3.70	4.72	0.91
		DLAC-CNN-RF	3.48	3.84	0.97
	Particle	RF	3.74	4.75	0.90
		DLAC-CNN-RF	3.49	3.85	0.96
	Color + Texture	RF	3.63	4.61	0.97
		DLAC-CNN-RF	3.51	3.88	0.96
	Color + Particle	RF	3.64	4.65	0.95
		DLAC-CNN-RF	3.41	3.77	0.97
	Particles + Texture	RF	3.67	4.66	0.95
		DLAC-CNN-RF	3.45	3.81	0.98
	Color + Particle + Texture	RF	3.59	4.57	0.97
		DLAC-CNN-RF	3.46	3.83	0.98

Compared with the traditional method of predicting sand, soil, and clay, the proposed DLAC-CNN-RF model produces a better performance than that of Qi et al. [28] (R^2 values of 0.77, 0.68, and 0.71 for sand, silt, and clay, respectively) and Swetha et al. [23] (R^2 values of 0.97–98, 0.96–98, and 0.62–0.75 for sand, silt, and clay, respectively). The proposed method also exhibited a better prediction performance than Aitkenhead et al. [17] (R^2 values of 0.25, 0.19, and 0.18 for sand, silt, and clay, respectively). Additionally, this study showed a lower RMSE than Minasny et al. [29] (RMSE values of 6.31% and 6.23% for sand and clay, respectively).

Figure 8 reports the RF and DLAC-CNN-RF model predicted plots using full image features. In general, the proposed DLAC-CNN-RF mode shows higher accuracy in predicting all the soil types; the measured and predicted values are closer compared to the distribution in the RF models, especially for sand prediction, where the DLAC-CNN-RF model predicted value and measured value almost reach an agreement. Thus, the proposed DLAC-CNN-RF model appears to be preferable for predicting soil types.

The confusion matrix used for evaluating the performance of the DLAC-CNN-RF model is shown in Figure 9. The elements on the diagonal indicate that the predicted values equal the actual values, and these samples are classified correctly. The nonzero elements on the off-diagonal part of the matrix demonstrate a wrong classification. In most cases, the proposed model performs well in soil type identification. However, six images were classified incorrectly. For example, images in the clay, loam, sand, sandy clay loam, silt, and silty clay loam class were classified as loam, sandy loam, silty loam, loamy sand, clay loam, and silty loam, respectively.

Table 2 shows the mean values of the classified performance parameters, including accuracy, precision, sensitivity, specificity, and area under the curve (AUC). An average accuracy of 99.67% was obtained using the proposed DLAC-CNN-RF model. The key parameters that determine the accuracy are sensitivity and specificity. Generally, the sensitivity parameter denotes how a model detects a positive sample, while specificity parameters show how a model detects negative samples. AUC describes an efficient relationship for evaluating the performance of the proposed DLAC-CNN-RF classifier model. A higher AUC value illustrates a better performance from the model. It is noted that the silty loam and sand clay were classified 100% correctly, and the ACU value for all predictions is over 97.5%.

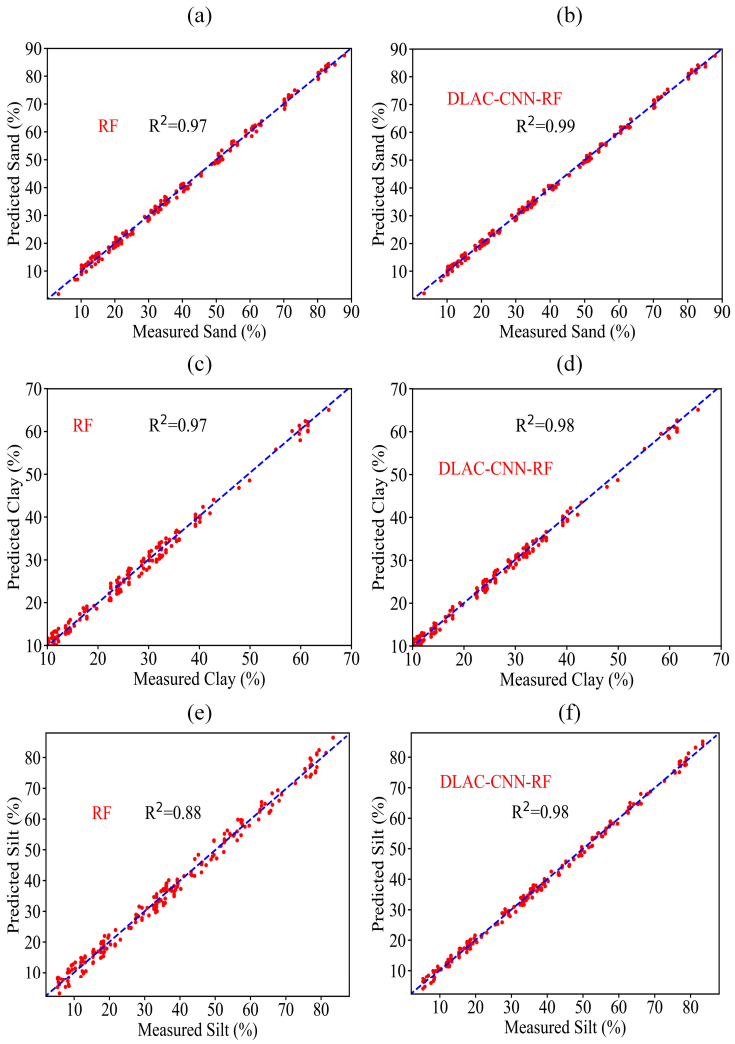


Figure 8. The measured and predicted plots for (**a**) sand, using full features (RF), (**b**) sand, using full features (DLAC-CNN-RF), (**c**) clay, using full features (RF), (**d**) clay, using full features (DLAC-CNN-RF), (**e**) silt, using full features (RF), and (**f**) silt, using full features (DLAC-CNN-RF).

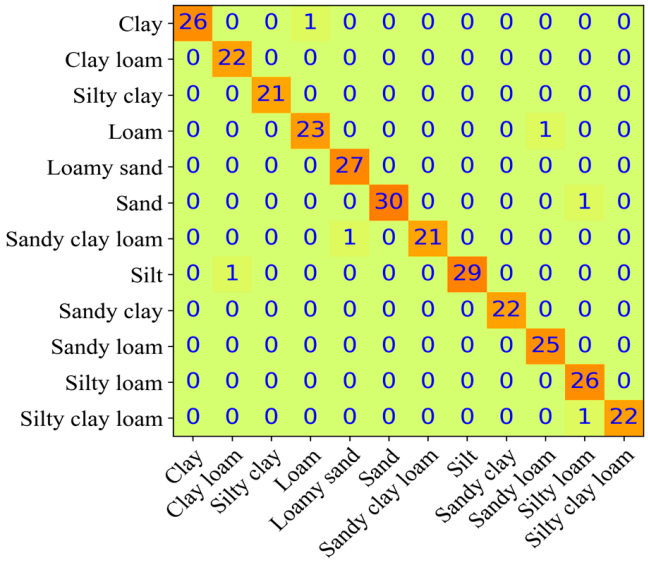


Figure 9. Predicted distributions of 12 soil textures classes via the proposed DLAC-CNN-RF model.

Table 2. Performance parameters of the proposed DLAC-CNN-RF model for classification and prediction using full features. The number of the clay, silty clay, silty clay loam, sandy clay, sandy clay loam, clay loam, silt, silty loam, loam, sand, loamy sand, and sandy loam samples used for testing were 26, 23, 21, 24, 28, 30, 21, 29, 22, 26, 28, and 22, respectively, and a total number of 300 samples was employed.

Soil Textures	Accuracy	Precision	Sensitivity	Specificity	AUC
Clay	99.67%	100%	96.3%	100%	98.15%
Clay loam	99.67%	95.65%	100%	99.64%	99.82%
Silty loam	100%	100%	100%	100%	100%
Loam	99.33%	95.65%	95.83%	99.64%	97.74%
Loamy sand	99.67%	96.43%	100%	99.63%	99.82%
Sand	99.67%	100%	96.77%	100%	98.39%
Sandy clay loam	99.67%	100%	95.45%	100%	97.73%
Silt	99.67%	100%	96.67%	100%	98.34%
Sandy clay	100%	100%	100%	100%	100%
Sandy loam	99.67%	96.15%	100%	99.64%	99.82%
Silty loam	99.33%	92.86%	100%	99.28%	99.64%
Silty clay loam	99.67%	100%	95.65%	100%	97.83%
average	99.67%	98.06%	98.06%	99.82%	98.94%

Figure 10 reports the loss function of the proposed DLAC-CNN-RF model for training and testing the soil images. An epoch represents one cycle of updating the weights through the complete training soil image dataset. The loss value shows how well our model reacts after each iteration of optimization. A downward trend in the loss curves was observed, indicating the proposed model performs well for soil image classification.

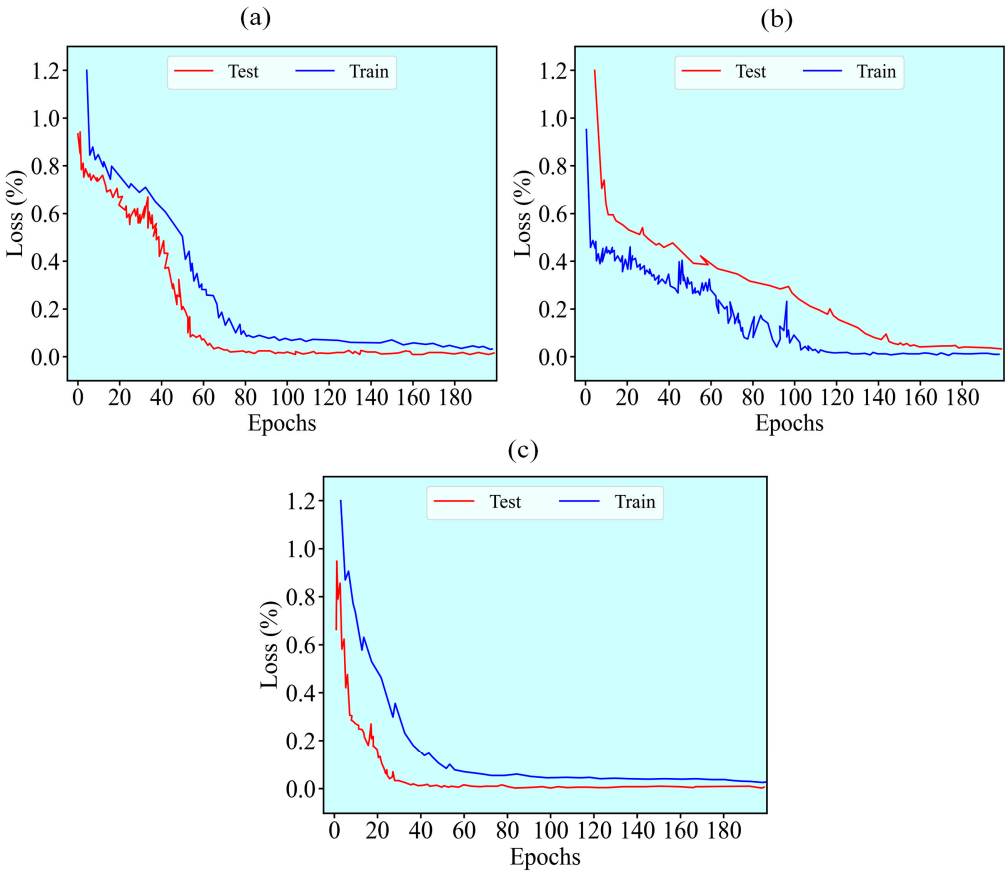


Figure 10. Classification accuracy and error rate using the proposed DLAC-CNN-RF model for (**a**) sand images, (**b**) clay images, and (**c**) silt images.

Table 3 shows the accuracy of the clay, sand, and silt soil image classifications of the proposed DLAC-CNN-RF model and other machine learning methods, including KNN and VGG16-RF. The full features of the prepared soil images were extracted in these models to evaluate their performance. As shown in Table 3, the conventional KNN model realizes an accuracy of 95% for clay image classification, which is relatively higher than the VGG16-RF models, but in terms of sand and silt classification, it is worse than the VGG16-RF model. In contrast, the proposed DLAC-CNN-RF model shows very good performance in classifying all the soil textures, exhibiting the highest R^2 and the lowest RMSE, which can be interpreted as the convolutional networks automatically and simultaneously extracting and selecting the features and reducing data over-fitting and complex computation.

Table 3. The comparison between the proposed DLAC-CNN-RF model and other models.

Model	Soil Types	Feature	Test	
			R^2	RMSE
KNN	Clay	Color + particle + texture	0.95	4.59
	Sand	Color + particle + texture	0.85	4.62
	Silt	Color + particle + texture	0.94	4.60
VGG16-RF	Clay	Color + particle + texture	0.85	4.23
	Sand	Color + particle + texture	0.93	3.85
	Silt	Color + particle + texture	0.97	3.95
Proposed DLAC-CNN-RF model	Clay	Color + particle + texture	0.99	3.76
	Sand	Color + particle + texture	0.99	3.71
	Silt	Color + particle + texture	0.98	3.79

4. Conclusions

This study demonstrated a cheap and environmentally friendly image acquisition system consisting of a smartphone, a customized chamber, and a mobile application for predicting soil images. The particle (threshold binarization), color (HSV (hue, saturation, and value) and Hu moments), and texture features (local binary patterns (LBP) and Harallick features) were extracted and used in random combinations to predict clay, silt, and sand content via RF and DLAC-CNN-RF algorithms. The results indicated that the proposed DLAC-CNN-RF model has better performance. Particularly, when the full features were extracted, an average accuracy of 99.67% was obtained when predicting all the soil textures. A user-friendly interface based on the calibrated DLAC-CNN-RF model has been designed, which clearly presents the prediction results. When compared with other commonly used models, the proposed DLAC-CNN-RF model is a promising solution that benefits from rapid and low-cost soil identification and classification. Our research may have two limitations. The first is that the used algorithm cannot predict soil moisture and organic carbon content. The second is the distance between the camera and the sample, which cannot be moved freely within the image acquisition system. Future research will study the effect of soil moisture and imaging distance on the prediction performance of the proposed model. Over the past few years, an unmanned aerial system (UAS)-based soil image acquisition method has received much attention due to its simple and fast implementation and the ability to take images with multiple elevation points in remote areas. Many UAS-based soil texture identification works have been reported [30–32]. Therefore, it is also worth exploring the classification and prediction performance of the proposed model by imputing images taken by the drone.

Author Contributions: Formal analysis and writing original draft, Z.Z.; data curation and validation, W.F.; investigation and visualization, J.X.; review and editing, S.P.; supervision, X.L.; conceptualization, project administration, and funding acquisition, Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The National Natural Science Foundation of China (51975136, 52075109), the Science and Technology Innovative Research Team Program in Higher Educational Universities of Guangdong Province (2017KCXTD025), Special Research Projects in the Key Fields of Guangdong Higher Educational Universities (2019KZDZX1009), Natural Science Foundation of Guangdong Province (2022A010102014), the Tertiary Education Scientific research project of Guangzhou Municipal Education Bureau (202235139), and Guangzhou University Research Project (YJ2021002).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to business restrictions.

Acknowledgments: We thank for the support from the Guangdong Engineering Research Centre for Highly Efficient Utility of Water/Fertilizers and Solar-Energy Intelligent Irrigation.

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

References

1. Phogat, V.K.; Tomar, V.S.; Dahiya, R. Soil physical properties. *Soil Sci. Introd.* **2015**, 135–171.
2. Rahimi-Ajdadi, F.; Abbaspour-Gilandeh, Y.; Mollazade, K.; Hasanzadeh, R.P.R. Development of a novel machine vision procedure for rapid and non-contact measurement of soil moisture content. *Measurement* **2018**, *121*, 179–189. [CrossRef]
3. Klute, A. *Methods of Soil Analysis. Part 1. Physical and Mineralogical Methods*; Sssa book series 5; Soil Science Society of America: Madison, WI, USA, 1986.
4. Robinson, G.W. A new method for the mechanical analysis of soils and other dispersions. *J. Agric. Sci.* **1922**, *12*, 306–321. [CrossRef]
5. Di Stefano, C.; Ferro, V.; Mirabile, S. Comparison between grain-size analyses using laser diffraction and sedimentation methods. *Biosyst. Eng.* **2010**, *106*, 205–215. [CrossRef]
6. Chakraborty, S.; Weindorf, D.C.; Deb, S.; Li, B.; Paul, S.; Choudhury, A.; Ray, D.P. Rapid assessment of regional soil arsenic pollution risk via diffuse reflectance spectroscopy. *Geoderma* **2017**, *289*, 72–81. [CrossRef]
7. Fu, Y.; Taneja, P.; Lin, S.; Ji, W.; Adamchuk, V.; Daggupati, P.; Biswas, A. Predicting soil organic matter from cellular phone images under varying soil moisture. *Geoderma* **2020**, *361*, 114020. [CrossRef]
8. Andrenelli, M.C.; Fiori, V.; Pellegrini, S. Soil particle-size analysis up to 250 μm by x-ray granulometer: Device set-up and regressions for data conversion into pipette-equivalent values. *Geoderma* **2013**, *192*, 380–393. [CrossRef]
9. Fisher, P.; Aumann, C.; Chia, K.; O'Halloran, N.; Chandra, S. Adequacy of laser diffraction for soil particle size analysis. *PLoS ONE* **2017**, *12*, e0176510. [CrossRef]
10. Jaconi, A.; Vos, C.; Don, A. Near infrared spectroscopy as an easy and precise method to estimate soil texture. *Geoderma* **2019**, *337*, 906–913. [CrossRef]
11. Vaz, C.M.P.; de Mendonça Naime, J.; Macedo, Á. Soil particle size fractions determined by gamma-ray attenuation. *Soil Sci.* **1999**, *164*, 403–410. [CrossRef]
12. Vohland, M.; Ludwig, M.; Thiele-Bruhn, S.; Ludwig, B. Determination of soil properties with visible to near-and mid-infrared spectroscopy: Effects of spectral variable selection. *Geoderma* **2014**, *223–225*, 88–96. [CrossRef]
13. El Hourani, M.; Broll, G. Soil protection in floodplains—A review. *Land* **2021**, *10*, 149. [CrossRef]
14. Sofou, A.; Evangelopoulos, G.; Maragos, P. Soil image segmentation and texture analysis: A computer vision approach. *IEEE Geosci. Remote Sens. Lett.* **2005**, *2*, 394–398. [CrossRef]
15. Sudarsan, B.; Ji, W.; Adamchuk, V.; Biswas, A. Characterizing soil particle sizes using wavelet analysis of microscope images. *Comput. Electron. Agric.* **2018**, *148*, 217–225. [CrossRef]
16. Aitkenhead, M.; Coull, M.; Gwatkin, R.; Donnelly, D. Automated soil physical parameter assessment using smartphone and digital camera imagery. *J. Imaging* **2016**, *2*, 35. [CrossRef]
17. Aitkenhead, M.; Cameron, C.; Gaskin, G.; Choisy, B.; Coull, M.; Black, H. Digital rgb photography and visible-range spectroscopy for soil composition analysis. *Geoderma* **2018**, *313*, 265–275. [CrossRef]
18. de Oliveira Morais, P.A.; de Souza, D.M.; de Melo Carvalho, M.T.; Madari, B.E.; de Oliveira, A.E. Predicting soil texture using image analysis. *Microchem. J.* **2019**, *146*, 455–463. [CrossRef]
19. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
20. Dornik, A.; DrĂGuŢ, L.; Urdea, P. Classification of soil types using geographic object-based image analysis and random forests. *Pedosphere* **2018**, *28*, 913–925. [CrossRef]
21. Fan, R.; Bocus, M.J.; Zhu, Y.; Jiao, J.; Wang, L.; Ma, F.; Cheng, S.; Liu, M. Road crack detection using deep convolutional neural network and adaptive thresholding. In Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019.
22. Vardhana, M.; Arunkumar, N.; Lasrado, S.; Abdulhay, E.; Ramirez-Gonzalez, G. Convolutional neural network for bio-medical image segmentation with hardware acceleration. *Cogn. Syst. Res.* **2018**, *50*, 10–14. [CrossRef]
23. Swetha, R.K.; Bende, P.; Singh, K.; Gorthi, S.; Biswas, A.; Li, B.; Weindorf, D.C.; Chakraborty, S. Predicting soil texture from smartphone-captured digital images and an application. *Geoderma* **2020**, *376*, 114562. [CrossRef]

24. Azadnia, R.; Jahanbakhshi, A.; Rashidi, S. Developing an automated monitoring system for fast and accurate prediction of soil texture using an image-based deep learning network and machine vision system. *Measurement* **2022**, *190*, 110669. [CrossRef]
25. He, R.; Dai, Y.; Lu, J.; Mou, C. Developing ladder network for intelligent evaluation system: Case of remaining useful life prediction for centrifugal pumps. *Reliab. Eng. Syst. Saf.* **2018**, *180*, 385–393. [CrossRef]
26. Gee, G.W.; Or, D. 2.4 particle-size analysis. *Methods Soil Anal. Part 4 Phys. Methods* **2002**, *5*, 255–293.
27. Soil Survey Staff. *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys*; USDA, Natural Resources Conservation Service: Washington, DC, USA, 1999.
28. Qi, L.; Adamchuk, V.; Huang, H.-H.; Leclerc, M.; Jiang, Y.; Biswas, A. Proximal sensing of soil particle sizes using a microscope-based sensor and bag of visual words model. *Geoderma* **2019**, *351*, 144–152. [CrossRef]
29. Minasny, B.; McBratney, A.B.; Tranter, G.; Murphy, B.W. Using soil knowledge for the evaluation of mid-infrared diffuse reflectance spectroscopy for predicting soil physical and mechanical properties. *Eur. J. Soil Sci.* **2008**, *59*, 960–971. [CrossRef]
30. Aboutalebi, M.; Allen, L.N.; Torres-Rua, A.F.; McKee, M.; Coopmans, C. *Estimation of Soil Moisture at Different Soil Levels Using Machine Learning Techniques and Unmanned Aerial Vehicle (UAV) Multispectral Imagery*; SPIE: Bellingham, WA, USA, 2019; pp. 216–226.
31. Marcu, I.; Suciu, G.; Bălăceanu, C.; Vulpe, A.; Drăgulinescu, A.-M. Arrowhead technology for digitalization and automation solution: Smart cities and smart agriculture. *Sensors* **2020**, *20*, 1464. [CrossRef]
32. Maimaitijiang, M.; Sagan, V.; Sidike, P.; Hartling, S.; Esposito, F.; Fritschi, F.B. Soybean yield prediction from uav using multimodal data fusion and deep learning. *Remote Sens. Environ.* **2020**, *237*, 111599. [CrossRef]

agronomy

MDPI

Communication

Smart Automation for Production of Panchagavya Natural Fertilizer

Sumathi V. [1,*] and Mohamed Abdullah J. [2]

[1] Centre for Automation, School of Electrical Engineering, VIT, Chennai 600127, Tamil Nadu, India
[2] School of Electrical Engineering, VIT, Chennai 600127, Tamil Nadu, India
* Correspondence: vsumathi@vit.ac.in

Abstract: Modern agricultural farming techniques employ the usage of chemical supplements to improve crop yield in terms of quantity and quality. This practice has brought down the fertility of the soil and has led to secondary impacts and necessitates a significant financial investment. Awareness of the side effects of artificially enriched food has made people move towards organically grown food, and the consumption has also increased significantly. One of the ancient organic fertilizers used in India is panchagavya. As the name implies, pancha means five and gavya means cow. The five products of the cow are combined as per the compositions and procedure described in the literature, to provide economical and meaningful value to organic farming. The objective of this work is to design, develop, and implement an automated system to manufacture panchagavya with reduced operator assistance. The system implements an ATmega 328 microcontroller to automate the entire process by interfacing sensors such as pH, moisture, temperature, and pressure. The system is also provided with a SIM900A GSM modem to provide information to the user regarding the status of the process. The developed pilot scale design discussed in this work has several advantages in the world of farming technologies in terms of enriching the soil, thereby improving the crop yield. This technology will benefit the farmers as this natural fertilizer can be mass-produced and turn them into entrepreneurs, which benefits society at large.

Keywords: panchagavya; organic fertilizer; liquid fertilizer; automated fertilizer production; drip irrigation system; automated irrigation

Citation: V., S.; J., M.A. Smart Automation for Production of Panchagavya Natural Fertilizer. *Agronomy* **2022**, *12*, 3044. https://doi.org/10.3390/agronomy12123044

Academic Editors: Jian Zhang, Randy G. Goebel, Zhihai Wu and Pablo Martín-Ramos

Received: 7 October 2022
Accepted: 24 November 2022
Published: 1 December 2022

1. Introduction

The use of chemicals in farming improves the crop yield, but this has led to secondary effects as well, such as drastic reduction in the fertility of the soil and health-related issues in humans. These factors have presently led to adapting the age-old organic farming system. An ideal organic farming system enhances the process of nutrient cycling and reduces the usage of external inputs [1]. The most widely practiced organic farming system in ancient India uses panchagavya, vermicompost, and farmyard manure for nutrient management and soil enrichment. Panchagavya, derived from ancient Indian times, has always been treasured as a valuable possession of the country. It has been used for generations in enriching and improving the soil, and provides health benefits for both humans and animals. It has proved to be successful within a short duration. Among several benefits of this natural fertilizer, the agricultural era has benefited the most. The five cow products use three dairy products: cow milk, curd, and ghee, which is mixed with cow dung and urine [2]. The usage of chemicals adopted during the industrial revolution has led to the diminishing use of the natural fertilizer panchagavya. A study carried out in Andhra Pradesh, India [3], reveals the benefits of panchagavya as a fertilizer and also a seed storage treatment entity. It also characterizes three different organic preparations concerning microbiological aspects and the impacts made on crop growth and yield given in Figure 1.

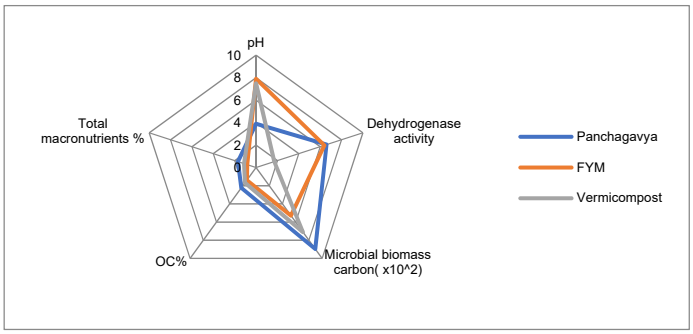


Figure 1. Microbiological quality and impacts of different organic fertilizers [3].

The use of panchagavya in a field experiment conducted on brinjal (*Solanum melongena*) in 2008–2009 by the University of Agricultural Sciences, Dharwad (Karnataka), India, has exhibited good results in brinjal yield. The nutrient uptake has been reported to be N (92.86 kg ha^{-1}), P (22.16 kg ha^{-1}), K (110.62 kg ha^{-1}), and S (29.48 kg ha^{-1}) [4]. Apart from several advantages related to plantation and crop yield, the most specific application is in the use of seeds or seedling treatment and can be used as a catalyst for organic manure for speeding up its decomposition [5]. Experienced personnel use panchagavya as a medical supplement to treat people for various diseases and deficiencies. Though the different applications listed require a 3% concentration of panchagavya, some plants, such as cashew rootstocks, require a 5% concentration for beneficial growth parameters and graft success [6]. Effects of the application of panchagavya in the form of seed treatment and foliar spray to southern sunn-hemp mosaic virus-infected sun hemp plants have been reported [7]. Since panchagavya not only holds itself useful in the field of agriculture but also plays a vital role in the health of humans and animals, researchers can bring out the possible benefits of panchagavya as a medicine [8–10].

The conventional method in practice for the preparation of panchagavya involves the mixing of cow dung and ghee thoroughly in a mud pot (or any other vessel except metal containers as panchagavya is acidic). This mixture is stirred every 12 h for 3 days. Once the process is completed, the remaining three ingredients are added, and again, the mixture is stirred thoroughly. At this stage, any one of the catalysts may be added, such as sugarcane juice and jaggery for speeding up the process. This mixture is stirred every 12 h for 15 days, and it is ready for use after the 18th day. The entire process is carried out manually, which requires physical effort and tolerance of uncomfortable smell and timekeeping. The ingredient proportions mentioned in Table 1 yield approximately 20 L of panchagavya, as reported in the literature.

Table 1. Composition of panchagavya.

S. No.	Ingredients Required for the Preparation of Panchagavya	Composition for Making 20 L of Panchagavya
1	Cow dung	5 kg
2	Cow urine	3 L
3	Cow milk	2 L
4	Cow ghee (clarified butter)	500 g
5	Curd	2 L

This paper proposes the implementation of a new method for modern organic farming techniques, bringing sustainable changes in the field of agriculture technology by developing a cost-effective design for the production of panchagavya with minimal human effort. An attempt is made to design, develop, and automate a microcontroller-based system in pilot scale to overcome the disadvantages of the manual method in practice. The proposed work, when implemented in the irrigation field, uses a fertigation technique, that is, mixing of liquid fertilizer with irrigation water [11] and with more focus on applying in the surface

irrigation method [12]. The first section of the article discusses the methodology involved in the design and development of the natural fertilizer panchagavya's preparation. The scaled-down values for pilot-scale implementation is also discussed. The next section highlights the working of the developed system, followed by the design implementation of the experimental setup. Finally, the results and conclusion of this research are discussed.

2. Methodology

The pilot design shown in Figure 2 for the preparation of panchagavya consists of three stages, which yields 1 L of panchagavya.

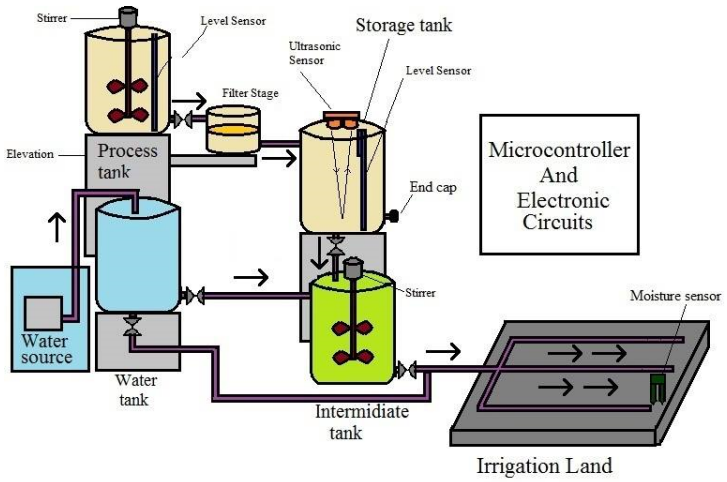


Figure 2. Block diagram of the designed system.

The block diagram shown above consists of the process tank, storage tank, water tank, and intermediate tank. The process tank is attached to a feeder, blender assembly, level sensor, and GSM module. The blender assembly is mounted vertically on top of the process tank. The assembly consists of a DC motor coupled with an elongated shaft and blades for rotary action. The solenoid valves used in this system have a larger outlet orifice compared with other valves used since unfiltered panchagavya passes through this valve. The end-to-user communication is carried out using the GSM module. The storage tank is used for storing the liquid fertilizer. Since the panchagavya is acidic in nature and cannot be directly irrigated into the field, the liquid fertilizer is mixed with water from the water tank and stored in the intermediate tank. The Atmel ATmega 328 microcontroller is integrated into the system to monitor the physical parameters of the soil, such as moisture and pH.

3. Working of the Desired System

The ingredients used for the preparation of 1 L of panchagavya have been scaled from Table 1 and are prescribed in Table 2 and placed in the process tank. The process tank's inner layer is made of a nonmetallic body, and the input is loaded through a hopper feeder placed above the top lid of the tank and is sealed. The pH and the temperature values are displayed as sensed by the pH sensor and temperature sensor inside the process tank. A horizontally placed motor on top of the process tank acts as the blender to mix the composition evenly at regular intervals. Continuous mixing of the composition at regular intervals aids the culturing process and avoids sedimentation. Once the process ends, the prepared mixture is processed, the microprocessor sends a command to the outlet solenoid valve placed in the process tank to open, and the liquid flows gradually to the filter tank. In the filter tank, using multiple layers of filter, the mixture is filtered and the liquid fertilizer is separated from the residue and passed into storage tank. The storage tank is monitored by using a pH sensor and level sensor given in Table 3. The level of the liquid fertilizer in the storage tank is measured. At every stage of the process, the level in each of the tanks

is monitored by the level sensor placed in the tanks interfaced with the microcontroller. The available panchagavya in the storage tank has to be mixed appropriately with water. The water storage system for irrigation is controlled by the microcontroller by turning ON and OFF the pump using the relay. The suggested ratio of mixing panchagavya with water is 30:70, which is stored in the intermediate tank for irrigation. The sensors interfaced to every stage of the process is controlled by the Atmel ATmega 328 microcontroller.

Table 2. Scaled-down ingredients.

S. No.	Ingredients	Quantity
1	Fresh cow dung	500 g
2	Cow urine	300 mL
3	Cow milk	200 mL
4	Cow curd	200 mL
5	Cow ghee	50 g
6	Ripe banana	1 no

Table 3. Ultrasonic-level sensor readings.

Level of Panchagavyam	Level Sensor Readings (in cm)
Empty	Above 16
Low	13–16
Medium	10–12
High	7–9

4. Experimental Setup of the Implemented Design

The liquid fertilizer ready to be fed into the field cannot be transferred as such. A pilot study was conducted to farmers from a small village in Kandigai, Chennai, Tamil Nadu, India, on the time constraints followed in irrigating the field when drip irrigation is adopted. Based on the inputs received from them, it is ideal to supply water for 4 h using the drip irrigation method to irrigate the field. The work focuses on transferring the liquid fertilizer to the soil by drip irrigation, and thus, the 4 hours of supplying water to the field by drip irrigation is divided into four slots. As per these proportionated time slots, an algorithm for drip irrigation is designed. In the first slot, only water is fed to the crops. After an hour, the value of the pH sensor readings is considered to supply the liquid fertilizer; if the pH value is within the nominal range, the microprocessor turns ON the intermediate tank valve to open. The reason behind not feeding the panchagavya-mixed water directly to the first time slot is to avoid the supplied enriched water from directly being absorbed by the soil, and the crops will be left with very few nutrients. Two slots are used to irrigate the land with the liquid fertilizer. Finally, when the countdown timer is left with the remaining 1 hour, the intermediate tank valve is closed and water is fed to the irrigation system. The algorithm for drip irrigation is shown in Figure 3.

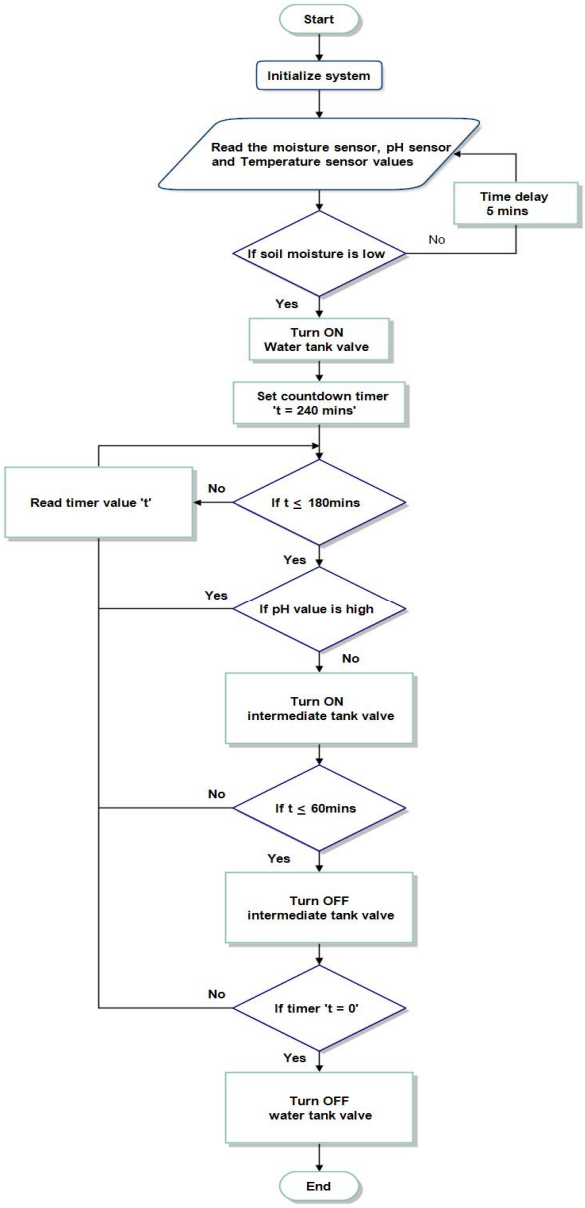


Figure 3. Flowchart of the automated drip irrigation process.

5. Results and Discussion

The proposed work in pilot scale is implemented, and the setup is shown in Figure 4. This model is designed to produce 1 L of panchagavya. For the requirement of a nonmetallic tank surface, all the tanks used in this model are chosen to be plastic-based containers. The introduction of preparation and irrigation modules has led to the inclusion of two microcontrollers. Two Atmel ATmega 328 microcontrollers are used, one as master and the other as slave. Each microcontroller carries 14 digital input/output pins and 6 analog inputs, which support the signals from level sensors, pH sensors, moisture sensors, and ultrasonic sensors and support communication through the GSM module given in Figure 4a.

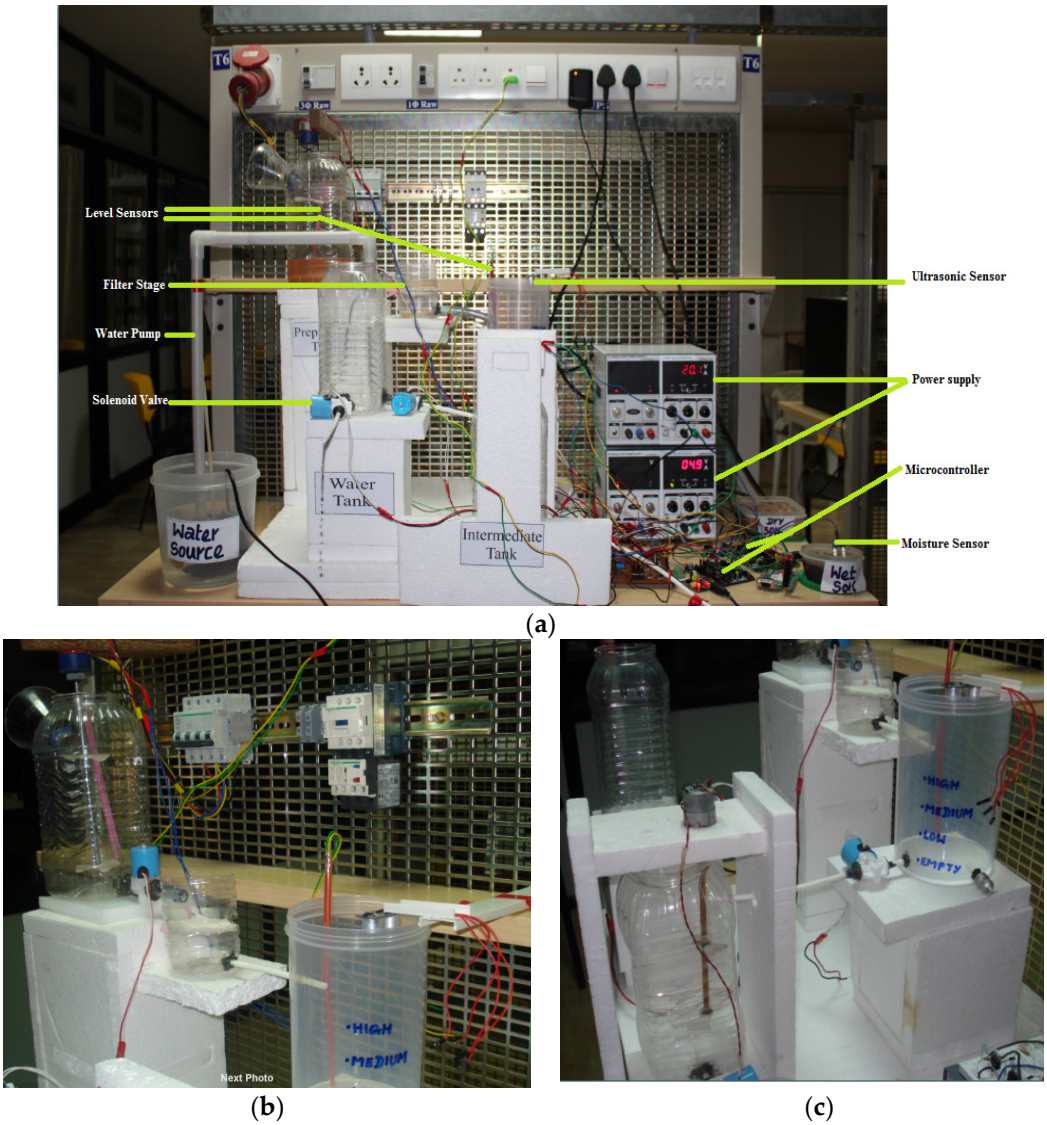


Figure 4. (**a**) Hardware setup of the proposed pilot scale model, (**b**) implementation of the preparation module, and (**c**) implementation of the irrigation module.

Since the production is planned for a small scale, a DC motor with high torque and low rpm is used for blender operation. A driver circuit is included to operate the DC motor interfaced to the microcontroller. Solenoid valves used in various levels interconnect different stages of the tank, and the HR1-type solenoid valve of a different size is used in this model. A conductance level sensor is used in both the preparation tank and the storage tank, which prevents overflowing given in Figure 4b,c. The working of this sensor is similar to the float-type sensor to indicate the brim level. An HC-SR04 model ultrasonic sensor is used in the storage tank to measure the level of the liquid fertilizer. A SIM900A module GSM modem is connected to the preparation to serve the purpose of communicating between the processing end and user end. The start and end of the blender operation and irrigation process at regular intervals are notified to the user by a text message and sent through the SMS service.

The model has been successfully implemented for the automated preparation of panchagavya with a surface drip irrigation system given in Figure 5a, and the installation of a moisture sensor is given in Figure 5b. Moisture sensor reading in voltage is shown in Figure 6. From this study, we understand that the positioning of a moisture sensor and pH sensor on the surface of the soil is to be wisely chosen, depending upon the crop's

requirement, and it is proven from [13] that it is ideal to measure under 10–15 cm from the soil surface. Depending upon the different crop needs, the type of irrigation with a liquid fertilizer is recommended as surface, subsurface, and root infiltration, as suggested in [12,14,15]. The drip irrigation systems are usually powered by a gravity tank for the flow of water to reach each node of the drip system. If the field to be covered is large, then the ordinary tank setup would not support the extent; hence, an intermediate pump would be required to cover the field. The sensors should be placed widely in the field to ensure that the region is properly covered for optimal utilization.

(**a**) (**b**)

Figure 5. (**a**) Soil moisture sensor and pH sensor installed and tested; (**b**) soil moisture sensor placed in panchagavya.

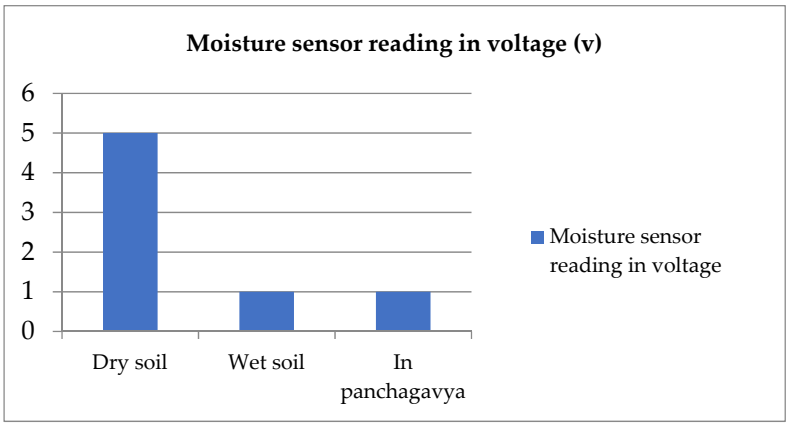


Figure 6. Moisture sensor reading in voltage.

6. Conclusions

The prototype of the ATmega 328 microcontroller-based automatic preparation of panchagavya natural fertilizer supported with drip irrigation is developed in pilot scale, and the results are reported. It is clear from the results that the panchagavya prepared by the automatic method has more advantage than the conventional method. In the preparation stage, the manual method requires much human effort, very tedious preparation, foul smell, and no information on the physical parameters of the liquid fertilizer; that is, moisture, pH, and so on are all overcome in the present development. The interfacing of pH sensors, moisture sensor, ultrasonic sensor, and GSM module gives immediate update on the process as the user is notified at every stage and, thus, eliminates the need for humans to be tied to the process. The developed prototype is low cost and, if implemented in a large scale, can be upgraded easily with minor modifications. The research concludes that if this method is adopted, the use of the diminishing natural fertilizer panchagavya can be brought to the limelight, which is much beneficial to the agricultural sector, improving more yield economically. Despite the several health benefits being reported in the literature, the farmers at large are benefitted, turning them into entrepreneurs and improving their livelihood.

Funding: The authors wish to thank VIT management for providing sufficient funds towards publication of the work.

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

References

1. Natrajan, K. *Book on "Panchagavya", 2nd ed*; Distributed by Other India Bookstore; Organic Farming Association of India: Mapusa, Goa, India, 2008.
2. Math, K.K.; Babalad, H.; Jyothi, T.V.; Bhat, S.N. Growth, yield and nutrient uptake of brinjal (*Solanum melongena* Linn.) as influenced by organic nutrient management practices. *Plant Arch.* **2013**, *13*, 363–367.
3. Amalraj, L.D.; Kumar, G.P.; Hassan, S.K.; Shaik, R. Microbiological analysis of Panchagavya, vermicompost, and FYM and their effect on plant growth promotion of pigeon pea (*Cajanus cajan* L.). *Org. Agric.* **2013**, *3*, 23–29. [CrossRef]
4. Kumar, K.; Verma, G.; Veer, R.; Kumar, S.; Kumar, P. Exploitation of Panchagavya, benefits and eco-friendly management of plant diseases: A review. *J. Entomol. Zool. Stud.* **2020**, *8*, 2360–2364.
5. Thimmaiah, A. Training Manual on Low cost Organic Agriculture. In *Nutrient Management in Organic Agriculture*; SNV: Thimpu, Bhutan, 2010; Volume 3, pp. 36–38.
6. Singh, L.S.; Pariari, A. Effect of foliar application of plant bio-regulators (PBRs) on growth and graft success of cashew rootstocks. *Res. Crops* **2013**, *14*, 530–534.
7. Vallimayil, J.; Sekar, R. Investigation on the Effect of Panchagavya on Sounthern Sunnhemp Mosaic Virus (SSMV) Infected Plant Systems Department of Botany with specialization in Plant biotechnology. *Glob. J. Environ. Res.* **2012**, *6*, 75–79.
8. Dhama, K.; Chakraborty, S.; Wani, M.Y.; Verma, A.K.; Deb, R.; Tiwari, R.; Kapoor, S. Novel and emerging therapies safeguarding health of humans and their companion animals: A review. *Pak. J. Biol. Sci. PJBS* **2013**, *16*, 101–111. [CrossRef] [PubMed]
9. Miller, E.H. A note on reflector arrays. *IEEE Trans. Antennas Propagat.* **2012**, *15*, 692–693.
10. Sachithanandam, P.; Muniyandi, K. All About Panchagavya for Human usage—A Review. *Indian J. Nat. Sci.* **2021**, *11*, 29173–29181.
11. Bajaj, K.K.; Chavhan, V.; Raut, N.A.; Gurav, S. Panchgavya: A precious gift to humankind. *J. Ayurveda Integr. Med.* **2022**, *13*, 100525. [CrossRef] [PubMed]
12. Rubio-Asensio, J.S.; Abbatantuono, F.; Miguel, J.; Hortelano, D.; Ru, L.; Parra, M.; Mar, R.; Intrigliolo, D.S.; Buesa, I. Effects of Cover Crops and Drip Fertigation Regime in a Young Almond Agroecosystem. *Agronomy* **2022**, *12*, 2606. [CrossRef]
13. Zhang, Y.; Liu, H.; Gong, P.; He, X.; Wang, J.; Wang, Z.; Zhang, J. Irrigation Method and Volume for Korla Fragrant Pear: Impact on Soil Water and Salinity, Yield, and Fruit Quality. *Agronomy* **2022**, *12*, 1980. [CrossRef]
14. Holatko, J.; Hammerschmiedt, T.; Kintl, A.; Kucerik, J.; Malicek, O.; Latal, O.; Baltazar, T.; Brtnicky, M. Effects of Strip-Till and Simultaneous Fertilization at Three Soil Depths on Soil Biochemical and Biological Properties. *Agronomy* **2022**, *12*, 2597. [CrossRef]
15. An, S.; Yang, F.; Yang, Y.; Huang, Y.; Zhangzhong, L.; Wei, X.; Yu, J. Water Demand Pattern and Irrigation Decision-Making Support Model for Drip-Irrigated Tomato Crop in a Solar Greenhouse. *Agronomy* **2022**, *12*, 1668. [CrossRef]

agronomy

MDPI

Article

A Grading Method of Ginseng (*Panax ginseng C. A. Meyer*) Appearance Quality Based on an Improved ResNet50 Model

Dongming Li [1], Xinru Piao [1], Yu Lei [1], Wei Li [2], Lijuan Zhang [3] and Li Ma [1,*]

[1] College of Information Technology, Jilin Agricultural University, Changchun 130118, China
[2] College of Chinese Medicinal Materials, Jilin Agricultural University, Changchun 130118, China
[3] College of Computer Science and Engineering, Changchun University of Technology, Changchun 130012, China
* Correspondence: mali@jlau.edu.cn; Tel.: +86-13-5008-83257

Abstract: In the academic world, ginseng (*Panax ginseng C. A. Meyer*) has received much attention as the most representative element of Chinese medicine. To address the lack of traditional algorithms in the identification of ginseng appearance quality and further improve the manual identification on ginseng, we propose a grading method of ginseng appearance quality based on deep learning, taking advantage of the benefits of deep learning in the image identification. Firstly, we substituted *Leaky ReLU* for the conventional activation function *ReLU* to enhance the predictive power of the model. Secondly, we added an ECA module to the residual block, which allowed attention to be focused on the input object to capture more precise and detailed features. Thirdly, we used the focal loss function to solve the problem of an imbalanced dataset. Then, the self-constructed dataset was processed with data enhancement and divided into four different classes of ginseng. The dataset was trained on a model with transfer learning to finally obtain the best model applicable to the identification of ginseng appearance quality. The experiments showed that, compared with the classical convolutional neural network models VGG16, GoogLeNet, ResNet50 and Densenet121, the proposed model reported the best performance, its accuracy in the test set was as high as 97.39%, and the loss value was 0.035. This method can efficiently classify the appearance quality of ginseng, and has a significant value in the field of ginseng appearance quality identification.

Keywords: appearance quality identification of ginseng; deep learning; attention mechanism; activation function; loss function

Citation: Li, D.; Piao, X.; Lei, Y.; Li, W.; Zhang, L.; Ma, L. A Grading Method of Ginseng (*Panax ginseng C. A. Meyer*) Appearance Quality Based on an Improved ResNet50 Model. *Agronomy* **2022**, *12*, 2925. https://doi.org/10.3390/agronomy12122925

Academic Editors: Jian Zhang, Randy G. Goebel and Zhihai Wu

Received: 18 October 2022
Accepted: 16 November 2022
Published: 23 November 2022

1. Introduction

Ginseng (*Panax ginseng C. A. Meyer*) is the dried root and rhizome of ginseng of the family Wujia, which has the effects of tonifying the kidneys, calming the mind, nourishing the brain, brightening the eyes, and beneficial to intellectual development [1]. It is referred to as the "king of all herbs" since it is a priceless Chinese herb and tonic [2]. The appearance of particular ginseng features is the key factor used in the traditional method of ginseng identification to assess its quality. The literature [3] also demonstrates that there is a correlation between the appearance of ginseng's characteristic traits and its main chemical components. Empirical identification of ginseng quality based on "identifying the appearance and shape to discuss the quality" is a scientific method [4–6] that is particularly effective. However, it has shortcomings. To identify a ginseng specimen, a connoisseur must first carefully observe all aspects of the ginseng plant under exam, which requires a lot of time and effort. Secondly, ginseng with high economic and therapeutic value may be damaged, since the manual identification process has a tendency to harm the plant. In the end, manual recognition is subject to some degree of subjective influence. This frequently leads to inconsistent identification criteria and incorrect identification outcomes [7]. At present, there is a lack of sufficient attention to the study of ginseng characteristics. The 2020 edition of the Chinese Pharmacopoeia also includes descriptions and specificities of ginseng phenotypic features;

however, in practice, the general public will still not be able to tell ginseng quality by its appearance. With the rapid advancement of computer technology, artificial intelligence technology has slowly begun to be integrated in traditional Chinese medicine identification. The "Fourteenth Five-Year Plan for the Development of Chinese Medicine Informatization" announced in 2022 makes it very obvious that informational means must be used to advance the modernization of Chinese medicine.

Deep learning is close to the way humans learn by mimicking the neural networks of the human brain, using a hierarchical network model structure to gather information about the appearance and sound of things, to perceive and understand them, and to generate appropriate behaviors. Its development in the field of computer vision is more complete than in other areas and has produced some outcomes [8]. LECUN et al. [9] proposed LeNet, which is the most representative convolutional neural network and consists of a Conv layer, a pooling layer, and a fully connected layer. When the network was initially used to accurately classify handwritten datasets, it established the stage for the later development of convolutional neural networks. With the AlexNet [10] network winning the ImageNet (large visualization database) competition in 2012, a number of classical deep convolutional neural networks have emerged, such as VGG [11], GoogLeNet [12], ResNet [13], DenseNet [14], etc. They outperform conventional approaches in picture identification and offer greater advantages. These networks are already extensively utilized in the fields of Chinese medicine and plant identification, despite the fact that they have not yet been employed to identify the appearance and quality of ginseng. Dyrmann et al. [15] adopted a method for identifying plant species in color images using convolutional neural networks and were able to identify 22 weed and crop species with an accuracy of 86.2%. Lee et al. [16] designed a deep learning method for the quantitative discrimination of leaves by gathering data on leaf features and analyzing them through convolutional neural networks and deconvolutional networks. The results of the study demonstrated that using deep learning methods can further increase the identification efficiency of plant classification based on the leaf. Liu et al. [17] applied the Inception structure and introduced dense connectivity to successfully achieve the classification and identification of six grape diseases, and the final model accuracy was 97.22%. Liu Wei et al. [18] combined Xception and DenseNet to propose the new image identification model DxFusion, which was able to accurately identify 60 Chinese herbs. Li Dongming et al. [19] used a deep learning strategy with the residual network and densely linked network to eventually identify and categorize five distinct origins of the Saposhnikovia divaricate.

Ginseng appearance quality identification is a fine-grained classification task. Therefore, considering the above research analysis, we attempted to apply deep learning techniques to the field of ginseng appearance quality identification and enhance ResNet50 according to the characteristics of ginseng. We finally propose an algorithm that can be effective, quick, and precise for grading the appearance and quality of ginseng.

2. Data Pre-Processing

2.1. Dataset Construction

After analyzing various ginseng varieties, we chose to employ white ginseng (a product made from fresh garden ginseng that was cleaned and dried or dried in the sun) as the experimental subject due to the complex morphological structure of the herb and market demand. The experimental data were gathered in September 2021 at the School of Chinese Herbal Medicine, Jilin Agricultural University, from the same batch of white ginseng. To facilitate the image collection, the white ginseng batch was assigned a serial number and scored by experts in accordance with the guidelines for white ginseng in the document "Group Standard for Ginseng of Jilin Daoji Herbs" which was released by the Tonghua Ginseng Industry Association on 10 September 2021 (as in Table 1). Ginseng of a standard lower than the principal, first-class, and second-class was rated as inferior.

The specimens were then photographed according to their ratings by using a small HD folding studio box (Sutefoto., Guangdong, China). A mobile phone camera (Apple.,

Cupertino, CA, USA) was positioned at the top of the studio box, perpendicular to the ginseng, at a height of 40 cm, guaranteeing that all samples were essentially photographed in the same position. The photographs of the white ginseng were gathered from various angles and backgrounds, and each image had a resolution of 1024 × 1024 pixels, with distinct image details. A total of 549 images of the principal white ginsengs, 790 images of the first-class white ginsengs, 600 images of the second-class white ginsengs, and 290 images of the inferior white ginsengs were saved after numerous photos were shot and gathered.

Table 1. Scoring criteria.

Project	Principal	First-Class	Second-Class
Rutabaga	Complete with rutabaga and ginseng *(Panax ginseng C. A. Meye)* fibrous roots	The rutabaga and ginseng fibrous roots are more complete	Rutabaga and ginseng with incomplete fibrous roots
Surface	Yellowish white or greyish yellow, without water embroidery or guttering	Yellowish white or greyish yellow, or with water rust or guttering	Yellowish white or greyish yellow with rust and guttering
Breakage Scar	None	Mild	Have
Branching out	No small ginseng or ginseng whiskers are allowed to be caught by those with branching roots, no tied tails or lightly tied tails or tied tails		
Section	Section pale yellowish white, pink		
Texture	Hard, powdery, no hollow		
Odor	Unique aroma, taste slightly bitter, sweet		
Main Root	Cylindrical		
Insects Mildew Impurities	None		

2.2. Data Enhancement

When the number and quality of the data were higher during the deep learning training phase, the model was more generalizable. As a result, the upper limit of the model learning was directly determined by the data. During the collection of the data, most of the photographs had difficulty in reproducing the full scene of the environment in which the samples were positioned, and the number of photographs taken was small. Because of this, we needed to perform data enhancement on the collected photos [20]. In this paper, two types of enhancement were adopted: offline enhancement and online enhancement.

The offline enhancement method applies random rotation, random adjustment of contrast, and random adjustment of brightness to expand the dataset. Following offline augmentation, the data were increased to 6131 frames. Information on the dataset is shown in Figure 1.

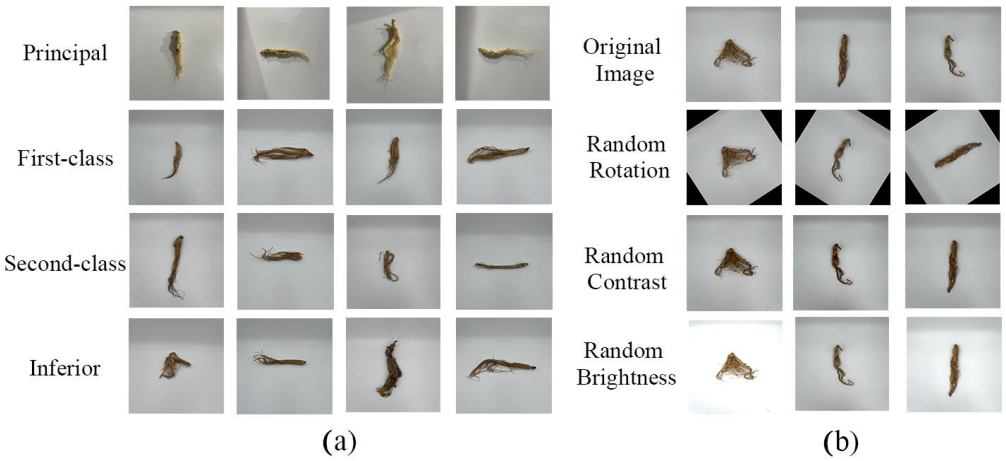


Figure 1. Ginseng *(Panax ginseng C. A. Meye)* dataset. (**a**) Different levels of ginseng images; (**b**) Image enhancement example diagram.

The online enhancement method uses the PIL (Python Image Library) module in the Python image library to uniformly crop a given image to 256 pixels by 256 pixels before each training round; then the images are processed using the center crop method, and finally the normalization is completed. This could indirectly increase the amount of training data by increasing the diversity of the image samples used in the training process.

2.3. Dataset Partitioning

To eliminate the serendipity of the experimental results, all training in this paper was conducted by using the five-fold cross-validation method. A randomly selected 80% of the dataset was used as the training set, and 20% of it as the test set. That is, 1225 images were used as validation data each time, while 4906 images were extracted from the training set as training data, and the validation data were not repeated. The final results were the average of five experiments, and the breakdown of the data for each class is shown in Table 2.

Table 2. Dataset partition.

Level	Training Set Image	Verification Set Image
Principal	1225	306
First-class	1396	348
Second-class	1589	397
Inferior	696	174

3. Building the Network Model

3.1. Resnet50 Model

He et al. proposed ResNet in 2015 [13], in which the residual block is an important structure. It enhances the transfer of features by introducing shortcut connections in the convolutional neural network, so that the next layer contains more information about the image. The + residual block consists of a Convolution Layer (Conv), a Batch Normalization Layer (BN), and a Rectified Linear Units Layer (*ReLU*). The structure of the residual block used by ResNet50 is shown in Figure 2. Here, x is the input to the network, and $F(x)$ represents the output after three Conv layers of processing. As the depth of the neural network increases, the image information in the feature map decreases layer by layer. Therefore, it is merged with the original output through the mapping of shortcut connections before being delivered to the subsequent layer $F(x) + x$ The issue that typical deep learning algorithms cause the network to burst and disappear as the number of layers grows, making the model difficult to converge, is resolved by this operation. The main branch contains three Conv layers, the first being a 1×1 Conv layer to compress dimensionality, the second a 3×3 Conv layer, and the third a 1×1 Conv layer to reduce dimensionality. This greatly maintains accuracy and reduces the computational effort [21].

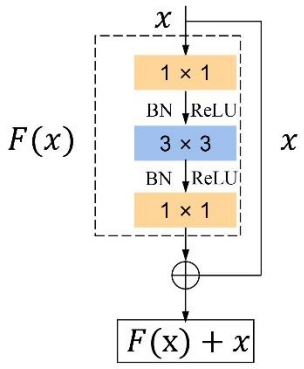


Figure 2. Residual block.

3.2. Improving the Resnet50 Model

3.2.1. Using the Leaky ReLU Activation Function

Using activation functions in neural networks can increase the non-linear variability of the neural network model. The traditional ResNet uses a *ReLU* activation function. It is characterized by fast computing speed and good performance. Its function expression is shown in Equation (1).

$$ReLU(x) = \begin{cases} 0, x \leq 0 \\ x, x > 0 \end{cases} \quad (1)$$

As can be seen in Figure 3, the gradient of this interval is constant when x of the *ReLU* function is greater than 0, thus alleviating the gradient disappearance problem. However, when the input value is negative during training, the function will reach hard saturation, resulting in the weights not being updated, i.e., the neuron death phenomenon.

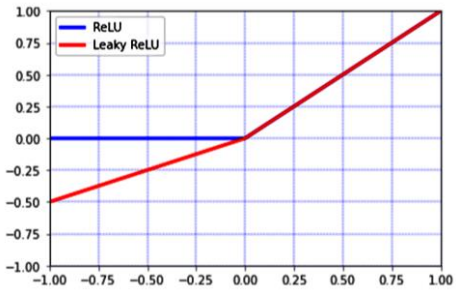


Figure 3. Comparison of the *ReLU* and *Leaky ReLU* functions.

To address this problem, we replaced the *ReLU* activation function with a *Leaky ReLU* [22]. The *Leaky ReLU* function has the characteristics of being able to achieve hard saturation, fast computation, and quick convergence compared to the conventional activation functions Sigmoid and Tanh. Its expression is shown in Equation (2).

$$Leaky\ ReLU(\mathrm{x}) = \begin{cases} x, x \geq 0 \\ \alpha x, x < 0 \end{cases} \quad (2)$$

where α is the value of the gradient, the default setting being $\alpha = 0.5$. The capacity of the *Leaky ReLU* function to preserve the gradient when the input information is less than 0 gives it an advantage over the *ReLU* function, allowing the parameters to continue to be updated, as can also be seen from the comparison chart in Figure 3. By doing this, the network's interpretability will be enhanced, and any data loss will be prevented.

3.2.2. Adding an Attention Mechanism

The use of an attention mechanism in convolutional neural networks has been much favored in recent years, as it can substantially improve a network performance by refining the feature mapping [23–25]. The channel attention mechanism has demonstrated the most potential for enhancing network performance among the available attention mechanisms. One of its representative networks is the Squeeze and Excitation Network (SENet) that incorporates the SE module (Squeeze and Excitation) [26]. However, the feature extraction process's dimensionality reduction has negative effects on channel attention prediction and ineffectively captures the dependencies between all channels. Therefore, Wang et al. then proposed a new efficient channel attention network, i.e., the ECA-Net (Efficient Channel Attention for Deep Convolutional Neural Networks), in 2019 [27]. The structure of the ECA module is shown in Figure 4. Without dimensionality reduction, the ECA module has the ability to adaptively choose one-dimensional convolutional kernels. Here is how it works.

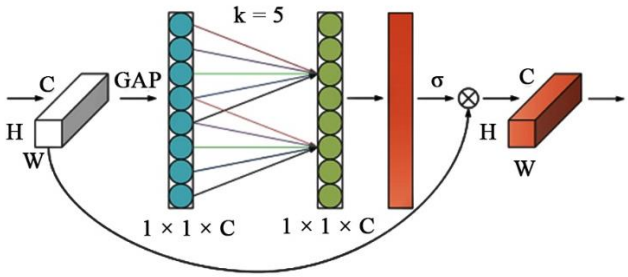


Figure 4. Structure of the Efficient Channel Attention Module.

(1) Input features: Given an input feature image X of size $H \times W \times C$, the input image is subjected to Global Average Pooling (GAP) to obtain all the feature information without dimensionality reduction.

(2) Calculation: The one-dimensional convolution operation with a convolution kernel of size k is used to efficiently complete the cross-channel information interaction and obtain the weights of each channel ω, as shown in Equation (3).

$$\omega = \delta(C1D_k(y)) \tag{3}$$

where δ represents the sigmoid activation function, and $C1D$ represents the one-dimensional convolution. The number of channels C is proportional to the one-dimensional convolution with kernel k, as shown in Equation (4).

$$C = 2^{(\gamma * k - b)} \tag{4}$$

The final kernel size k can be determined adaptively using Equation (5).

$$k = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{5}$$

where t is the nearest odd number to $|t|_{odd}$, b is 1, and γ is 2.

(3) Output features: The weights of each channel are obtained using the sigmoid function, and the input feature maps are channel-weighted to finally obtain the feature maps under different weights.

Since backpropagating the feature information is vulnerable to gradient dissipation close to the input layer, it is challenging to boost the network model's efficiency. The majority of ginseng plants share a similar shape and have fine textures and dense roots, which can affect their identification after downsampling and makes it difficult to extract detailed features from the network. Therefore, we improved the original residual block by adding the ECA module before the feature overlay, as shown in Figure 5. With the addition of the ECA module, the network enhanced the learning of channel attention features for each Conv block, and focusing the attention on hard-to-identify ginseng images allowed the features to be reused. As a result, the network model performed significantly better.

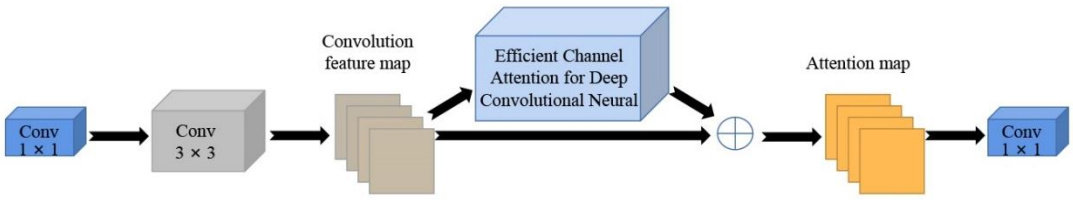


Figure 5. Adding the residual block structure of the ECA attention mechanism.

3.2.3. Introduction of a Focal Loss Function

The focal loss function (Focal Loss) [28] is the main solution to the problem of the existence of imbalances in the classification dataset and of mining difficult samples. It is based on an enhancement of the cross-entropy loss function [29]. Ordinary cross entropy states that for positive samples, the output probability increases, and the loss decreases; for negative samples, the output probability decreases, and the loss increases. In this case, the loss function improved slowly over the iterative process of a large number of simple samples and might not be able to optimize to the optimum. The original base was increased by the adjustment factor $\gamma(\gamma > 0)$. Its formula is shown in Equation (6); by reducing the loss of easy-to-classify samples, it concentrates more on complex, incorrectly classified samples. When γ is 0, we have the cross-entropy loss function.

$$L_{fl} = \begin{cases} -(1-y')^{\gamma} \log y', y = 1 \\ -y'^{\gamma} \log(1-y'), y = 0 \end{cases} \tag{6}$$

In addition to this, a balancing factor α was introduced to compensate the unequal distribution of positive and negative samples; its formula is given in Equation (7).

$$L_{fl} = \begin{cases} -\alpha(1-y')^{\gamma} \log y', y = 1 \\ -(1-\alpha) y'^{\gamma} log(1-y)', y = 0 \end{cases} \tag{7}$$

Due to the minimal number of inferior samples in the constructed ginseng dataset and the lack of distinctive features between different ginseng species, to reduce the weight of the negative samples in the training, a focal loss function was introduced; this allowed the model to obtain more accurate classification results.

3.2.4. Structure of the Ginseng Appearance Quality Grading Model

The classical ResNet consists of three parts: the input, the convolution, and the output. First, the data enter the network and go through the input section. Then, they go through the intermediate 1, 2, 3, and 4 Conv layer sections. Finally, the data go to the averaging pooling and fully connected layers to acquire the result. In this study, the original ResNet50 was reconstructed using the three suggested strategies for improvement. The enhanced network structure is shown in Figure 6.

The input feature map of this network first passes through a 7×7 Conv layer, preserving the image's original features over a substantial area. Then, after the BN (BatchNormalization) layer and the *Leaky ReLU* activation function, the features are extracted from the feature map using a 3×3 maximum pooling layer and compressed into a feature map with a channel count of 64. The feature map enters four layers in turn. Each layer consists of a Conv block and a variable number of ID (Identity) blocks. One Conv block and two ID blocks make up the first layer, one Conv block and three ID blocks make up layer 2, one Conv block and five ID blocks make up layer 3, and one Conv block and two ID blocks make up layer 4. Each Conv block and ID block adds the ECA module as indicated in Figure 6b,c. Eventually, a feature map with a channel count of 2048 is produced by successively passing the shallow spatial information and the underlying semantic information. The constructed model employs a global average pooling layer to optimize the network structure and adds a discard layer before the fully connected layer to prevent overfitting. The feature map is finally sent to the fully connected layer prediction to obtain the classification result.

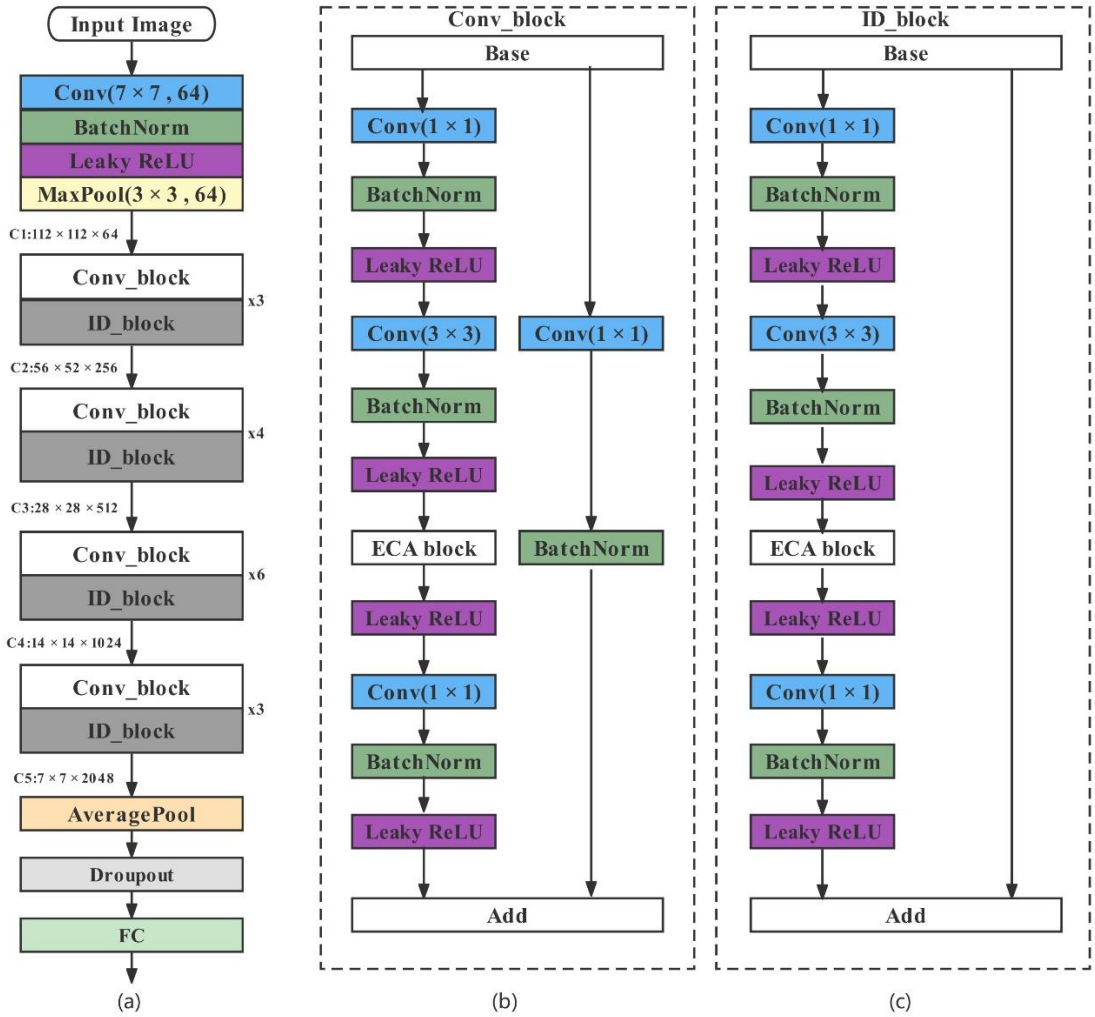


Figure 6. Structure of the improved model. (**a**) Backbone structure; (**b**) Conv_block; (**c**) ID_block. Note: Conv is Conv Layer; BatchNorm is Batch Normalization; ID_Block is identity block; Conv_block is convolution block; FC is full connection; ECA block is Efficient Channel Attention block.

3.3. Model Training for Transfer Learning

Transfer learning [30] involves moving pre-trained model parameters to a newly constructed model and sharing the newly learned knowledge with the new model in order to accelerate and optimize the learning speed of the model and bring it to convergence in the shortest possible time. Due to the fact that the model training used a supervised learning approach, a substantial amount of data samples were needed once the ginseng appearance quality grading model had been constructed. We had a limited number of samples, which made it challenging to set up the deep network training. However, the use of transfer learning methods could effectively improve the accuracy and generalization of the model. First, the constructed ginseng appearance quality grading model was initialized by loading the dataset from ImageNet. After that, the trained weights were transferred to the enhanced Conv layer. Finally, the ginseng dataset was loaded into the network, and then all layers aside from the fully connected layer were frozen and continuously fine-tuned for training until the optimal model was obtained [31].

4. Experimental Validation and Analysis of the Results

4.1. Experimental Setup and Analysis of the Results

We implemented our approach based on PyTorch. The processor of the experimental workstation was Xeon 4210 (8-core 2.45 GHz) (Intel., Santa Clara, CA, USA); the memory

was 64 G. the GPU was NVIDIA GeForce GTX 1080 ti (NVIDIA., Santa Clara, CA, USA); the running memory was 11 GBRAM. The software experimental configuration environment was Ubuntu 16.04; Python 3.7.0; Pytorch 1.10.1; CUDA 10.2. The specific parameter settings in the experiment are shown in Table 3.

Table 3. Model parameter settings.

Parameter	Set Up
Optimizer	Adam
Learning rate	0.0001
Weight decay	0.0001
Batch size	32
Epoch	50
Loss function	Focal Loss

To confirm the validity of the model, we compared the improved model with the classical convolutional neural network models VGG16, GoogLeNet, ResNet50, and DenseNet121. The accuracy rates increased by 3.76, 2.61, 2.45, and 1.88 percentage points, and the loss values decreased by 0.091, 0.042, 0.031, and 0.022 percentage points, as shown in Table 4. The comparison graph in Figure 7 shows that the improved model showed the highest recognition accuracy and the lowest loss value when compared to the other models as well as the shortest training time per round, and could quickly converge to find the optimal value; it also reached convergence quickly to find the optimal value. This demonstrated the superior performance of the model, indicating that it would provide a valuable reference for the subsequent application in ginseng appearance quality identification.

Table 4. Comparison of the experimental results of various models.

Model	Accuracy/%	Loss	Convergence/Epoch	Training Time per Round/s
VGG16	93.63	0.126	22	20
GoogLeNet	94.78	0.077	18	18
ResNet50	94.94	0.066	15	22
DenseNet121	95.51	0.057	12	25
Our model	97.39	0.035	5	19

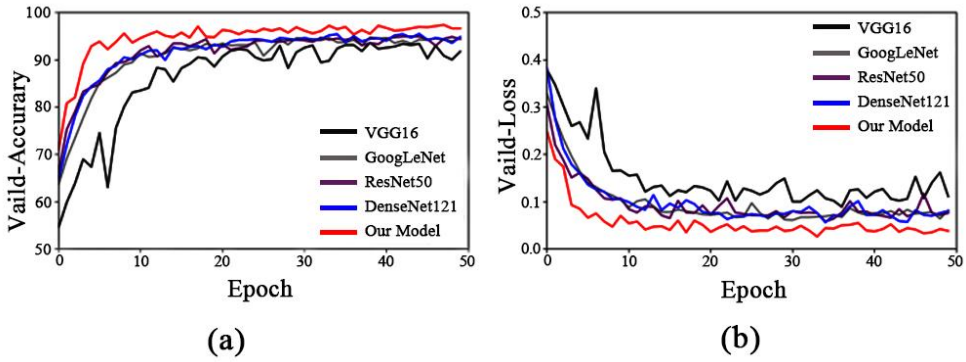


Figure 7. Identification results of each mode. (**a**) Model accuracy; (**b**) model loss.

4.2. Impact of Attention Mechanisms on the Model Performance

In our experiments, we found that the addition of different attention mechanisms also had some effects on the model's performance when all other factors were equal. As shown in Table 5, the model improved its recognition accuracy by 0.74, 1.31, and 1.72 percentage

points, and the loss values decreased by 0.002, 0.02, and 0.021 percentage points, respectively, with the addition of the *Leaky ReLU* activation function alone and then the addition of the SE channel attention mechanism [26], the CBAM (Convolutional Block Attention Module) convolutional attention mechanism [32], and the ECA efficient channel attention mechanism. While the loss values decreased by 0.002, 0.02 and 0.021 percentage points, the recognition accuracy increased by 0.74, 1.31 and 1.72 percentage points, respectively. In contrast, the addition of the SK (Selective Kernel) attention mechanism [33] resulted in a 2.857 percentage point decrease in identification accuracy and a concomitant 0.019 increase in loss value. The results showed that the ECA mechanism had a more significant role in improving the performance of the network model than the other attention mechanisms. However, the results of incorporating the SK attention mechanism also demonstrated that the introduction of an attention mechanism into the network model does not guarantee an accurate identification; therefore the choice of the attention mechanism needs to be made appropriately for different tasks.

Table 5. Comparison of the results obtained with different attention mechanisms.

Model	Accuracy/%	Loss
No attention	95.67	0.056
SE	96.41	0.054
CBAM	96.98	0.036
SK	92.82	0.075
ECA	97.39	0.035

Note: SE is Squeeze and Excitation; CBAM is Convolutional Block Attention Module; SK is Selective Kernel.

Furthermore, the visualization images obtained by the tool Grad-Cam [34] provided a more intuitive view of the regions of the network's attention to the feature maps before and after the addition of the ECA mechanism. Figure 8 clearly demonstrates how effectively the ECA mechanism module we introduced could capture the ginseng features. It overcame the conflict between performance and complexity by using fewer parameters. To enhance the feature representation of the network model, it efficiently made use of the ability of convolutional neural networks to capture details about neighborhoods. The complex texture of ginseng and slight variations in appearance could also be used to effectively address the negative impact on experimental accuracy.

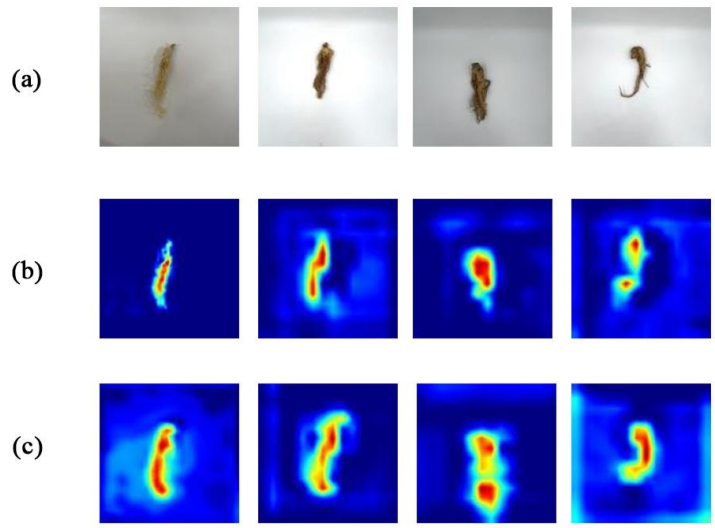


Figure 8. Visualization results of the thermal characteristic diagram of the new network before and after adding the ECA module. (**a**) Input image; (**b**) image before joining the ECA module; (**c**) image after adding the ECA module.

4.3. Effect of the Activation Function on the Model Performance

The selection of an activation function is also crucial during the training process, since it has a significant impact on how well the same model performs. During the experiment, we replaced the *Leaky ReLU* activation function with three activation functions, *ReLU*, Sigmoid, and Tanh. The accuracy and loss values employing the two activation functions Sigmoid and Tanh differed significantly from those obtained with *ReLU*, as shown in Table 6. In contrast to the *ReLU* activation function, the *Leaky ReLU* function showed an improvement in accuracy of 0.63 and a reduction in loss value of 0.023. There are two reasons for this result: (1) the *Leaky ReLU* activation function could transmit negative weights, which makes the model more capable of pushing information; (2) the *Leaky ReLU* activation function transmitted more detail information, such as texture, line, and color of ginseng, to extract feature detail information that was not easily extracted by the network, which led to a substantial improvement in model performance.

Table 6. Experimental comparison results of different activation functions.

Activate Function	Accuracy/%	Loss
Sigmoid	63.84	0.308
Tanh	75.76	0.216
ReLU	96.76	0.058
Leaky ReLU	97.39	0.035

4.4. Effect of the Loss Function on the Model Performance

To investigate the effect of various loss functions on the model performance, we selected a cross-entropy loss function for comparison. Although the cross-entropy loss function convergence effect is very good, but it still has limitations: in real applications, containing the complex background information of an image—such as, in this case, the fact that the number of ginseng pixels was much smaller than the number of background pixels—can lead to prevailing background information in the cross-entropy loss function composition; therefore, the network model is clearly biased, resulting in a decline in recognition effect. As can be seen in Table 7, the accuracy of the model utilizing the focal loss function increased to some extent both before and after the improvement, with a significant decrease in the loss values. This indicated that the introduction of a focal loss function could adequately address the problem of sample imbalance in the dataset and has a positive influence on the final classification of the model.

Table 7. Experimental comparison of the results obtained with different loss functions.

Model	Loss Function	Accuracy/%	Loss
ResNet50	Cross entropy Loss	94.78	0.176
	Focal Loss	94.94	0.066
Our model	Cross entropy Loss	97.22	0.119
	Focal Loss	97.39	0.035

4.5. Model Evaluation

In image classification, the prediction results of a classification model are frequently represented as a confusion matrix, from which the metrics Accuracy, Recall, and Precision are derived:

$$Accuracy = \frac{TP}{TP + FN + FP + TN} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

where TP (true positive) is the number of ginseng specimens accurately identified, FP (false positive) is the number of ginseng specimens incorrectly identified, FN (false negative) is the number of ginseng specimens identified as other types of ginseng, and TN (true negative) is the number of other ginseng specimens accurately identified.

In the confusion matrix in Figure 9, 1 represents the principal ginseng, 2 the first-class ginseng, 3 the second-class ginseng, and 4 the inferior ginseng; the horizontal coordinates represent the true category, and the vertical coordinates represent the predicted category. According to the confusion matrix in Figure 9, the ResNet50 model has an overall accuracy of 94.94% for the test set, an overall recall of 94.37%, and an accuracy of 94.99%. The overall accuracy of our proposed model's test set was 97.39%, while its overall recall was 97.03%, and its accuracy was 97.29%.

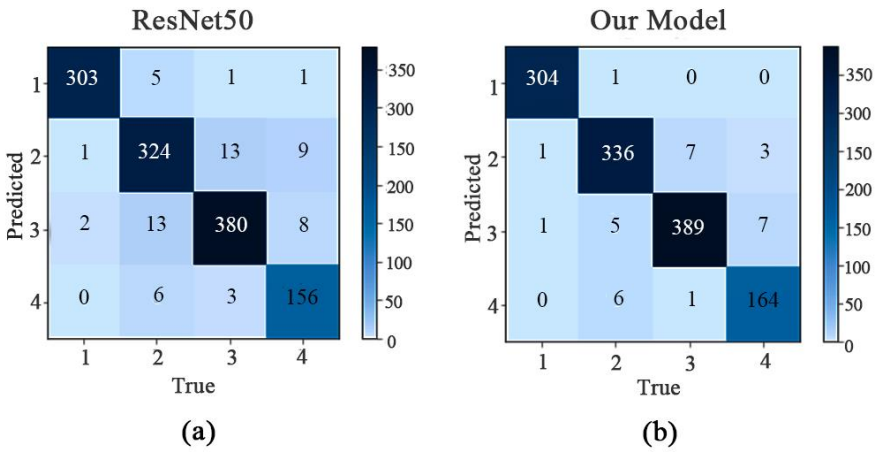


Figure 9. Confusion matrix before and after the model improvement. (**a**) ResNet50 confusion matrix; (**b**) our model confusion matrix. Note: 1 is principal; 2 is first-class; 3 is second-class; 4 is inferior. The x-axis is the real category, and the y-axis is the predicted category.

In Table 8, the improved model enhanced the recall of all four types of ginseng to different degrees compared to the original model. However, the model still showed a large error in the identification of "inferior" ginseng; thus, improving the identification rate of inferior ginseng will be the focus of our future research.

Table 8. Recall and accuracy of ginseng classification by class before and after model improvement.

Level	Model	Recall/%	Precision%
Principal	ResNet50	99.02	97.74
	Our model	99.35	99.67
First-class	ResNet50	93.10	93.37
	Our model	96.56	96.83
Second-class	ResNet50	95.72	94.29
	Our model	97.98	96.77
Inferior	ResNet50	89.66	94.55
	Our model	94.25	95.91

5. Discussion

Ginseng is a traditional Chinese medicine also used as food and is also a common health care medicine, very appreciated by people. The identification of the quality of ginseng is therefore of paramount importance and is an important area of concern for the majority of researchers in this field, today. The traditional methods of identifying the appearance and quality of ginseng do not offer advantages in the age of artificial intelligence; therefore, after learning about the achievements of convolutional neural networks in the field of image processing, we decided to use ResNet50 as a base model to improve the

evaluation of ginseng quality based on the small differences in features and details found in ginseng plants of different levels of quality. However, as can be seen from the confusion matrix with the original model, inferior ginseng had the lowest recognition accuracy, which also reflects the need to further improve the model's ability to extract features in order to provide better recognition results when using a ginseng dataset with highly similar specimens. We subjected the improved model to multiple ablation experiments on the ginseng dataset, and the results showed that the improved ResNet50 outperformed other models mainly in terms of accuracy and loss values and offered greater advantages in terms of convergence speed and stability; it also showed a slightly faster training time per round than the other models. The model performance was further improved with a recognition accuracy of 97.39%. We also proved through experiments that the method could improve the identification accuracy without damaging the appearance of the ginseng specimens and basically met the identification needs.

6. Conclusions

The focus of our future research is on the further optimization of this model structure, with reference to the needs of realistic applications, focusing on reducing the weight of the model so that it could still achieve effective classification tasks with fewer parameters. As this study was only about the classification of white ginseng specimens of different quality, we will next construct a dataset for ginseng from different origins and different years to analyze the effect of the appearance characteristics of ginseng from different backgrounds on its classification. In addition, we will deploy the model on mobile devices for use in public promotion, market supervision, and industrial production. We will make it easier and faster for people to identify the quality of ginseng, thus improving public awareness, market supervision, and industrial production efficiency. This will provide a better solution to the problem of identifying the appearance of ginseng quality in theory and practice.

In this paper, the traditional ResNet50 network was improved according to the characteristics of ginseng. Our conclusions are as follows.

First of all, we constructed a ginseng dataset and classified the data to identify samples of four different quality levels: principal, first-class, second-class, and inferior ginseng. Data augmentation was then performed on the dataset to extend it. Then, we replaced the traditional activation function *ReLU* with the *Leaky ReLU* activation function to enhance the expressiveness of the model. We also introduced the ECA mechanism module on the residual block to increase the model's sensitivity to ginseng pixels and better capture the ginseng specimens' features. Additionally, the focal loss function was introduced to balance the dataset, and the idea of transfer learning was used to train the model. Finally, our proposed model showed greater advantages in different aspects compared to the classical convolutional neural network models Vgg16, GoogLeNet, ResNet50, and Densenet121. Many comparison experiments clearly showed that the method proposed in this paper had a beneficial impact on the model performance. Compared with the original model, the accuracy and loss value of the improved model were the best. In addition, with the significant improvement of the model performance indicators, the accuracy and overall recall rate were also improved. These results also validated the effectiveness and feasibility of the improvements to the original model. In future studies, we plan to further optimize the model and combine it with a mobile terminal to provide technical support for ginseng appearance quality recognition.

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

Funding: This research was funded by the National Natural Science Foundation of China (No. 61801439); the Department of Science and Technology of Jilin Province (20210204050YY); Jilin Provincial Education Department Scientific Research Project (JJKH20210747KJ); Jilin Provincial Environmental Protection Department Project (202107); Jilin Provincial Middle and Young Leaders Team and Innovative Talents Support Program (No. 20200301037RQ).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All relevant data are included in the manuscript. Raw images are available on request from the corresponding author.

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

References

1. Li, F.; Bao, H. Herbal Textual Research and progress on pharmacological actions of Ginseng Radix et Rhizoma. *Ginseng Res.* **2017**, *29*, 43–46.
2. Liu, W.; Li, W. Review on industrialization development status and prospect of panax ginseng processing. *J. Jilin Agric. Univ.* **2022**, 1–11.
3. Dai, Q.; Li, L.; Xu, G.; Tang, D.; Ma, H. Study on the Correlation between Morphological Characteristics and Chemical Constituents of GardenGinseng, Mountain Transplanted Ginseng and Mountain-grown Ginseng Based on "Evaluating Qualityfrom Morphological Characteristics". *China Pharm.* **2020**, *31*, 650–655.
4. Li, J.; Wei, X.; Wan, G.; Yang, X.; Shi, J.L. Historical Evolution and Modern Research Progress of Quality Evaluation Based onCharacter Identification of Traditional Chinese Medicinal Materials. *Chin. J. Exp. Tradit. Med.* **2021**, *27*, 189–196.
5. Liu, T.-R.; Jin, Y.; Meng, H.-B.; Zhao, Y.-Y.; Zhou, J.-H.; Yuan, Y.; Huang, L.-Q. Biological research of color and quality evaluation in"quality discriminationby character"of Chinese medicine. *China J. Chin. Mater. Med.* **2020**, *45*, 4545–4554.
6. Wang, H.; Tian, Y.; Liu, D.; Ma, X.; Zhan, Z.; Huang, L.; Du, H. History, Development and Application of the Traditional Chinese Medicine "Quality Evaluation ThroughMorphological Identification". *J. Chin. Med. Mater.* **2021**, *44*, 513–519.
7. Li, M.; Zhang, X.; Liu, S.; Chen, X.; Huang, L.; Shi, T.; Yang, R.; Liu, S.; Zheng, F. Partly Interpretable Machine Learning Method of Ginseng Geographical Origins Recognition and Analysis by Hyperspectral Measurements. *Spectrosc. Spectr. Anal.* **2022**, *42*, 1217–1221.
8. Geng, L.; Huang, Y.; Guo, Y. Apple variety classification method based on fusion attention mechanism. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 304–310 + 369.
9. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
10. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
11. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
12. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
13. 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.
14. Huang, G.; Liu, Z.; Pleiss, G.; van der Maaten, L.; Weinberger, K.Q. Convolutional Networks with Dense Connectivity. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 8704–8716. [CrossRef]
15. Dyrmann, M.; Karstoft, H.; Midtiby, H.S. Plant species classification using deep convolutional neural network. *Biosyst. Eng.* **2016**, *151*, 72–80. [CrossRef]
16. Lee, S.H.; Chan, C.S.; Mayo, S.J.; Remagnino, P. How deep learning extracts and learns leaf features for plant classification. *Pattern Recognit.* **2017**, *71*, 1–13. [CrossRef]
17. Liu, B.; Ding, Z.; Tian, L.; He, D.; Li, S.; Wang, H. Grape Leaf Disease Identification Using Improved Deep Convolutional Neural Networks. *Front. Plant. Sci.* **2020**, *11*, 1082. [CrossRef]
18. Liu, W.; Zou, W.; Lu, Y.; Hu, W.; Liu, T. Research on image recognition of chinese medicinal materials based on transfer learning and model fusion. *J. Hunan Univ. Chin. Med.* **2022**, *42*, 809–814.
19. Li, D.; Tang, P.; Zhang, L.; Lei, Y.; Liu, S. Genuine identification for Saposhnikovia divaricata based on improved DenseNet. *Trans. Chin. Soc. Agric. Eng.* **2022**, *38*, 276–285.
20. He, J.; Shi, Y.; Liu, B.; He, D. External Quality Grading Method of Fuji Apple Based onDeep Learning. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 379–385.
21. He, D.; Wang, P.; Niu, T.; Mao, Y.; Zhao, Y. Classification Model of Grape Downy Mildew Disease Degree in FieldBased on Improved Residual Network. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 235–243.

22. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models. In Proceedings of the 30th International Conference on International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; JMLR: Cambridge, MA, USA, 2013; p. 3.
23. Zhu, X.; Cheng, D.; Zhang, Z.; Lin, S.; Dai, J. An Empirical Study of Spatial Attention Mechanisms in Deep Networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6688–6697.
24. Yan, H.; Liu, Z.; Cui, Q.; Hu, Z. Multi-target detection based on feature pyramid attention and deep convolution network for pigs. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 193–202.
25. Choi, H.; Cho, K.; Bengio, Y. Fine-grained attention mechanism for neural machine translation. *Neurocomputing* **2018**, *284*, 171–176. [CrossRef]
26. 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.
27. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. Supplementary Material for 'ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; IEEE: Piscataway, NJ, USA; pp. 13–19.
28. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
29. De Boer, P.-T.; Kroese, D.P.; Mannor, S.; Rubinstein, R.Y. A tutorial on the cross-entropy method. *Ann. Oper. Res.* **2005**, *134*, 19–67. [CrossRef]
30. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
31. Zhang, R.; Li, Z.; Hao, J.; Sun, L.; Li, H.; Han, P. Image recognition of peanut pod grades based on transfer learning with convolutional neural network. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 171–180.
32. 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.
33. Li, X.; Wang, W.; Hu, X.; Yang, J. Selective Kernel Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 510–519.
34. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-Cam: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.

agronomy

MDPI

Article

Development of Weed Detection Method in Soybean Fields Utilizing Improved DeepLabv3+ Platform

Helong Yu [1,2], Minghang Che [1], Han Yu [1,*] and Jian Zhang [1,2,3,*]

[1] College of Information Technology, Jilin Agricultural University, Changchun 130118, China
[2] Faculty of Agronomy, Jilin Agricultural University, Changchun 130118, China
[3] Department of Biology, University of British Columbia, Kelowna, BC V1V 1V7, Canada
* Correspondence: yuhan@jlau.edu.cn (H.Y.); jian.zhang@ubc.ca (J.Z.)

Abstract: Accurately identifying weeds in crop fields is key to achieving selective herbicide spraying. Weed identification is made difficult by the dense distribution of weeds and crops, which makes boundary segmentation at the overlap inaccurate, and thus pixels cannot be correctly classified. To solve this problem, this study proposes a soybean field weed recognition model based on an improved DeepLabv3+ model, which uses a Swin transformer as the feature extraction backbone to enhance the model's utilization of global information relationships, fuses feature maps of different sizes in the decoding section to enhance the utilization of features of different dimensions, and adds a convolution block attention module (CBAM) after each feature fusion to enhance the model's utilization of focused information in the feature maps, resulting in a new weed recognition model, Swin-DeepLab. Using this model to identify a dataset containing a large number of densely distributed weedy soybean seedlings, the average intersection ratio reached 91.53%, the accuracy improved by 2.94% compared with that before the improvement with only a 48 ms increase in recognition time, and the accuracy was superior to those of other classical semantic segmentation models. The results showed that the Swin-DeepLab network proposed in this paper can successfully solve the problems of incorrect boundary contour recognition when weeds are densely distributed with crops and incorrect classification when recognition targets overlap, providing a direction for the further application of transformers in weed recognition.

Keywords: attention mechanism; improved DeepLabv3+ model; semantic segmentation; transformer; weed recognition

Citation: Yu, H.; Che, M.; Yu, H.; Zhang, J. Development of Weed Detection Method in Soybean Fields Utilizing Improved DeepLabv3+ Platform. *Agronomy* **2022**, *12*, 2889. https://doi.org/10.3390/agronomy12112889

Academic Editor: Enrico Borgogno-Mondino

Received: 28 August 2022
Accepted: 7 November 2022
Published: 18 November 2022

1. Introduction

Crop growth may be affected by climatic, soil and biological factors, among which weeds are one of the main biological influences. Weeds compete with crops for nutrients, sunlight, growing space and water during their growth, and may adversely affect the growth of crops if they are not removed in a timely manner [1]. Crop yield and quality are affected by pests, diseases, and weeds, with global crop yield losses approaching 30% per year [2,3]. Effectively removing weeds is critical for growing and increasing crop yield. Traditional chemical weed control involves spraying excessive amounts of herbicides indiscriminately across a given crop field, leading to the overuse of herbicides, which is both wasteful and harmful to environment and consumer health [4–6]. If it is possible to achieve the targeted and quantitative spraying of appropriate herbicides on different kinds of weeds, weed control efficiency could be improved, and herbicide use reduced. The key issue that needs to be solved to achieve selective spraying is determining how to accomplish the real-time detection and differentiation of crops and weeds [5].

To achieve the accurate detection of crops and weeds, a few problems need to be solved [7], such as the uneven density and distribution of weeds, different lighting conditions under different weather conditions resulting in reduced recognition accuracy, similar shapes and colors of crops and weeds, and the existence of shading and overlapping

leaves [8]. In recent years, machine learning has been widely used in weed identification [9]. Traditional machine-learning-based algorithms use feature descriptors to extract object features from sensory data and use machine-learning-based classifiers for classification, detection, or segmentation [10]; these mainly include supervised learning algorithms such as k-nearest neighbor algorithm and logistic regression, as well as unsupervised learning algorithms such as clustering and principal component analysis (PCA). In detecting crops and weeds, color features, texture features, location information, and multispectral features are mainly used [11]. Deng et al. [12] combined the color, shape, and texture features of weed images to solve the problem of the low accuracy of single-feature recognition of weeds in rice fields. Ashraf et al. [13] proposed two techniques for weed-density-based image classification. The first technique used texture features extracted from a grayscale cogeneration matrix (GLCM) and obtained 73% accuracy using a radial basis function (RBF) kernel in a support vector machine (SVM), while the second method outperformed the first technique using a random forest classifier with 86% accuracy. Le et al. [14] proposed a combination of local binary pattern (LBP)-based operators for extracting crop leaf texture features. The accuracy of the combined LBP algorithm was investigated, and an accuracy of 91.85% was achieved on a multiclass plant classification task. Wendel et al. [15] used vegetation separation techniques to remove the background, followed by principal component analysis (PCA) for different spectral preprocessing to extract features; finally, an SVM was used for classification, achieving an F1-score of 0.91. Traditional machine learning techniques can be used as tools for weed identification, but these techniques require significant domain expertise to construct feature extractors from raw data [16]. This increases the difficulty of model reuse between different domains, and the recognition accuracy is not as high as that of deep learning.

Deep learning is an important branch of machine learning and convolutional neural networks are the foundation of deep learning. A convolutional neural network is a complex mesh of multiple convolutional layers, pooling layers, fully connected layers and nonlinear transforms that can automatically learn from labeled data to acquire complex intrinsic features and use them to recognize unlabeled data. For image classification, target detection, and segmentation problems, DL algorithms have many more advantages than traditional machine learning methods. Due to the similarity of crops and weeds, it is difficult to extract and select features using ML methods; deep-learning-based methods have strong feature learning ability, which can effectively solve this problem [6,16]. Huang et al. [17] used a fully convolutional network (FCN) [18] semantic segmentation model with AlexNet [19], VGGNet [20], and GoogLeNet [21] as the backbone to identify weeds in rice fields, where VGGNet had the highest accuracy. They further compared this model with patch-based and pixel-based CNN structures. The results showed that the VGG-16-based FCN model had the highest classification accuracy. Fawakherji et al. [22] proposed a three-step method based on crop and weed classification. The first step was pixel-level segmentation using ResNet-101 to separate plants [23] from the soil background, the second step was the extraction of image blocks containing plants, and the third step was the classification of crops and weeds using the UNet model with VGG-16 as the backbone network; this method achieved 80% accuracy for sugar beet. Lottes et al. [24] used an FCN with an encoder–decoder structure and merged spatial information when considering image sequences, and the model was trained using RGB and NIR images. The results showed that the method greatly improved the accuracy of crop weed classification. Similarly, Li et al. [25] proposed a method to use spatial information. They proposed a new network to identify crops in the case of dense weed distribution in a field. The network used ResNet-101 as the backbone, and introduced a short link structure to achieve multiscale feature fusion to improve the segmentation of target boundaries and fine structure. The method is general, and can be applied to different crops.

Although these deep-learning-based weed recognition methods can achieve high accuracy, there is still room for improvement. Due to the short growth cycle of weeds, weeds with different growth periods may exist simultaneously in a growth area, which requires the

model to recognize multiscale targets, and there is also the problem of insufficient features due to the partial obstruction of weed features. People have started to focus on spatial information acquisition and multiscale feature extraction as the research focus [24,25]. DeepLabv3+ [26] is the first model that introduces null convolution into the field of semantic segmentation, and achieves multiscale feature extraction through null convolution with different null rates. Ramirez et al. [1] attempted to identify weeds using a DeepLabv3 model. They compared SegNet [27], U-Net, and DeepLabv3 models for the identification of crop segmentation on sugar beet farms, and found that DeepLabv3 obtained the highest accuracy with an AUC of 89% and F1-score of 81%. Wu et al. [28] designed a vision system for segmenting abnormal leaves of hydroponic lettuce using a DeepLabv3+ model with different backbone networks, and after an experimental study, ResNet-101 produced the best results, with an average intersection ratio of 83.26%.

Previous research utilized multiscale features and spatial information to improve detection accuracy, but due to the narrow perceptual field of the convolutional operation itself, the utilization of global information is still insufficient, and there is still room to improve detection accuracy. This study aims to explore the solutions to the problems of mutual occlusion caused by the dense distribution of crops and weeds in weed segmentation, the degradation of recognition accuracy caused by similar colors of crops and weeds, and the inaccuracy of kernel edge segmentation. First, some photos of the dense distribution of weeds and crops in soybean fields were collected as datasets, and data enhancement was performed to address changes in different environments. Then, this dataset was used to investigate the above issues, and a deep learning model is proposed to make improvements for the above problem.

2. Materials and Methods

2.1. Image Acquisition

The soybean weed dataset used in this experiment was collected from a soybean experimental field at Jilin Agricultural University in Changchun, Jilin Province, China, between 9:00 and 15:00 on 10 and 16 June 2021, which corresponded to the main working hours of weeding work. The device used was a Huawei mate30 cell phone, with a shooting angle perpendicular to the ground, 60 cm from the ground, a resolution of 3000 × 4000 pixels, and JPG format images.

2.2. Image Preprocessing

A total of 520 images of 512 × 512 pixels were collected in this experiment, and some unclear images were eliminated, resulting in 502 images. The data images mainly contain soybean crops; graminoid weeds such as *Digitaria sanguinalis* (L.) Scop and *Setaria viridis* (L.) Beauv; broadleaf weeds such as *Chenopodium glaucum* L., *Acalypha australis* L., and *Amaranthus retroflexus* L.; as well as background consisting of soil, stones, and dead plants. Therefore, the experiment divides the segmented target into soybeans, graminoid weeds, broadleaf weeds, and the remaining unsegmented target into background. The distribution of weeds and crops in the dataset used in this experiment is complex, containing many weeds and crops shading each other, which makes identification difficult. Some pictures of the dataset are shown in Figure 1.

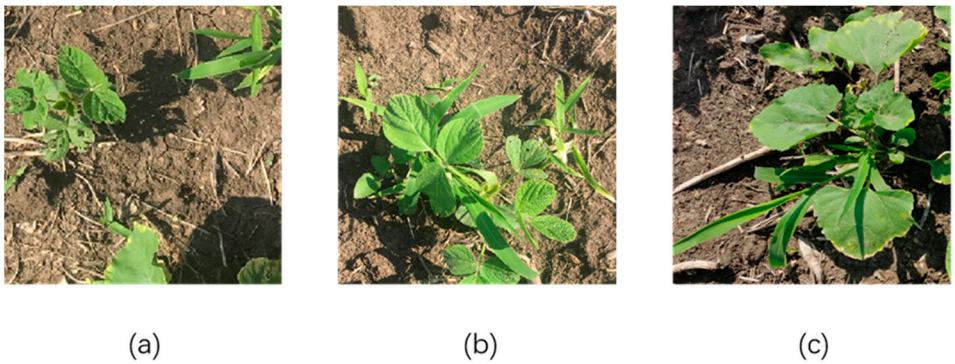


Figure 1. Examples of images of selected datasets: (**a**) image showing a sparse distribution of weeds and crops; (**b**,**c**) images showing the dense distribution of weeds and crops.

In this study, the LabelMe tool was used to label the images, and the image pixels were classified into four categories: soybeans, graminoid weeds, broadleaf weeds, and background. The labeling results are shown in Figure 2. In general, effective data expansion can better improve the robustness of a model and enable the model to obtain stronger generalization ability [29]. For this purpose, this study expanded the images using random rotation, random flipping, random cropping, adding Gaussian noise, and increasing contrast to obtain 2410 new images, which were randomly divided into a training set and a test set with a ratio of 7:3.

Figure 2. The images on the top are the original images, and the images on the bottom are their labeled images. The RGB values in the labeled images are (0,0,0) for the background area (black), (128,0,0) for soybeans (red), (0,128,0) for graminoid weeds (green), and (128,128,0) for broadleaf weeds (yellow).

2.3. Model Structure

Semantic segmentation has proven its effectiveness in a large number of experiments for detecting the contours of irregularly sized and shaped objects; in this study, the DeepLabv3+ model, which can fuse multiscale features, was used as the base segmentation model for segmenting soybean and weeds. DeepLabv3+ [26] is the classical semantic segmentation model, it is a more powerful encoder–decoder structure proposed by the original authors of the DeepLab series based on the DeepLabv3 [30] model as the encoder, which uses the Xception [31] model as the backbone network for feature extraction

and successively obtains a shallow feature map and a deep feature map. After obtaining the deep feature map, an atrous spatial pyramid pooling (ASPP) module with different void rate convolution operations is used to collect information at different scales in the deep feature map to enhance the recognition of features at different scales. Thus, the deep feature map processed by the ASPP module is fused with the shallow feature map to reduce the loss of edge information at the decoder stage; this could improve the accuracy of the segmentation boundary. Finally, upsampling is performed to obtain the pixel-level classification results.

The network structure proposed in this study is shown in Figure 3; it consists of two stages: encoder and decoder. The distribution of weed leaves in the image data is radial, and the leaves are far apart from each other, which will lead to the insufficient extraction of model contextual feature information. To address this issue, we used the Swin transformer to replace the backbone network of the original DeepLabv3+ model. The Swin transformer can capture the remote dependencies between pixels, utilize the spatial information, and increase the segmentation accuracy.

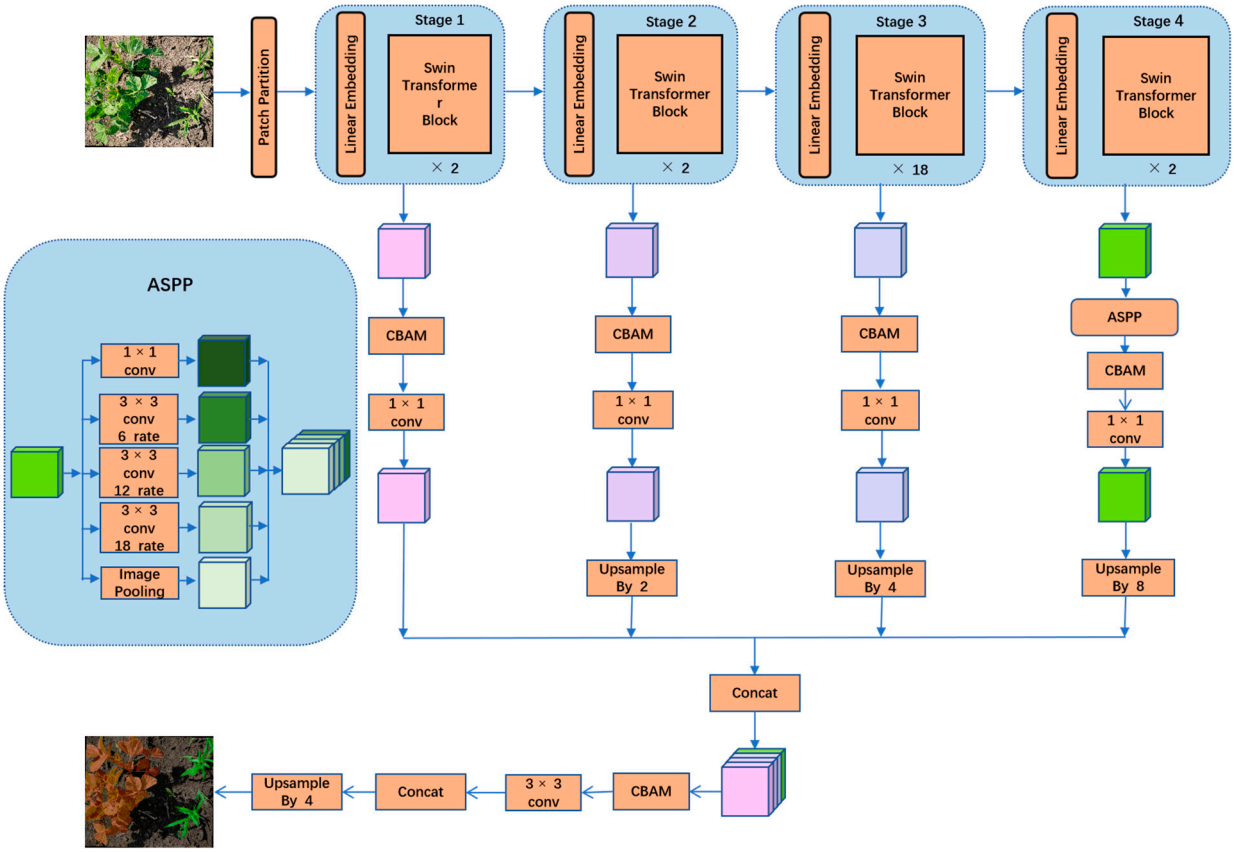


Figure 3. Structural diagram of the Swin-DeepLab model.

The encoder stage is divided into 4 building layers with 2, 2, 18, and 2 Swin transformer blocks, respectively. In the first layer, the size of the input feature map is 512 × 512 × 3. We use the Patch Partition layer to cut the feature map in units of 4 pixels along its height (H) and width (W), and the input feature map is divided into patches with feature size of 4 × 4 × 3; then, each patch is flattened along the channel direction, and the size of the input feature map becomes 128 × 128 × 48. The output feature map is computed using a Linear Embedding layer to project its feature dimension to a specific dimension C (the size of C in this experiment is 96), then the Swin transformer block is used to compute the self-attentiveness of the feature map and generate the output feature map.

In the last three construction layers, to generate multi-dimensional feature maps, the feature maps are first cut into multiple patches using Patch Merging, then they are

reorganized and linearly transformed to halve the feature map dimension and double the number of channels; finally, the output feature maps are computed using the corresponding number of Swin transformer blocks. The input feature maps of the four construction layers are as follows. The feature maps generated by the four stages are a shallow feature map $\frac{H}{4} \times \frac{W}{4} \times C$, a subshallow feature map $\frac{H}{8} \times \frac{W}{8} \times 2C$, a subdepth feature map $\frac{H}{16} \times \frac{W}{16} \times 4C$, and a deep feature map $\frac{H}{32} \times \frac{W}{32} \times 8C$, and the four feature maps are used to recover the feature maps in the decoding stage.

In the decoding stage, the deep feature map is used as the input to the ASPP module. Then, the four feature maps are upsampled using the deconvolution layer. Next, the feature maps are all resized to $\frac{H}{4} \times \frac{W}{4} \times 48$ and stitched by channel dimension using the concat operation to obtain a total feature map of $\frac{H}{4} \times \frac{W}{4} \times 192$, followed by a 3×3 convolution kernel to compress their channels to $\frac{H}{4} \times \frac{W}{4} \times 48$. Finally, quadruple upsampling is used to restore the feature maps to their original size. Among them, after the four feature maps are generated and after the four feature maps are stitched together, the CBAM is added to increase the focus of the model on the channels and the key regions in space. This model is named Swin-DeepLab.

2.3.1. CBAM

In the weed identification task, the color features of the individual segmented objects are very similar, the shape features change due to the leaves shading each other, and the accuracy of the segmentation of the leaf edges is insufficient. To capture key information in the images, allow the algorithm to focus more on plant morphological features in space, and enhance the use of focused channel information, the model uses a CBAM [32] to deepen the attention of feature maps of different sizes generated by the backbone network. The CBAM structure is shown in Figure 4. The CBAM is a channel–space dual-attention lightweight module Similar to SENet [33], which can improve the model performance with only a small increase in computation and can be directly inserted into the existing network architecture. Therefore, we use the CBAM attention module for each dimensional feature map after its generation to refine the feature map and increase the effect of feature fusion.

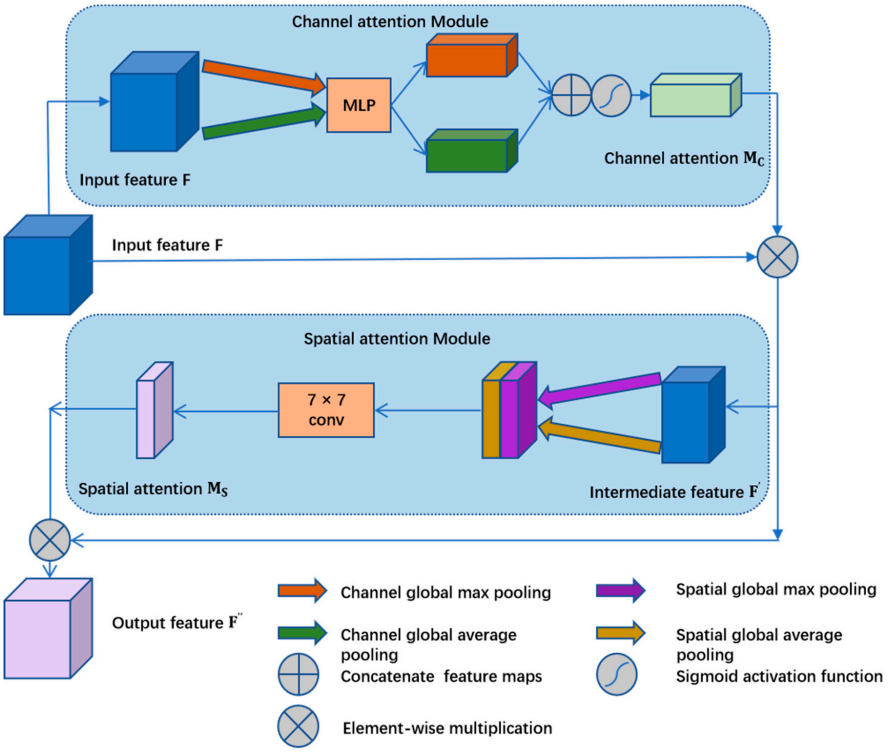


Figure 4. CBAM structure diagram.

The CBAM infers the attention map sequentially along two separate dimensions (channel and space), and first captures the channel features through the channel attention computation module to make the model pay more attention to the channels with more important semantic information. In the channel attention computation module, the original feature map $F^{H\times W\times C}$ is first computed by global average pooling (AvgPool) and global maximum pooling (MaxPool), separately, to generate two different feature maps F^c_{avg} and F^c_{max} of size $C\times 1\times 1$, which extract channel features with their own focus. Then, the two feature maps are dimensionally compressed and recovered using a multilayer perceptron with shared weights, first compressing to the size of $\frac{C}{16}\times 1\times 1$ and then restored to the size of $C\times 1\times 1$. This process can increase the correlation between the channels, and finally the two feature maps are added elementwise after the activation of the sigmoid activation function to obtain the channel weight vector. The generated channel weight vector performs an elementwise summation operation with the original feature map to increase the weight of the key channel in the original feature map and generate the intermediate feature map, where the channel attention mechanism is calculated as follows:

$$M_c(F) = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MLP}(\mathrm{MaxPool}(F))) \quad (1)$$

where σ denotes the sigmoid function and MLP is a fully connected layer.

Next, we use the spatial attention calculation module to calculate the spatial attention of the intermediate feature map, and the calculation process involves making the global average pooling (AvgPool) calculation and the global maximum pooling (MaxPool) calculation for the channel dimension to achieve the compression of the spatial dimension, obtain two feature maps $F^s_{avg}\in R^{H\times W\times 1}$ and $F^s_{max}\in R^{H\times W\times 1}$, then concatenate them in the channel dimension and perform convolution operations on them using a convolutional kernel of size 7 (a large convolutional kernel is used here to extract more contextual information). Finally, after activation by the sigmoid activation function, the spatial attention weight map $M^s\in R^{H\times W\times 1}$ is obtained, and the spatial attention map is used to conduct an elementwise multiplication operation with the intermediate feature map to obtain the final feature map containing more plant morphological features map F'', where the spatial attention mechanism is calculated by the following formula:

$$M_s(F) = \sigma\Big(f^{7\times 7}([\mathrm{AvgPool}(F);\mathrm{MaxPool}(F)])\Big) \quad (2)$$

where σ denotes the sigmoid function and $f^{7\times 7}$ represents a convolution operation with a convolutional kernel size of 7.

2.3.2. Swin Transformer Block

A Swin transformer block is a new transformer model that includes sliding window theory, and the structure is shown in Figure 5. To reduce the computational effort and increase the computation of self-attention in locally focused regions, two consecutive Swin transformer blocks with different functions are used for each complete self-attention computation. One block is based on multihead self-attention (W-MSA) and one is based on shifted-window-based multihead self-attention (SW-MSA). The W-MSA module uses the window as the basic unit for the calculation of self-attention. This reduces the amount of calculation, but will lead to a lack of information exchange between adjacent windows; therefore, the feature map is then input into the SW-MSA module, which breaks the new window formed by the original window and performs the attention calculation within the newly formed window, thus breaking the blockage of information exchange between the original windows and enabling the exchange of information across windows.

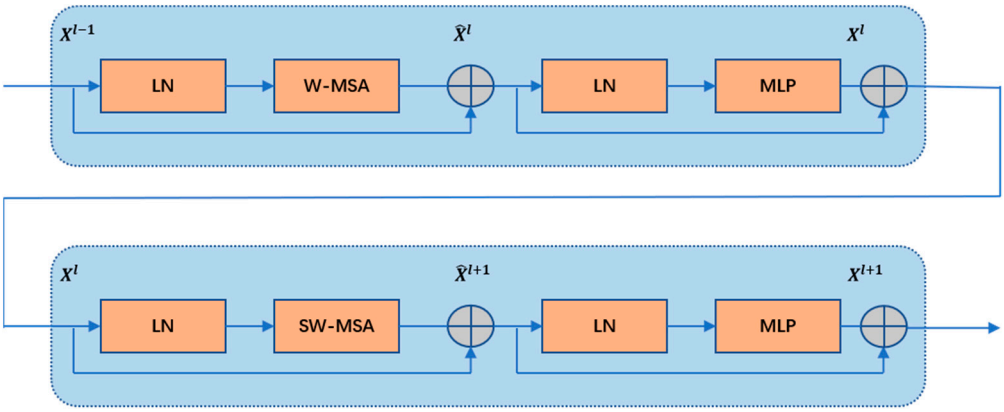


Figure 5. Structure diagram of the Swin transformer block.

In each Swin transformer block, the W-MSA module or SW-MSA module is followed by a multilayer perceptron, and the feature maps are normalized with a LayerNorm operation before being processed by the MSA and multilayer perceptron to prevent gradient disappearance. A jump connection is used in each Swin transformer block, and the two consecutive Swin transformer blocks are calculated as follows:

$$\hat{X}^l = W - \mathrm{MSA}\left(\mathrm{LN}\left(X^{l-1}\right)\right) + X^{l-1} \tag{3}$$

$$X^l = \mathrm{MLP}\left(\mathrm{LN}\left(\hat{X}^l\right)\right) + \hat{X}^l \tag{4}$$

$$\hat{X}^{l+1} = \mathrm{SW} - \mathrm{MSA}\left(\mathrm{LN}\left(X^l\right)\right) + X^l \tag{5}$$

$$X^{l+1} = \mathrm{MLP}\left(\mathrm{LN}\left(\hat{X}^{l+1}\right)\right) + \hat{X}^{l+1} \tag{6}$$

$\hat{X}^l$ and $\hat{X}^{l+1}$ represent the calculation results of W-MSA and SW-MSA, X^l and X^{l+1} represent the calculation results of MLP, LN is the LayerNorm layer, MLP is a fully connected layer after one cycle to complete one intrawindow and interwindow attention calculation.

3. Experiments and Analysis of Results

3.1. Model Training

The hardware environment was an Intel(R) Xeon(R) Gold 6246R CPU,16 GB memory, and GPU was NVIDIA Quadro RTX 8000 GPU with an NVIDIA Quadro RTX 8000 dedicated graphics card with 48 GB of graphics memory. The software environment was Windows 10, Python version 3.8.13, PyTorch version 1.7.1, and CUDA version 11.3.

The slender leaves of grassy weeds occupy a small area in the pictures, which can lead to problems of positive and negative sample imbalance. Therefore, the cross-entropy loss function was used to calculate the loss in this experiment, and its calculation formula is as follows. To prevent the network parameters from falling into local minima during the training process, a stochastic gradient descent algorithm was used, with the initial learning rate set to 0.0001, the learning rate decay factor to 0.001, the batch size set to 16, and the number of training iterations to 100×10^3 Iter times (where one Iter is equal to one training using batch size samples).

$$Cross\_entropy = -\frac{1}{N}\sum_i \sum_{c=1}^{M} y_{ic} \log(p_{ic}) \tag{7}$$

where M is the number of categories; y_{ic} is the sign function, taking 1 if the true category of sample i is equal to c and 0 otherwise; and p_{ic} is the predicted probability that the observed sample i belongs to category c.

3.2. Evaluation Indicators

In this study, four common metrics for semantic segmentation are used to evaluate the accuracy of the model segmentation results: the mean intersection over union (mIoU), accuracy rate (Acc), precision (Pr), recall (Re), and segmentation time (ST, i.e., the time taken by the model to segment a single image in the test set). The formulae for several indicators are as follows:

$$\mathrm{mIoU} = \frac{\sum \mathrm{TP}}{\sum \mathrm{TP} + \sum \mathrm{FN} + \sum \mathrm{FP}} \times 100\% \tag{8}$$

$$\mathrm{Acc} = \frac{\sum \mathrm{TP} + \sum \mathrm{TN}}{\sum \mathrm{TP} + \sum \mathrm{TN} + \sum \mathrm{FN} + \sum \mathrm{FP}} \times 100\% \tag{9}$$

$$\mathrm{Pr} = \frac{\sum \mathrm{TP}}{\sum \mathrm{TP} + \sum \mathrm{FP}} \times 100\% \tag{10}$$

$$\mathrm{Re} = \frac{\sum \mathrm{TP}}{\sum \mathrm{TP} + \sum \mathrm{FN}} \times 100\% \tag{11}$$

TP is true positive; TN is true negative; FP is false positive; FN is false negative.

3.3. Analysis of Results

3.3.1. Ablation Experiments

To explore the impact of different improvement points on the performance of the model, we compared the original DeepLabv3+ model (with a ResNet50 backbone), the DeepLabv3+ model with a modified backbone (with the backbone replaced by a Swin transformer with a similar number of parameters as ResNet50), and Swin-DeepLab (which uses a Swin transformer as the backbone and adds different dimensional feature extraction and a CBAM to the decoding part) for image recognition. The results of the split are shown in Figure 6.

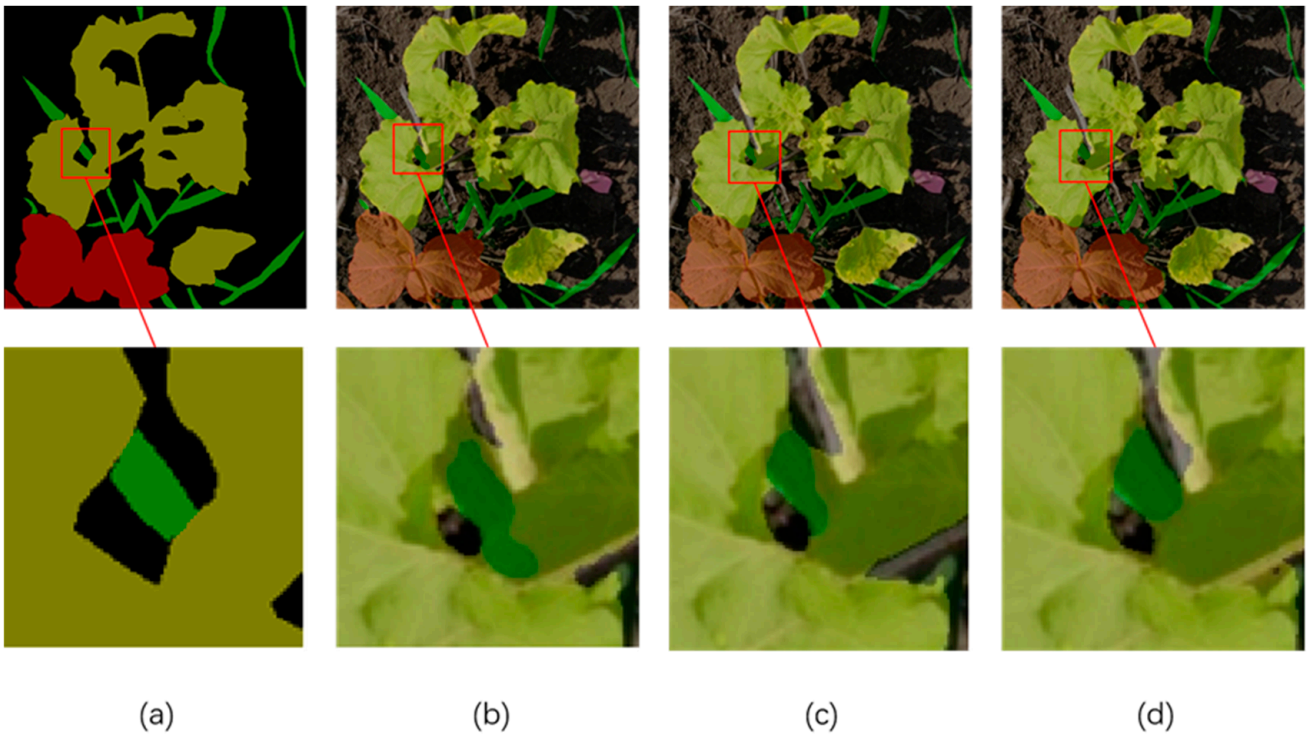


Figure 6. Recognition results of the model adding different improvement points: (**a**) labeled images; (**b**) DeepLabv3+ model (backbone is ResNet50); (**c**) DeepLabv3+ model with backbone modification only (backbone is Swin); (**d**) Swin-DeepLab.

In the original DeepLabv3+ model, a small number of graminoid weed pixels sandwiched between broadleaf weeds were not correctly segmented, resulting in the misclassification of neighboring broadleaf weeds. After replacing the model backbone in DeepLabv3+, the misclassification of broadleaf weeds was improved, but the graminoid weeds were still misclassified. In contrast, after adding different dimensional feature extraction and incorporating the CBAM, the contour boundaries of both weeds were better segmented, and the best results were obtained. The segmentation accuracy of the three models is shown in Table 1.

Table 1. Performance comparison of the models with different improvement points added.

Model	mIoU (%)	Acc (%)	Pr (%)	Re (%)	ST (ms)
DeepLabv3+	88.59	93.40	94.32	93.10	319
Swin+DeepLabv3+	91.10	95.15	95.75	94.33	341
Swin-DeepLab	**91.53**	**95.48**	**95.92**	**94.58**	367

3.3.2. Comparison of Swin-DeepLab with Other Algorithms

To verify the advantages of the Swin-DeepLab model for weed detection in soybean fields, the classical semantic segmentation models U-Net, DeepLabv3+, and PSPNet as well as the recently proposed semantic segmentation models UPerNet, CCNet, and OCRNet were selected for comparison experiments and compared with the results of the Swin-DeepLab model using the same experimental environment and data preprocessing. First, the loss curves of the models were compared, and the results are shown in Figure 7. The loss of the training set of the Swin-DeepLab model reached convergence after 50×10^3 iterations, and the loss of the test set converged relatively slowly, after 80×10^3 iterations. After reaching convergence, the loss of the validation set of the Swin-DeepLab model was the smallest. Next, the prediction result accuracy of the models was compared, as shown in Table 2. Swin-DeepLab model obtained the best results in several major metrics of semantic segmentation, reaching 91.53% for mIoU, 95.48% for Acc, 95.92% for Pr, and 94.58% for Re, surpassing the results obtained using the original model DeepLabv3+ and other comparison models. Swin-DeepLab achieved the highest segmentation accuracy at a slightly higher ST.

To investigate the effectiveness of the Swin-DeepLab model for densely distributed weed detection, the image detection results were analyzed, and the results are described below for different levels of weed distribution complexity. First, the comparative segmentation results under the weed sparse distribution condition are shown in Figure 8. In this image, the two weeds are somewhat spaced apart, not shaded by each other, and the whole plant features are more complete. In the segmentation results, the PSPNet and U-Net models still had some shortcomings in the recognition accuracy of broadleaf weeds. The Swin-DeepLab, DeepLabv3+, OCRNet and UPerNet models all basically achieved the pixel-level semantic segmentation of weeds. This figure shown that most of the evaluated segmentation algorithms could achieve good results under the conditions of sparse weed distribution.

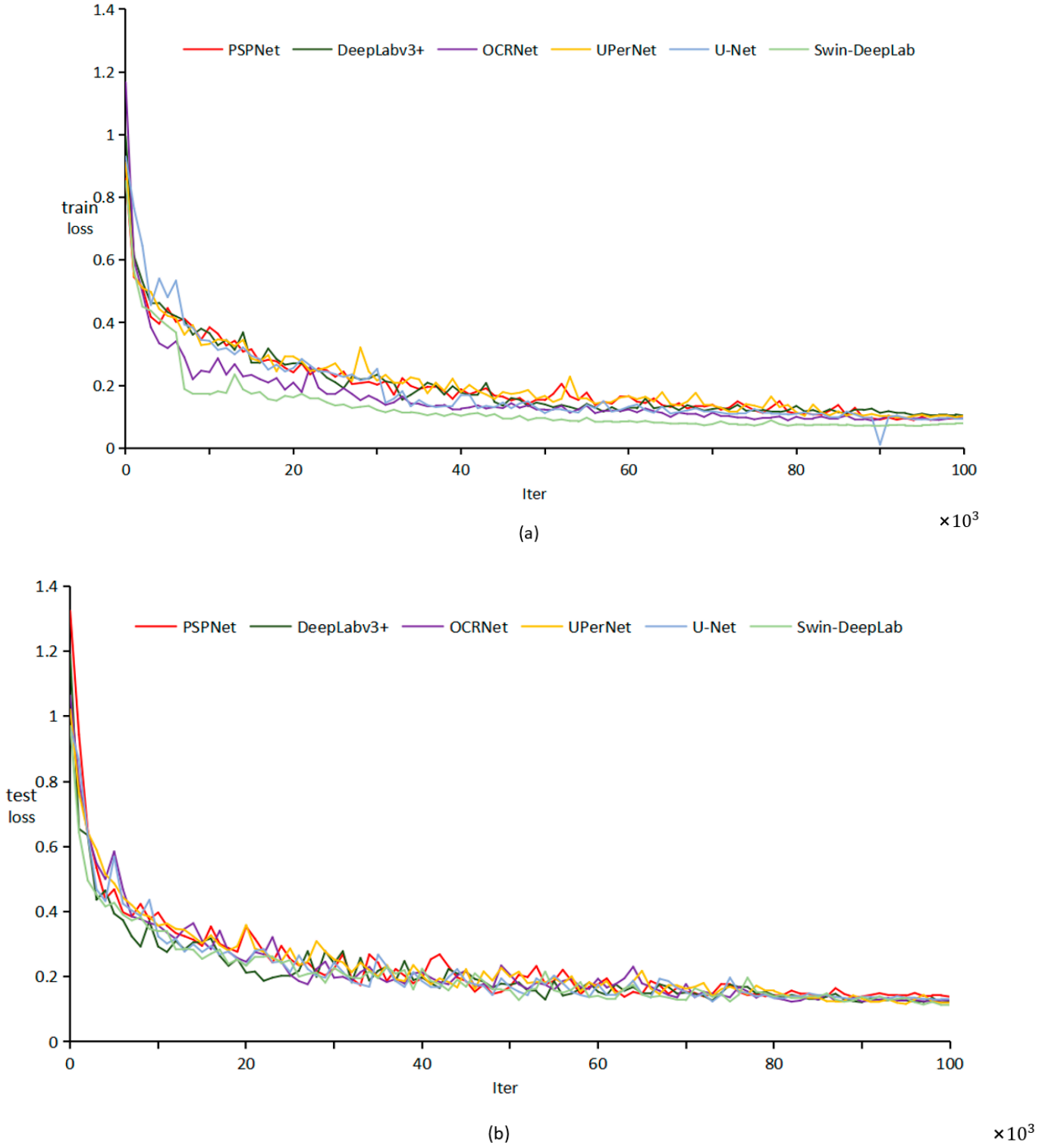


Figure 7. Loss variation plots: (**a**) loss plots for the training set; (**b**) loss plots for the test set.

Table 2. Comparison of performance metrics of the different models.

Model	mIoU (%)	Acc (%)	Pr (%)	Re (%)	ST (ms)
U-Net	86.61	92.65	93.58	92.63	356
PSPNet	88.18	93.01	94.11	93.36	301
DeepLabv3+	88.59	93.40	94.32	93.10	319
UPerNet	88.75	92.87	95.14	92.72	338
CCNet	87.62	93.17	93.20	93.19	340
OCRNet	88.73	93.15	94.69	93.13	335
Swin-DeepLab	**91.53**	**95.48**	**95.92**	**94.58**	367

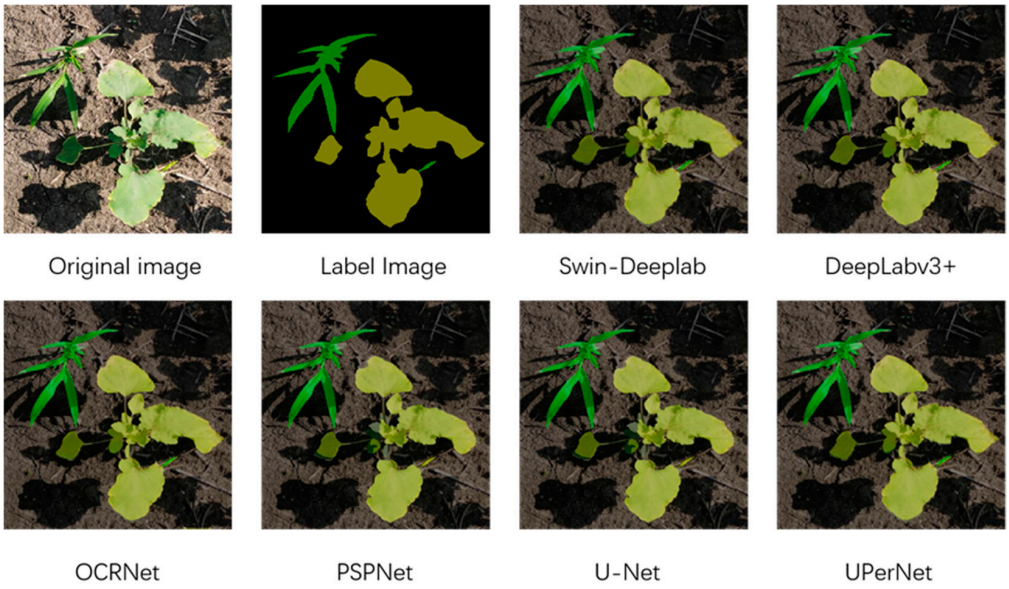


Figure 8. Comparison of segmentation results of different models under sparse weed distribution.

The segmentation results were then compared for the more complex spatial distribution of weeds. In this case, the soybean and weed plants were close to each other, the leaves shaded each other, the shape information of the plants was incomplete, and the segmentation accuracy decreased. The segmentation effect of different models is shown in Figure 9. PSPNet, U-Net, and DeepLabv3+ models all failed to identify weeds due to insufficient information about their leaf features. Due to the small range of spatial information extracted by the convolution operation, it is highly likely that different species of plants in proximity are regarded as the same species. In both the OCRNet model and the UPerNet model, a portion of the middle graminoid weeds was identified as the soybean leaf that surrounded it. Using the Swin-DeepLab model with the Swin transformer block as the backbone feature extraction network instead of the convolution operation, the model extracted more distant feature information, resulting in better segmentation of the overlapping portion. The use of the CBAM and the use of more intermediate feature maps for feature recovery in the decoding process resulted in more accurate leaf contours in the reduced result maps. Given the above analysis, it can be concluded that Swin-DeepLab provides better segmentation results in the case of a more complex spatial distribution of weeds.

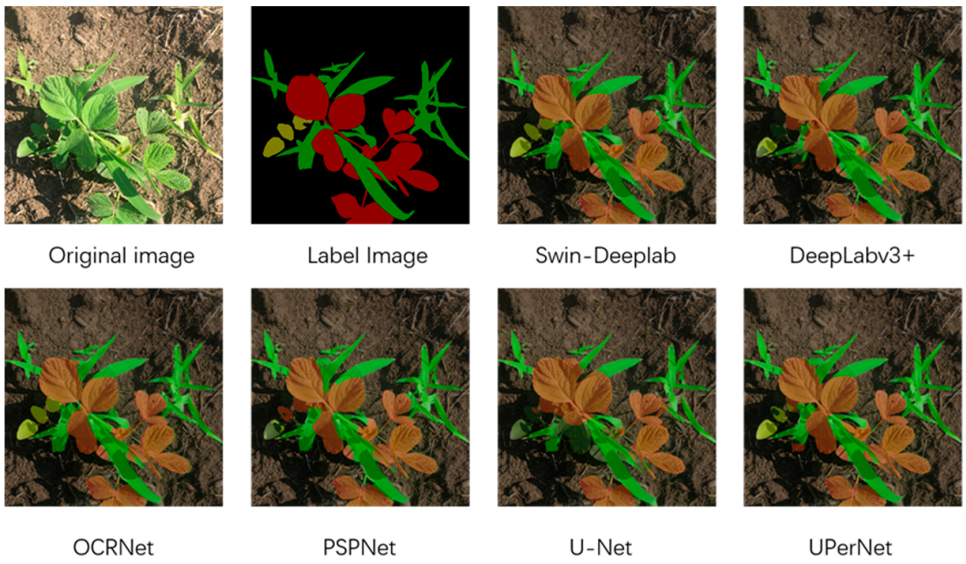


Figure 9. Comparison of segmentation results of different models under a dense distribution of weeds.

4. Discussion

In this paper, in order to solve the problem of dense distribution of weeds commonly found in actual production, we first collected a dataset of weed images from soybean fields which were collected and preprocessed to simulate different real-world conditions, then proposed a new segmentation model. Experiments were conducted using weed pictures. The model is based on the DeepLabv3+ model and replaces its original visual backbone Xception with a Swin transformer to enhance the model's key information extraction capability, which expands the perceptual field without losing the advantages of translation invariance and hierarchy of convolution operation, extracts different levels of feature maps in the decoding section, and introduces a CBAM to enhance the effect of feature recovery. This expands the perceptual field without losing the advantages of translation invariance and hierarchy of convolution operation, extracts different levels of feature maps in the decoding section, and then the CBAM attention mechanism module is used for all feature maps of different sizes so that it increases the weight of the focused semantic channels and the utilization of spatial information by the model, which effectively improves the segmentation accuracy at the junction of target contours and the segmentation accuracy in the densely distributed weed area. In the experiments, the proposed Swin-DeepLab model achieved a 91.53% mIoU and a 95.48% Acc on a complexly distributed soybean weed dataset, which is 2.94% and 2.08% better than the original DeepLabv3+ model, respectively. The new model had a detection speed of 367 ms for 512 $\times$ 512 images, which can meet the needs of practical production. This study demonstrates the capability of a vision transformer structure in weed identification, and provides direction for further applications of a vision transformer in weed identification tasks. Further research will be conducted on the application of a vision transformer in the field of crop identification, applying a vision transformer to different crop weed datasets and further improving the model to enhance the accuracy of weed segmentation, providing directions for further applications in precision agriculture.

Author Contributions: H.Y. (Helong Yu), M.C., H.Y. (Han Yu) and J.Z. conceived and designed the manuscript. H.Y. (Helong Yu) and M.C. analyzed the data. H.Y. (Helong Yu), H.Y. (Han Yu) and J.Z. wrote the paper. H.Y. (Han Yu) and J.Z. revised the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number 42001112) and the Science and Technology Development Program of Jilin Province (grant number 20200301047RQ).

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. Ramirez, W.; Achanccaray, P.; Mendoza, L.; Pacheco, M. Deep convolutional neural networks for weed detection in agricultural crops using optical aerial images. In Proceedings of the 2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS), Santiago, Chile, 21–26 March 2020; pp. 133–137.
2. Khan, S.; Tufail, M.; Khan, M.T.; Khan, Z.A.; Anwar, S. Deep learning-based identification system of weeds and crops in strawberry and pea fields for a precision agriculture sprayer. *Precis. Agric.* **2021**, *22*, 1711–1727. [CrossRef]
3. Olaniyi, O.M.; Daniya, E.; Kolo, J.G.; Bala, J.A.; Olanrewaju, A. A computer vision-based weed control system for low-land rice precision farming. *Int. J. Adv. Appl. Sci.* **2020**, *9*, 51–61. [CrossRef]
4. Brilhador, A.; Gutoski, M.; Hattori, L.T.; de Souza Inácio, A.; Lazzaretti, A.E.; Lopes, H.S. Classification of weeds and crops at the pixel-level using convolutional neural networks and data augmentation. In Proceedings of the 2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Guayaquil, Ecuador, 11–15 November 2019; pp. 1–6.
5. Gao, J.; Liao, W.; Nuyttens, D.; Lootens, P.; Vangeyte, J.; Pižurica, A.; He, Y.; Pieters, J.G. Fusion of pixel and object-based features for weed mapping using unmanned aerial vehicle imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *67*, 43–53. [CrossRef]
6. Wu, Z.; Chen, Y.; Zhao, B.; Kang, X.; Ding, Y.J.S. Review of weed detection methods based on computer vision. *Sensors* **2021**, *21*, 3647. [CrossRef]
7. Liu, B.; Bruch, R.J.C.R.R. Weed detection for selective spraying: A review. *Curr. Robot. Rep.* **2020**, *1*, 19–26. [CrossRef]
8. Wu, X.; Aravecchia, S.; Lottes, P.; Stachniss, C.; Pradalier, C. Robotic weed control using automated weed and crop classification. *J. Field Robot.* **2020**, *37*, 322–340. [CrossRef]
9. García-Santillán, I.D.; Pajares, G. On-line crop/weed discrimination through the Mahalanobis distance from images in maize fields. *Biosyst. Eng.* **2018**, *166*, 28–43. [CrossRef]
10. Bargoti, S.; Underwood, J.P. Image segmentation for fruit detection and yield estimation in apple orchards. *J. Field Robot.* **2017**, *34*, 1039–1060. [CrossRef]
11. Quan, L.; Wu, B.; Mao, S.; Yang, C.; Li, H. An Instance Segmentation-Based Method to Obtain the Leaf Age and Plant Centre of Weeds in Complex Field Environments. *Sensors* **2021**, *21*, 3389. [CrossRef]
12. Deng, X.; Qi, L.; Ma, X.; Jiang, Y.; Chen, X.; Liu, H.; Chen, W. Recognition of weeds at seedling stage in paddy fields using multi-feature fusion and deep belief networks. *Trans. Chin. Soc. Agric. Eng.* **2018**, *34*, 165–172.
13. Ashraf, T.; Khan, Y.N. Weed density classification in rice crop using computer vision. *Comput. Electron. Agric.* **2020**, *175*, 105590. [CrossRef]
14. Le, V.N.T.; Apopei, B.; Alameh, K. Effective plant discrimination based on the combination of local binary pattern operators and multiclass support vector machine methods. *Inf. Process. Agric.* **2019**, *6*, 116–131.
15. Wendel, A.; Underwood, J. Self-supervised weed detection in vegetable crops using ground based hyperspectral imaging. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 5128–5135.
16. Hasan, A.M.; Sohel, F.; Diepeveen, D.; Laga, H.; Jones, M.G. A survey of deep learning techniques for weed detection from images. *Comput. Electron. Agric.* **2021**, *184*, 106067. [CrossRef]
17. Huang, H.; Deng, J.; Lan, Y.; Yang, A.; Deng, X.; Zhang, L. A fully convolutional network for weed mapping of unmanned aerial vehicle (UAV) imagery. *PLoS ONE* **2018**, *13*, e0196302. [CrossRef] [PubMed]
18. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2012**, *25*, 84–90. [CrossRef]
19. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556v6.
20. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
21. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
22. Fawakherji, M.; Youssef, A.; Bloisi, D.; Pretto, A.; Nardi, D. Crop and weeds classification for precision agriculture using context-independent pixel-wise segmentation. In Proceedings of the 2019 Third IEEE International Conference on Robotic Computing (IRC), Naples, Italy, 25–27 February 2019; pp. 146–152.
23. 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.

24. Lottes, P.; Hoeferlin, M.; Sander, S.; Müter, M.; Schulze, P.; Stachniss, L.C. An effective classification system for separating sugar beets and weeds for precision farming applications. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 5157–5163.
25. Li, N.; Zhang, X.; Zhang, C.; Guo, H.; Sun, Z.; Wu, X. Real-time crop recognition in transplanted fields with prominent weed growth: A visual-attention-based approach. *IEEE Access* **2019**, *7*, 185310–185321. [CrossRef]
26. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
27. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
28. Wu, Z.; Yang, R.; Gao, F.; Wang, W.; Fu, L.; Li, R. Segmentation of abnormal leaves of hydroponic lettuce based on DeepLabV3+ for robotic sorting. *Comput. Electron. Agric.* **2021**, *190*, 106443. [CrossRef]
29. Kang, J.; Liu, L.; Zhang, F.; Shen, C.; Wang, N.; Shao, L. Semantic segmentation model of cotton roots in-situ image based on attention mechanism. *Comput. Electron. Agric.* **2021**, *189*, 106370. [CrossRef]
30. Chen, L.-C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
31. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
32. 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.
33. 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, 18–22 June 2018; pp. 7132–7141.

agronomy

Article

A Faster R-CNN-Based Model for the Identification of Weed Seedling

Ye Mu [1,2,3,4], Ruilong Feng [1], Ruiwen Ni [1], Ji Li [1], Tianye Luo [1], Tonghe Liu [1], Xue Li [1], He Gong [1,2,3,4], Ying Guo [1,2,3,4], Yu Sun [1,2,3,4], Yu Bao [5], Shijun Li [6,7], Yingkai Wang [8] and Tianli Hu [1,2,3,4,*]

[1] College of Information Technology, Jilin Agricultural University, Changchun 130118, China
[2] Jilin Province Agricultural Internet of Things Technology Collaborative Innovation Center, Changchun 130118, China
[3] Jilin Province Intelligent Environmental Engineering Research Center, Changchun 130118, China
[4] Jilin Province Colleges and Universities and the 13th Five-Year Engineering Research Center, Changchun 130118, China
[5] School of Management, Changchun University, Changchun 130022, China
[6] College of Information Technology, Wuzhou University, Wuzhou 543003, China
[7] Guangxi Key Laboratory of Machine Vision and Inteligent Control, Wuzhou 543003, China
[8] College of Agriculture, Jilin Agricultural University, Changchun 130118, China
* Correspondence: hutianli@jlau.edu.cn

Abstract: The accurate and rapid acquisition of crop and weed information is an important prerequisite for automated weeding operations. This paper proposes the application of a network model based on Faster R-CNN for weed identification in images of cropping areas. The feature pyramid network (FPN) algorithm is integrated into the Faster R-CNN network to improve recognition accuracy. The Faster R-CNN deep learning network model is used to share convolution features, and the ResNeXt network is fused with FPN for feature extractions. Tests using >3000 images for training and >1000 images for testing demonstrate a recognition accuracy of >95%. The proposed method can effectively detect weeds in images with complex backgrounds taken in the field, thereby facilitating accurate automated weed control systems.

Keywords: weed identification; Faster-R-CNN; FPN; ResNeXt

Citation: Mu, Y.; Feng, R.; Ni, R.; Li, J.; Luo, T.; Liu, T.; Li, X.; Gong, H.; Guo, Y.; Sun, Y.; et al. A Faster R-CNN-Based Model for the Identification of Weed Seedling. *Agronomy* **2022**, *12*, 2867. https://doi.org/10.3390/agronomy12112867

Academic Editors: Jian Zhang, Randy G. Goebel and Zhihai Wu

Received: 6 October 2022
Accepted: 11 November 2022
Published: 16 November 2022

1. Introduction

Weeds include all kinds of herbaceous plants that grow where they are not wanted [1,2]. Weeds cause much harm to the agricultural economy [3]. Not only do they compete with crops for sunlight, water and fertilizer, but they also compete for living space. If not dealt with in time, weeds will reduce crop growth, yield and quality, and can even cause crop failure. With the development of computer technology, rapid and accurate machine vision-based recognition technology has been increasingly used for weed identification [4–7]. Deep learning has achieved good results in the recognition of human behaviour, crop fruits and weeds. The image recognition method of kiwifruit based on a convolutional neural network (CNN) shows that CNNs have good application prospects in field fruit recognition. Recognition of corn weeds based on CNNs, hash codes and multi-scale hierarchical features has proven the effectiveness of CNNs in recognising weeds in field images. In a study on broccoli seedlings, a crop detection method based on the Faster R-CNN model was proposed. Using a dropout value of 0.6, the ResNet101 network was used as the feature extraction network, and an average precision of 91.73% was achieved. At present, the Faster R-CNN model is also widely applied in the field of vehicle detection [8], ground object recognition in remote sensing images [9,10], appearance defect detection [11], pedestrian detection and recognition [12], and field image detection [13], and has excellent recognition accuracy.

Since most current models have the shortcomings of high complexity and difficulty in modification, an improved Faster R-CNN field weed detection model is proposed in this study. This network model was fused with the Feature Pyramid Network (FPN) to improve the detection accuracy of the model in weed identification. On the basis of obtaining a large amount of weed and plant seedling data, a deep learning network is established, and the ResNeXt feature extraction network is used to generate target recognition models for different types of weeds. After training, it can effectively identify weeds and plant seedlings in images. In this way, a Faster R-CNN-based rapid weed identification system for use in the field is obtained, which can identify weeds in multiple crops.

2. Materials and Methods

2.1. Image Data Acquisition

Image data were obtained from the V2 Plant Seedlings Dataset [14]. This dataset consists of images of nine field weeds—scentless mayweed, common chickweed, shepherd's purse, cleavers, charlock, fat hen, small-flowered cranesbill, black-grass, and loose silky-bent—and three crop seedlings—maize, common wheat, and sugar beet (Figure 1). To ensure experimental accuracy, the images were collected under different conditions, including sunny, cloudy, and rainy days. There were 5539 images in total, including 598 of scentless mayweed, 713 of common chickweed, 274 of shepherd's purse, 335 of cleavers, 452 of charlock, 538 of fat hen, 576 of small-flowered cranesbill, 309 of black-grass, 762 of loose silky-bent, 257 of maize, 253 of common wheat, and 463 of sugar beet. The resolution of the images after processing was 227 × 227 pixels. All processed images were divided as follows: 3329 in a training set, 1107 in a validation set, and 1103 in a testing set. The pictures in these sets did not overlap each other. The training set was used to train the model parameters, the test set was used to evaluate the generalization error of the model applied to the samples after training, and the validation set was used to tune the hyperparameters of the model during the training process. After the experiment was completed, deep learning was used to apply evaluation indicators typically used in the field of object recognition to measure the performance of the detector model.

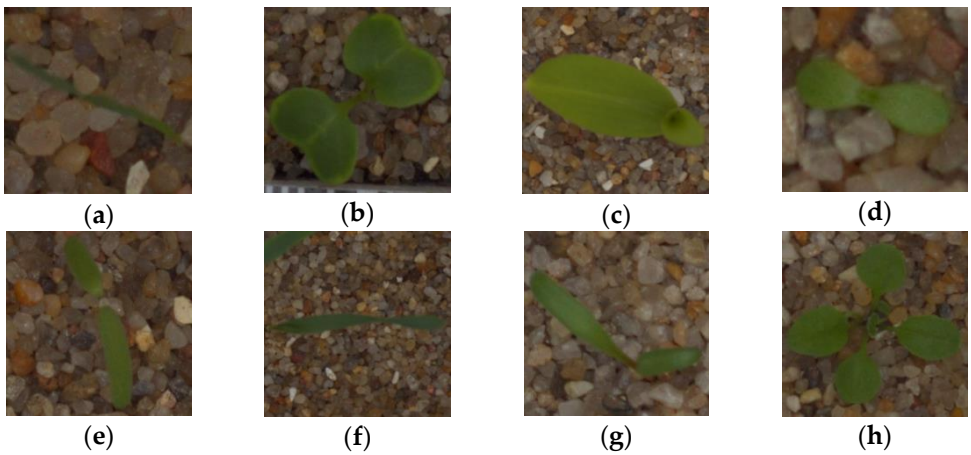


Figure 1. Examples of partial samples of (**a**) black-grass; (**b**) charlock; (**c**) maize; (**d**) scentless mayweed; (**e**) fat hen; (**f**) common wheat; (**g**) sugar beet; and (**h**) shepherd's purse.

This dataset is due to the use of public datasets, so there are some limitations. Because the pictures in the dataset exist as a single crop, they cannot be used for large-area crop identification between fields. Moreover, because the dataset collection site is under laboratory conditions, a large amount of sand and gravel is selected for plant cultivation, so the detection accuracy of soil-growing plants cannot be guaranteed.

2.2. Image Pre-Processing

Since the images were affected by the environmental conditions at the time of collection, extreme highlights and shadows could decrease the model's recognition success rate. Therefore, the images were processed and segmented; the green part of the plant was extracted and non-green areas were suppressed. The application of normalized colour components can effectively improve the effects of lighting and shadows on image quality [15,16]. After the greyscale images were obtained, the Otsu method [17,18] was used to convert them into binary images and separate the plants. After processing, clear plant images were obtained. They were then filled with noise to obtain single-leaf binary images after segmentation image processing, as shown in Figure 2.

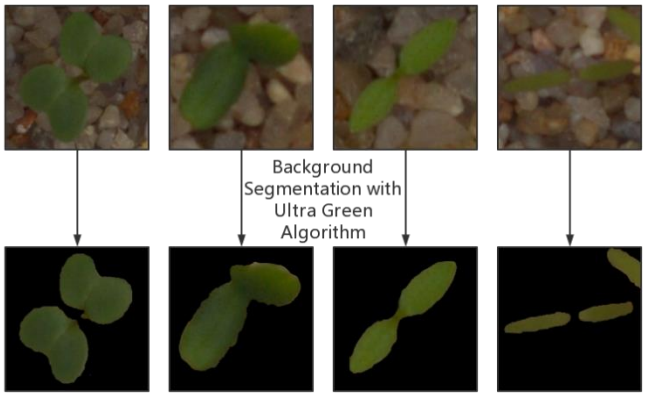


Figure 2. Example of single-leaf binary images after partial processing.

3. Model Refinement

Currently, there are two types of object detection models. The first is one complete object recognition and object positioning in two steps. Typical representatives of this approach are the R-CNN, Fast R-CNN, and Faster R-CNN families. This type of framework has a low recognition error rate and a low omission rate and can be used in real-time detection scenarios. The typical representative of the second category are YOLO (you only look once), SSD (Single Shot Multibox Detector), YOLOv2, and YOLOv3. To complete object classification and object localization in one step [19–21], SSD directly regresses the position and category of the target at the output layer. Although these methods have fast recognition speed, their accuracy rate is lower than that of Faster R-CNN. Therefore, this paper selected the Faster R-CNN model framework to identify weed images.

The network structure of Faster R-CNN is shown in Figure 3. The network can be roughly divided into four parts: (1) a feature extraction layer, (2) a Region Proposal Network (RPN), (3) a Region of Interest pooling (ROI pooling) layer, and (4) classification and regression.

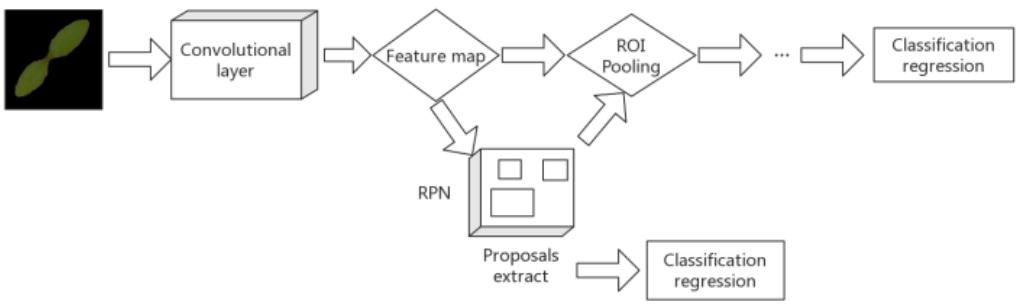


Figure 3. The Faster R-CNN network structure.

The deep network in the convolutional network is used to respond to the semantics, and the shallow layer responds to the image. However, in object detection, because the

feature map size is too small, the high-level network can respond to the semantic features but is not conducive to object detection [22]. Therefore, a Feature Pyramid Network (FPN) is introduced to improve the weed detection accuracy of the algorithm in the field identification process. The structure of the FPN network is shown in Figure 4.

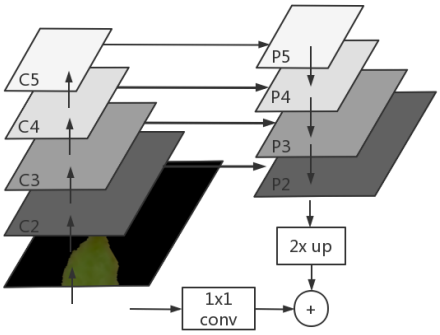


Figure 4. The FPN network structure diagram.

Bottom-to-top process: It can be seen that the bottom-to-top extraction process is a conventional feature extraction process. The output of the last layer of each stage is selected as the input of the feature map. The first layer is not used here due to its large memory footprint.

Top-down process: Upsampling starting from the highest layer yields better feature maps, and nearest-neighbour 2× upsampling is used for simplicity.

Horizontal connection process: Each horizontal connection fuses feature maps of the same spatial size, specifically, the upsampling result and the bottom-up C2, C3, C4, and C5 layers are fused with the same-sized feature maps generated by 1×1 convolution. The output channels are all set to the same 256 channels.

The ResNeXt target extraction network and FPN are selected for fusion in the Faster R-CNN network. This is used to extract target features in target detection so that the learning of the target features is more complete.

The building blocks of ResNet and ResNeXt are shown in Figure 5. While the ResNeXt network structure retains the basic stacking method of ResNet, it splits its single path into 32 independent paths, which perform convolution operations on the input image at the same time. Finally, the cumulative summation of outputs from different paths is used as the final result. This operation makes the division of labour of the network clearer and the local adaptability stronger. Since each path shares the same topology and convolution parameters, and the design method is the same, the network parameters will not increase, which is convenient for model transplantation.

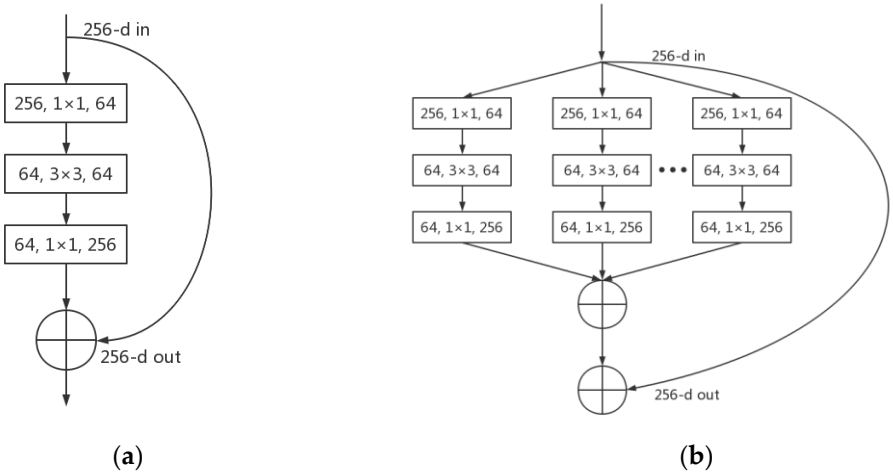


Figure 5. Building blocks of (**a**) ResNet and (**b**) ResNeXt.

The improved model is shown in Figure 6. At present, there are mainly two training methods for Faster R-CNN: alternating training and approximate joint training. The alternate training method trains two networks—RPN and Fast-RCNN—in a total of two stages, with each stage training the RPN and Fast-RCNN once. In the approximate joint training process, only one weighted network is trained, which requires slightly less memory. This training method saves 25–50% of the training time compared with the alternate training method, and the two training methods have similar accuracy, so the approximate joint training method was selected.

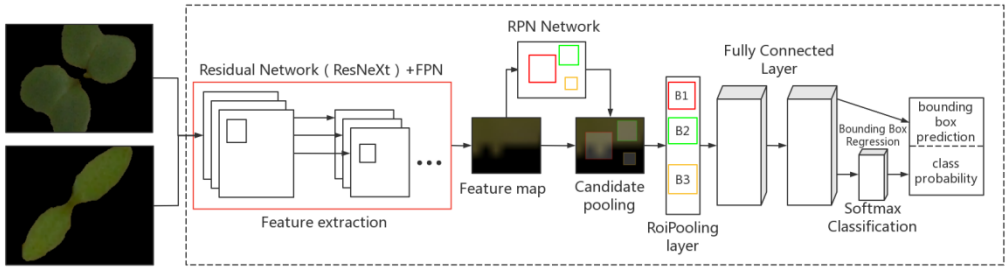


Figure 6. Improved model plots.

First, the weed image is input into the model and the improved ResNeXt network is used to fuse the FPN network to extract features from the image. The overall design of FPN is presented in a multi-scale pyramid structure, and each layer of FPN corresponds to {P2, P3, P4, P5, P6} in ResNeXt pyramid through anchors, as shown in Figure 7. Using 3 proportions {1:2, 1:1, 2:1}, 15 types of anchors were used to predict the target objects in weed images in the field. After the feature map is obtained, it is input into the RPN network and propagated forward to obtain a higher-dimensional feature map. When the feature map is passed to the RPN network, the proposed box is obtained and a non-maximum suppression value operation is performed on the proposed box. The top N highest-scoring proposal boxes are used as ROIs. The feature map and ROIs are passed to the ROI pooling layer for a pooling operation. Then, the ROI pooling layer is output to the full link layer and linear regression is performed on each area of interest to obtain accurate detection of weeds. Then, the Softmax regression model is used to perform multi-classification target detection and, finally, the categories of weeds and crop seedlings in different fields can be identified.

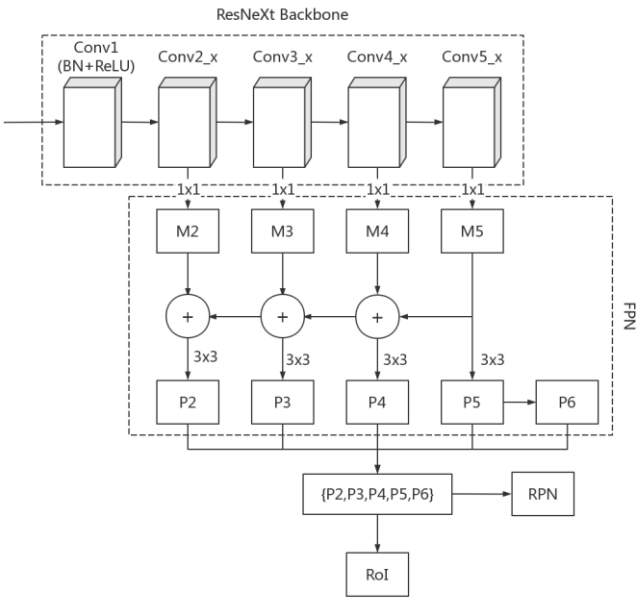


Figure 7. Schematic representation of the ResNeXt + FPN structure.

4. Experimental Results and Analysis

4.1. Test Platform

The operating environment for the test was a desktop computer with a Windows 10 64-bit operating system, and the computer memory was a 32 GB, NVIDIA GeForce RTX 3070 graphics card, equipped with an AMD Ryzen 7 5800H processor.

The software environments used are Anaconda 3.9.12 (Developed by Anaconda, Inc. in Austin, TX, USA), Python 3.7.9, CUDA 11.2 (Developed by NVIDIA Corporation of California, USA), and cuDNN 8.2.1 (Developed by NVIDIA Corporation of California, USA). The open-source deep learning framework Tensorflow 2.0 (Developed by Google, Inc. in California, USA) was used as the development environment.

4.2. Parameter Design

To improve the performance of the model and reduce overfitting, the pre-trained model was used to initialize the parameters and the stochastic gradient descent (SGD) method was used to improve the model. The learning rate was set to 0.001, the momentum factor was set to 0.9, the epoch was set to 1500, the maximum number of iterations was 200,000, and the learning rate was adjusted to 0.0001 after 80,000 iterations and to 0.00001 after 160,000 iterations. After the model calculation was completed, the model with the highest accuracy was selected for weed identification. The non-maximum suppression (NMS) value was set to 0.3 to obtain the best candidate box. Finally, to obtain the trained network model, the test set was used to further verify the modelling effect, and the recognition result was the output.

4.3. Evaluating Indicator

In this paper, two evaluation indexes, accuracy (P) and recall (R), were used to verify whether the model can be used for image recognition of weed seedlings. The range of both is [0, 1]. In addition, the F_1 value was used to harmonize the average evaluation of the calculation results [23], and the Mean Intersection over Union (MIoU) was used to evaluate the image segmentation results. The assessment is calculated as follows:

$$P = \frac{TP}{TP + FP} \times 100\%, \tag{1}$$

$$R = \frac{TP}{TP + FN} \times 100\%, \tag{2}$$

$$F_1 = \frac{2PR}{P + R} \times 100\%, \tag{3}$$

$$MIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{TP}{FN + FP + TP} \tag{4}$$

where P stands for accuracy for weed image recognition; R is a recall for weed image recognition; F_1 is the harmonic mean of P and R; TP is the true positive rate (number of correctly identified crop seedlings and weed targets); FP is the number of incorrectly identified crop seedlings and weed targets; FN is the number of unidentified crop seedlings and weed targets; k is how many categories there are in the dataset.

5. Results and Analysis

5.1. Model Training Results

According to the above experimental methods, a weed recognition network based on Faster R-CNN fused with FPN was trained and the improved ResNeXt-101 feature extraction network was adopted. The loss degree and accuracy of this model in recognition are shown in Figures 8 and 9.

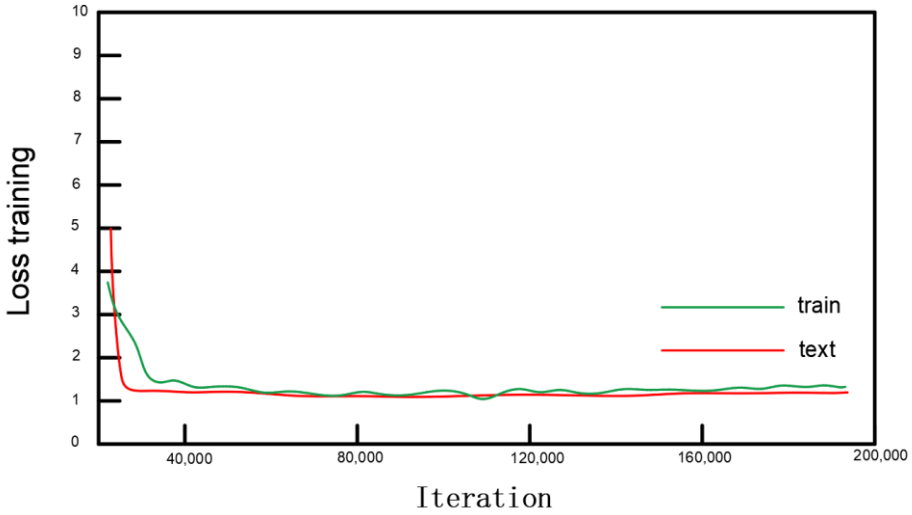


Figure 8. Loss of the Faster R-CNN model incorporating FPN.

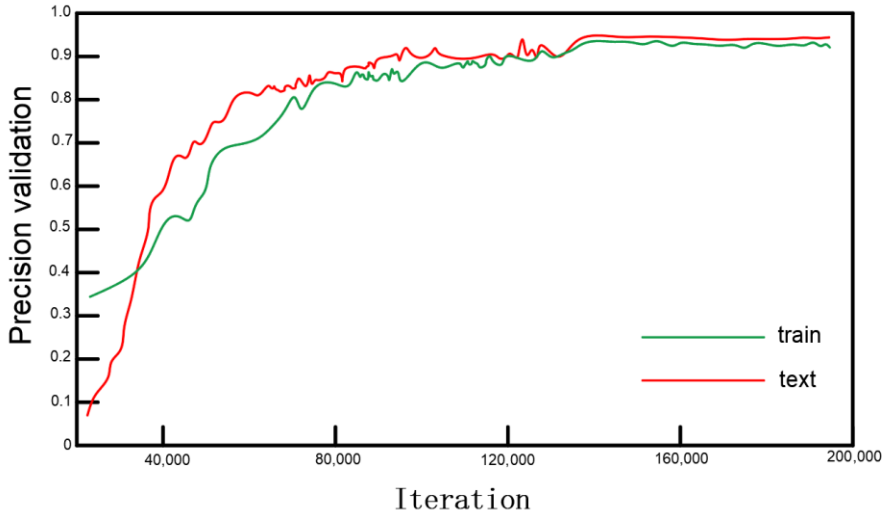


Figure 9. Precision of the Faster R-CNN model incorporating FPN.

5.2. Experimental Results

It can be seen from Figures 8 and 9 that the weed recognition network based on Faster R-CNN integrated with FPN has high recognition accuracy when using the ResNeXt-101 feature extraction network. To select a better deep network model and a suitable feature extraction network, this experiment compares the Faster R-CNN integrated with FPN and the ordinary Faster R-CNN network under the condition of the ResNeXt-101 feature extraction network. The results are shown in Table 1.

Table 1. Performance comparison of the faster R-CNN-FPN network and ordinary Faster R-CNN.

Model (With ResNeXt-101)	Accuracy (%)	Recall (%)	F_1-Value (%)	MIoU (%)	Detection Time (ms)
Faster R-CNN-FPN	95.61	87.26	91.24	93.7	330
Faster R-CNN	92.4	85.2	88.65	89.6	319

It can be seen from Table 1 that Faster R-CNN has a higher accuracy rate when it does not integrate the FPN and only uses the ResNeXt-101 feature extraction network. However, with integrated FPN, the accuracy is higher than that of the ordinary Faster R-CNN network, and the F_1-value is higher; in addition, the value of MIOU is about 4% higher than the

unimproved model. It can be seen that the FPN can optimize the network after being integrated into the Faster R-CNN network, which improves the recognition accuracy.

With the Faster R-CNN deep network model integrated with FPN, compared with the ResNet-50 feature extraction network in the literature [22] that proposed FPN, ResNet-50 and ResNeXt-101 were used for training. With increases in the number of model training iterations, the real-time losses of the overall loss functions of the two feature extraction networks used in the Faster R-CNN deep network model were compared (Figure 10 and Table 2).

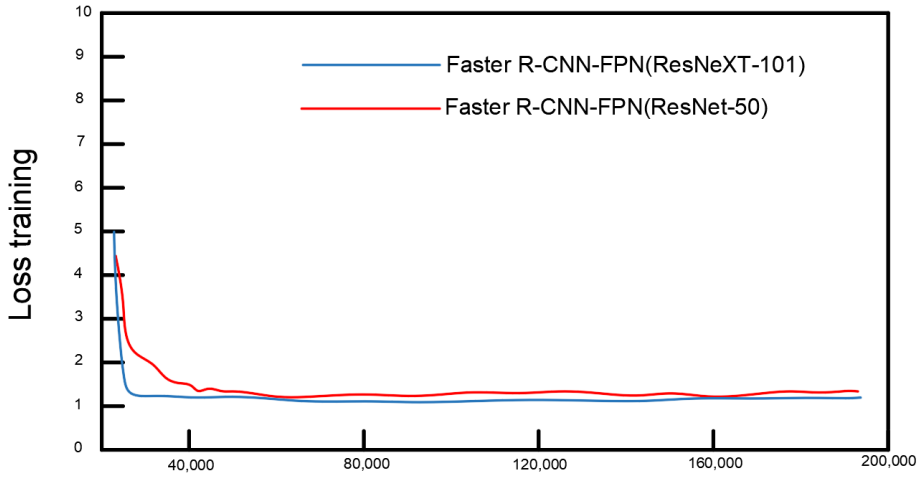


Figure 10. Overall loss comparison with the Faster R-CNN model.

Table 2. Feature extraction network comparison.

Model	Extraction Network	Accuracy (%)	Recall (%)	F_1-Value (%)	MIoU (%)	Detection Time (ms)
Faster R-CNN-FPN	ResNeXt-101	95.61	87.26	91.24	93.7	330
Faster R-CNN-FPN	ResNet-50	91.67	85.19	88.31	89.5	319

Figure 10 shows the losses of the overall loss functions of the two feature extraction networks with the Faster R-CNN deep network model. The Faster R-CNN deep network model based on the ResNeXt-101 feature extraction network can converge faster and is obviously better than the ResNet-50 feature extraction network model. The Faster R-CNN deep network model of the ResNet-50 feature extraction network oscillated to varying degrees after 120,000 iterations. The comparison in Table 2 shows that although the ResNeXt-101 feature extraction network uses 101 layers, there is basically no difference between the operation time, but the F_1-value and MIoU value of ResNeXt-101 as the result of the feature extraction network are higher than those of the ResNet-50 feature extraction network; therefore, the ResNeXt-101 network has a higher operation accuracy. Since the ResNeXt network replaces the three-layer convolutional blocks of the original ResNet with blocks of the same topology stacked in parallel, it can improve the accuracy of the model without significantly increasing the number of parameters. Compared with ResNet-50, it can be seen that ResNeXt-101 has certain advantages.

To determine the superiority of the deep network model comprised of the Faster R-CNN integrated with the FPN in weed recognition, the results of the three deep network models and the corresponding feature extraction networks were compared. The recognition performance of the ResNeXt-101 feature extraction network was the best. Therefore, on the premise of satisfying the recognition time, ResNeXt-101 was selected as the feature extraction network to compare the overall loss of the Faster R-CNN model, YOLOv3

model, and SSD model, and the accuracy of model verification. The results are shown in Figures 11 and 12 and Table 3.

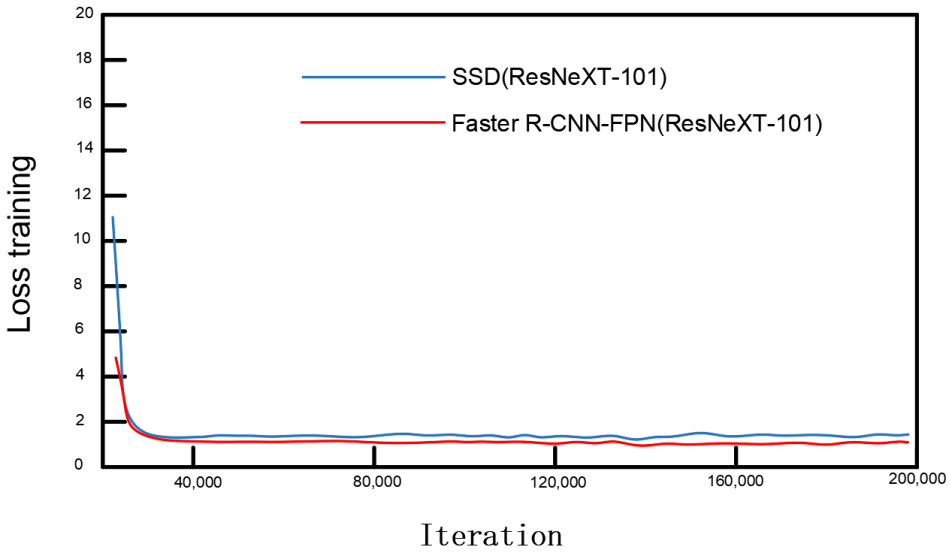


Figure 11. Comparison of the overall losses of the Faster R-CNN and SSD models.

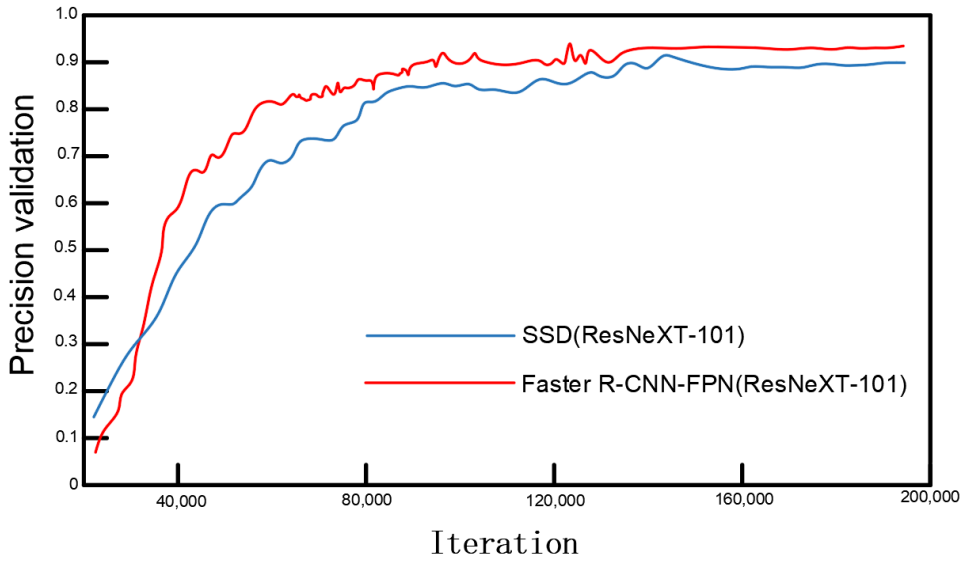


Figure 12. Comparison of the precision of the Faster R-CNN and SSD models.

Table 3. Deep network model comparison.

Model	Extraction Network	Accuracy (%)	Recall (%)	F_1-Value (%)	MIoU (%)	Detection Time (ms)
Faster R-CNN-FPN	ResNeXt-101	95.61	87.26	91.24	93.7	330
YOLOv3	ResNeXt-101	84.34	79.85	82.03	84.5	215
SSD	ResNeXt-101	88.12	79.69	83.69	85.3	203

The Faster R-CNN, SSD, and YOLOv3 deep network models are compared under the same image processing method and RESNEXT-101 feature extraction network. It can be seen from Figures 11 and 12 that when using the ResNeXt-101 feature extraction network, the SSD network has a slight oscillation at 150,000 steps, while the Faster R-CNN network

tends to stabilize after 120,000 steps. The network loss curve of Faster R-CNN with FPN is lower than that of the SSD deep network model after 4000 iterations, and the network model of Faster R-CNN integrated with FPN is better than the SSD model in terms of loss degree. By comparing the accuracy, recall rate and F-value, it can be seen that, compared with the SSD network, the accuracy rate of Faster R-CNN with the integrated FPN network is about 0.7 higher. Although the detection time is longer than that of the SSD network, the overall efficiency is still higher than that of the SSD network.

According to the model evaluation, this experiment makes a detailed comparison of the two indexes of accuracy and recall, F_1 value, as well as MIoU and detection time (single image). The comparative results of the deep network model are shown in Table 3. It can be concluded from Table 3 that the Faster R-CNN deep network model based on the ResNext-101 network shows better detection performance, while the accuracy and recall rate of the YOLOv3 deep network model are both somewhat lower than those of the Faster R-CNN network. The accuracy of the deep network model can only reach about 84% and the recall rate can only reach about 80%, the F_1 value is about 8% lower than that of Faster R-CNN and the difference in the value of MIoU can reach about 9%, although it has a shorter detection time. Hence, the advantages of the Faster R-CNN network are greater.

The Faster R-CNN network model integrated with FPN was analysed via a confusion matrix (Figure 13). The Faster R-CNN network model integrated with FPN was used for the detection of weeds and plant seedlings. The light-coloured areas show very low FN and PN values, most of which are 0 and a few are 1–4. The dark areas show very high values of TP and TN, which shows that the deep learning network predicts a high proportion of correct results; hence, it has high accuracy.

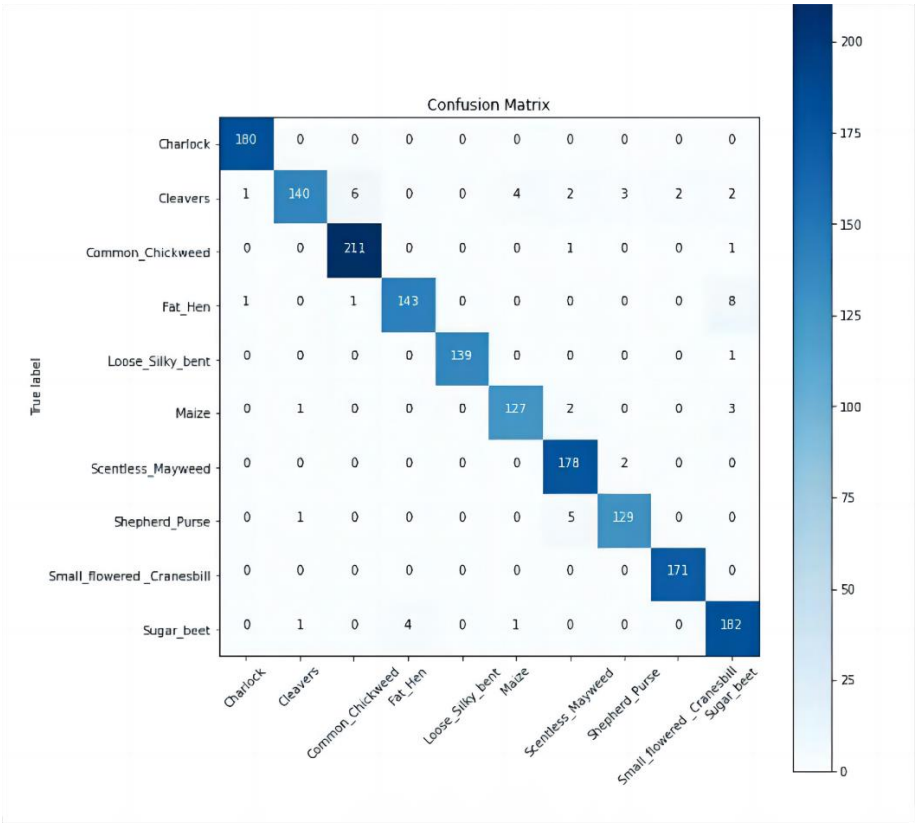


Figure 13. Confusion matrix diagram of identification results.

These model training results show the advantages and disadvantages of the two feature extraction networks (ResNet-50 and ResNeXt-101) in weed recognition. It is found that, compared with the Faster R-CNN deep network model, the YOLOv3 and SSD deep

network models have slightly inferior weed recognition accuracy. The Faster R-CNN network model obtains much better accuracy by integration with the FPN network, while the improved ResNeXt feature extraction network has higher computational efficiency. The Faster R-CNN network model integrates the FPN network into the feature extraction network and adopts the process of bottom-to-top, bottom-up, and horizontal connection to realize the simultaneous use of high-resolution low-level features and high-level features. Integrating the features of these different layers to achieve the prediction effect can greatly improve the detection accuracy. The selected ResNeXt-101 feature extraction network achieves more accurate target recognition by using a convolution operation with multiple paths. The experimental results show that the Faster R-CNN deep network model based on the ResNeXt-101 feature extraction network has obvious advantages for weed recognition in field images after being integrated with the FPN network.

6. Conclusions

This paper used corn beet and other crop seedling and weed image data to build a Faster R-CNN deep network model based on the ResNeXt-101 feature extraction network and the Tensorflow 2.0 (Developed by Google, Inc. in California, USA) deep learning framework. The FPN network and improved ResNeXt network were applied to the identification of weeds. The actual effects of the three feature extraction networks were compared and analysed, and the key parameters of the Faster R-CNN were optimized to make it more suitable for weed identification in field images with complex backgrounds.

The experimental results show that the Faster R-CNN deep network model obtains improved recognition accuracy by using the ResNeXt feature extraction network and incorporating the FPN network. It has obvious advantages compared with the ResNet feature extraction network in achieving rapid and effective target recognition, and demonstrates the great efficiency of deep learning methods in this field.

Author Contributions: Conceptualization, Y.M. and R.F.; methodology, Y.M.; software, R.F.; validation, R.N., J.L. and T.L. (Tianye Luo); formal analysis, Y.S.; investigation, T.L. (Tonghe Liu); resources, H.G.; data curation, X.L.; writing—original draft preparation, Y.M.; writing—review and editing, S.L. and Y.W.; visualization, Y.B.; supervision, Y.G.; project administration, T.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Jilin Province Science and Technology Development Plan (focuses on research and development projects), funding number 20200402006NC (http://kjt.jl.gov.cn) accessed on 1 January 2020, Key technology R&D project of Changchun Science and Technology Bureau of Jilin Province, funding number 21ZGN29 (http://kjj.changchun.gov.cn) accessed on 1 November 2021 and Science and Technology Research Project of Jilin Provincial Department of Education, funding number JJKH20220337KJ (http://jyt.jl.gov.cn/) accessed on 1 January 2022.

Data Availability Statement: All new research data were presented in this contribution.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Petrich, L.; Lohrmann, G.; Neumann, M.; Martin, F.; Frey, A.; Stoll, A.; Schmidt, V. Detection of Colchicum Autumnale in Drone Images, Using a Machine-Learning Approach. *Precis. Agric.* **2020**, *21*, 1291–1303. [CrossRef]
2. Osorio, K.; Puerto, A.; Pedraza, C.; Jamaica, D.; Rodríguez, L. A Deep Learning Approach for Weed Detection in Lettuce Crops Using Multispectral Images. *AgriEngineering* **2020**, *2*, 471–488. [CrossRef]
3. Sabzi, S.; Abbaspour-Gilandeh, Y. Using Video Processing to Classify Potato Plant and Three Types of Weed Using Hybrid of Artificial Neural Network and Particle Swarm Algorithm. *Measurement* **2018**, *126*, 22–36. [CrossRef]
4. Bakhshipour, A.; Jafari, A. Evaluation of Support Vector Machine and Artificial Neural Networks in Weed Detection Using Shape Features. *Comput. Electron. Agric.* **2018**, *145*, 153–160. [CrossRef]
5. Wu, G.; Zhu, M.; Jiang, Q.; Sun, X. Fast-Axis Collimating Lens Recognition Algorithm Based on Machine Vision. *J. Phys. Conf. Ser.* **2021**, *1820*, 012157. [CrossRef]
6. Wu, Z.; Chen, Y.; Zhao, B.; Kang, X.; Ding, Y. Review of Weed Detection Methods Based on Computer Vision. *Sensors* **2021**, *21*, 3647. [CrossRef] [PubMed]

7. Anthoniraj, S.; Karthikeyan, P.; Vivek, V. Weed Detection Model Using the Generative Adversarial Network and Deep Convolutional Neural Network. *J. Mob. Multimed.* **2021**, *18*, 275–292. [CrossRef]
8. Cui, F.; Ning, M.; Shen, J.; Shu, X. Automatic Recognition and Tracking of Highway Layer-Interface Using Faster R-CNN. *J. Appl. Geophy.* **2022**, *196*, 104477. [CrossRef]
9. Yan, D.; Li, G.; Li, X.; Zhang, H.; Lei, H.; Lu, K.; Cheng, M.; Zhu, F. An Improved Faster R-CNN Method to Detect Tailings Ponds from High-Resolution Remote Sensing Images. *Remote Sens.* **2021**, *13*, 2052. [CrossRef]
10. Yujian, C.; Guoming, X.; Guochuan, S. Low Altitude Armored Target Detection Based on Rotation Invariant Faster R-CNN. *Laser Optoelectron. Prog.* **2018**, *55*, 101501. [CrossRef]
11. Ma, X.; Zhu, D.; Jin, C.; Tong, X. Bullet Appearance Defect Detection Based on Improved Faster Region-Convolutional Neural Network. *Laser Optoelectron. Prog.* **2019**, *56*, 151202.
12. Chen, E.; Tang, X.; Fu, B. Pedestrian Search Method Based on Faster R-CNN with the Integration of Pedestrian Detection and Re-Identification. *J. Comput. Aided Des. Comput. Graph.* **2019**, *31*, 332–339.
13. Prasanna Kumar, A.; Mishra, D. Harnessing Feedback Region Proposals for Multi-object Tracking. *IET Comput. Vis.* **2020**, *14*, 434–442. [CrossRef]
14. Giselsson, T.M.; Jørgensen, R.N.; Jensen, P.K.; Dyrmann, M.; Midtiby, H.S. A Public Image Database for Benchmark of Plant Seedling Classification Algorithms. *arXiv* **2017**, arXiv:1711.05458.
15. Wang, S.; Yang, L. Feature dimension reduction and category identification of weeds in cotton field based on GA-ANN complex algorithm. *J. Henan Agric. Sci.* **2018**, *47*, 148–154, 160.
16. Wang, H.; Yang, J.; Xue, B.; Yan, X.; Tao, J. A Novel Color Calibration Method of Multi-Spectral Camera Based on Normalized RGB Color Model. *Results Phys.* **2020**, *19*, 103498. [CrossRef]
17. Liu, S.; Xue, Q.; Li, Z.; Li, C.; Gong, Z.; Li, N. An air temperature predict model based on BP neural networks for solar greenhouse in North China. *J. China Agric. Univ.* **2015**, *20*, 176–184.
18. Ranjitha, P.; Shreelakshmi, T.D. A Hybrid Ostu Based Niblack Binarization for Degraded Image Documents. In Proceedings of the 2021 2nd International Conference for Emerging Technology (INCET), Belagavi, India, 21–23 May 2021; IEEE: New York, NY, USA, 2021; pp. 914–920.
19. Xu, Z.-F.; Jia, R.-S.; Sun, H.-M.; Liu, Q.-M.; Cui, Z. Light-YOLOv3: Fast Method for Detecting Green Mangoes in Complex Scenes Using Picking Robots. *Appl. Intell.* **2020**, *50*, 4670–4687. [CrossRef]
20. Huang, Z.; Wang, J.; Fu, X.; Yu, T.; Guo, Y.; Wang, R. DC-SPP-YOLO: Dense Connection and Spatial Pyramid Pooling Based YOLO for Object Detection. *Inf. Sci.* **2020**, *522*, 241–258. [CrossRef]
21. Kuznetsova, A.; Maleva, T.; Soloviev, V. Detecting Apples in Orchards Using YOLOv3. In Proceedings of the Computational Science and Its Applications—ICCSA 2020, Cagliari, Italy, 1–4 July 2020; Springer International Publishing: Cham, Switzerland, 2020; pp. 923–934.
22. Lin, T.; Dollar, P.; Girshick, R. 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. 936–944.
23. Wang, D.D.; He, D.J. Apple Target Recognition of Robot before Fruit Thinning Based on R-FCN Deep Convolution Neural Network. *Trans. Chin. Soc. Agric. Eng.* **2019**, *35*, 156–163.

agronomy

MDPI

Article

Determining Strawberries' Varying Maturity Levels by Utilizing Image Segmentation Methods of Improved DeepLabV3+

Changqing Cai [1,2], Jianwen Tan [3], Peisen Zhang [4], Yuxin Ye [5,*] and Jian Zhang [6,7,*]

[1] Engineering Training Center, Changchun Institute of Technology, Changchun 130012, China
[2] National and Local Joint Engineering Research Center for Smart Distribution Network Measurement, Control and Safe Operation Technology, Changchun 130012, China
[3] College of Electrical and Information Engineering, Changchun Institute of Technology, Changchun 130012, China
[4] College of Energy and Power Engineering, Changchun Institute of Technology, Changchun 130012, China
[5] College of Computer Science and Technology, Jilin University, Changchun 130012, China
[6] Faculty of Agronomy, Jilin Agricultural University, Changchun 130018, China
[7] Department of Biology, University of British Columbia, Okanagan, Kelowna, BC V1V 1V7, Canada
* Correspondence: yeyx@jlu.edu.cn (Y.Y.); jian.zhang@ubc.ca (J.Z.)

Abstract: Aiming to determine the inaccurate image segmentation of strawberries with varying maturity levels due to fruit adhesion and stacking, this study proposed a strawberry image segmentation method based on the improved DeepLabV3+ model. The technique introduced the attention mechanism into the backbone network and the atrous spatial pyramid pooling module of the DeepLabV3+ network, adjusted the weights of feature channels in the neural network propagation process through the attention mechanism to enhance the feature information of strawberry images, reduced the interference of environmental factors, and improved the accuracy of strawberry image segmentation. The experimental results showed that the proposed method can accurately segment images of strawberries with different maturities; the mean pixel accuracy and mean intersection over union of the model were 90.9% and 83.05%, respectively, and the frames per second (FPS) was 7.67. The method can effectively reduce the influence of environmental factors on strawberry image segmentation and provide an effective approach for accurate operation of strawberry picking robots.

Keywords: improved DeepLabV3+; attention mechanism; image segmentation; strawberry

Citation: Cai, C.; Tan, J.; Zhang, P.; Ye, Y.; Zhang, J. Determining Strawberries' Varying Maturity Levels by Utilizing Image Segmentation Methods of Improved DeepLabV3+. *Agronomy* **2022**, *12*, 1875. https://doi.org/10.3390/agronomy12081875

Academic Editor: Andrea Peruzzi

Received: 25 June 2022
Accepted: 6 August 2022
Published: 9 August 2022

1. Introduction

Strawberry has been planted worldwide due to its strong environmental adaptability and high economic benefits, among which China ranks first in the world in terms of strawberry cultivation area [1]. With the increasing scale of strawberry cultivation, the traditional manual picking can no longer meet the harvesting/picking demand of the strawberry industry, and the automation of strawberry picking has become the focus of research in many countries [2]. Strawberry ripeness determination is used as a judgment condition for automated strawberry picking.

Researchers around the world have investigated various techniques to determine strawberry ripeness classes, including spectroscopic techniques and machine vision [3]. Spectroscopy mainly used optical information to obtain information about the chemical composition and physical properties of strawberry fruit, which is used as input data for classification models to determine strawberry fruit ripeness classes. Raj et al. [4] used a narrowband hyperspectral radiometer to collect the reflection characteristics of strawberry fruits of different maturity levels and subsequently took the drying method to obtain the water content of ground strawberries. They used the above two sets of data as input data for a linear support vector machine model. It was able to achieve 98% accuracy in strawberry

maturity classification; The classification accuracy of strawberry maturity is 71% when only the data set of strawberry fruit moisture content was available. Su [5] used a hyperspectral imager to collect one-dimensional spectral and three-dimensional spectral images and used the above data as input data in the residual network to build a one-dimensional model. Constructed networks for strawberry fruit classification at different maturity levels showed the accuracy of both networks reached 84%. However, this paper mainly adopted machine vision to obtain the image information of strawberry to determine the ripeness of strawberry fruit. Meanwhile, the strawberry fruit is in an unstructured environment, and there are factors such as dense distribution among fruits, leaf shading, and fruit stacking which make it difficult for the existing image recognition algorithms to perform accurately. Therefore, the research on strawberry image recognition, segmentation and localization methods with high recognition accuracy and strong environmental adaptability is the focus of this paper.

In recent years, as the research on deep learning technology in the field of computer vision has become very active [6,7], its recognition and segmentation of strawberry images has gradually become a hot topic for domestic and international research [8]. Semantic segmentation and instance segmentation are more popular in deep learning segmentation, and semantic segmentation networks mainly use models such as FCN [9,10], PSPNet [11], U-Net [12], SegNet [13], and DeepLab [14–16], while instance segmentation often uses Mask R-CNN [17–20] models. Ilyas [21] proposed a novel convolutional encoder/decoder network model, which combines dilated residual blocks (DRB), bottleneck blocks (BB) and an adaptive receptive field module (ARFM), reduces the network computational complexity, and enhances the network feature extraction capability. However, because the dataset is too small and usually does not contain occluded strawberry images, it is still necessary to increase the number of strawberry images and enrich the diversity of strawberry images to verify the general applicability of the model. Yu [22] proposed to use the Mask R-CNN model for recognition and segmentation of strawberry images in unstructured environments, combining Resnet50 with Feature Pyramid Network (FPN) as the backbone network to extract feature maps, then inputting the feature maps to a Region Proposal Network (RPN) to generate a Region of Interest (ROI), and finally inputting the region of interest to FCN; the target mask image was generated and the experimental results of this method for the masked strawberry image showed that the average detection accuracy is 95.78%; Jia [23] used the U-Net model as a prototype, and selected the convolution with the same improved VGG16 model to extract image feature information and retain more feature information to improve the model segmentation accuracy, and the experimental results of this method for obscured strawberry images showed that the average detection accuracy is 96.05%; Ge [24] proposed to use the Mask R-CNN model with DCNN network backbone to segment strawberry images. For the strawberry fruit occlusion problem, the occluded strawberries were detected and the occluded part was compensated by the enclosing box; the experimental results of this method on the occluded strawberry images showed that the average detection accuracy was 94%. The models in the literature [22–24] were focused on high detection accuracy when identifying images with occluded strawberries, but the real-time detection efficiency of the models was not mentioned, and experiments on model detection speed are still required to verify the feasibility of model integration into strawberry picking robots.

In summary, previous studies conducted valuable research on image segmentation of strawberry with different maturities to reduce the influence of environmental factors and improve the strawberry image segmentation accuracy, but have not considered improving the image segmentation accuracy without losing the model detection efficiency. Even though the image segmentation accuracy of the proposed method is acceptable, the recognition efficiency could not meet the real-time requirement of automatic strawberry picking. To address this issue, this paper proposes a method to improve DeepLabV3+ for strawberry image segmentation with different maturities, introducing the attention mechanism in the backbone feature extraction network and ASPP module, respectively,

and adjusting the weights of feature channels in the neural network propagation process through the attention mechanism, which could enhance the feature information of strawberry images with different maturities using fewer parameters, reducing the interference of environmental factors. The accuracy of strawberry image segmentation can be improved without losing the efficiency of model detection.

2. Materials and Methods

The overall workflow of this paper is shown in Figure 1, which mainly consists of 2 parts: (1) constructing the strawberry image dataset; (2) training the strawberry image segmentation model.

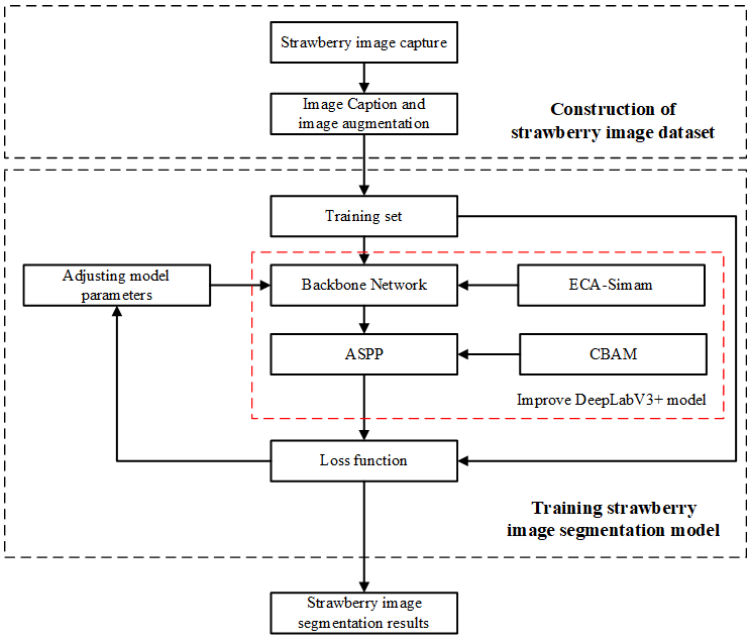


Figure 1. Overall workflow diagram.

2.1. Construction of Strawberry Image Dataset

Since there is no publicly available image dataset for strawberry cultivation on high shelves, this paper needed to construct an image dataset for strawberry cultivation on high shelves for research purposes. The strawberry image data in this paper were collected from the elevated strawberry cultivation base of Guoxin Modern Agriculture Company in Changchun, Jilin Province, China. A total of 1000 images of strawberries at different growth stages, different fruit numbers, different shading levels, and different light intensities were collected with a 40-megapixel HD camera, as shown in Figure 2, where Figure 2a–c show strawberry fruits at different growth stages and Figure 2d–f correspond to strawberry fruit images at different shading levels, respectively. During strawberry production, when the percentage of the red colored area is more than 75%, the fruit can be picked. Therefore, we classify the strawberry maturity into 3 stages: ripe ($s > 75\%$), semi-ripe ($25\% < s < 75\%$), and unripe ($s < 25\%$), according to the fruit coloring area (s), as shown in Figure 2a–c. The original image size of the acquired strawberry was 4624×3468 px, and to reduce the computational effort, we compressed the image size to 512×512 px during the training model.

The LabelMe [25] annotation tool was used to manually annotate each strawberry sample image to generate a Json file containing a large amount of label image information (location information of all annotated points, image size, and image category). We transformed each Json file into a masked label image by the Json command, as shown in Figure 3a. The immature strawberry mask image is green, the near-mature strawberry mask image is yellow, and the mature strawberry mask image is red, as shown in Figure 3b.

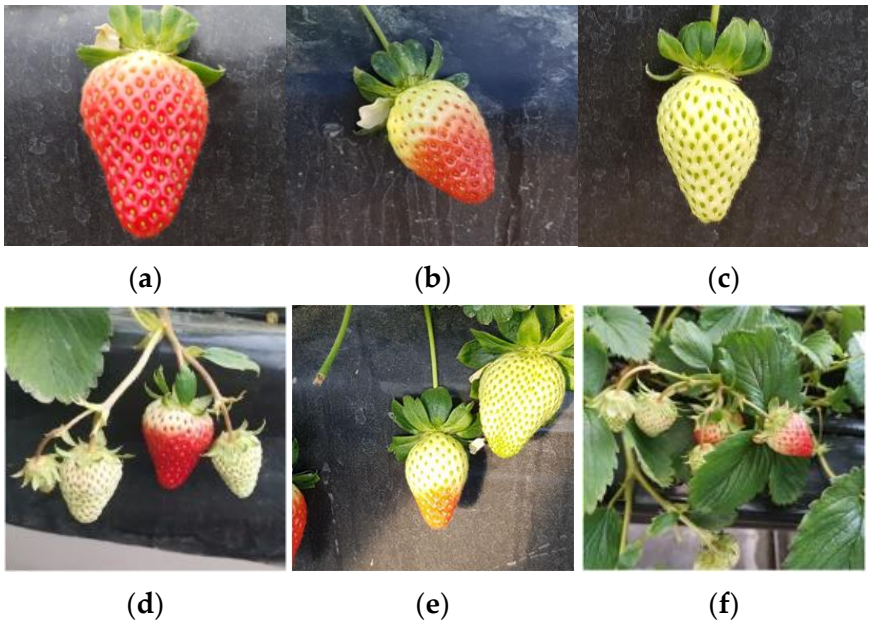


Figure 2. Images of strawberry samples collected under greenhouse environment. (**a**) Mature strawberry; (**b**) near-mature strawberry; (**c**) immature strawberry; (**d**) uncovered strawberry; (**e**) strawberry under strong light; (**f**) covered strawberry.

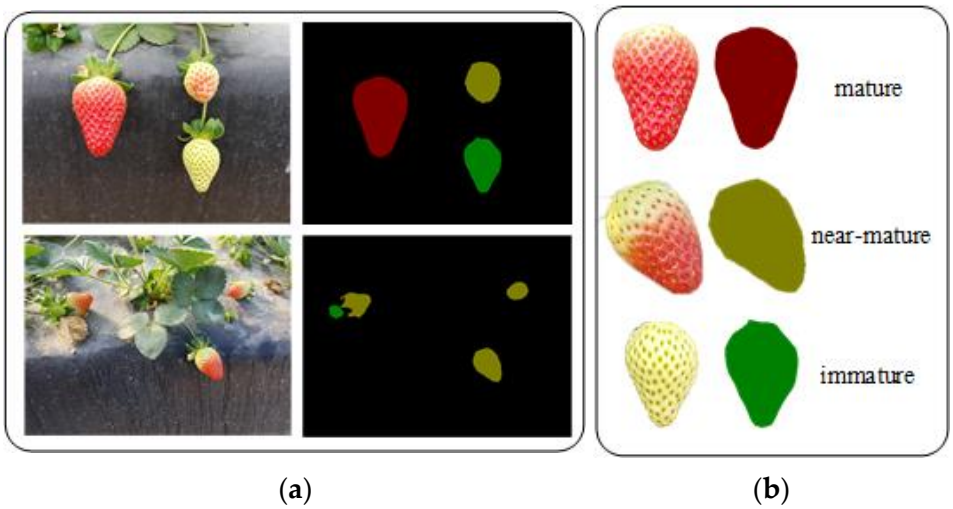


Figure 3. Labeled sample data set. (**a**) Strawberry image dataset and labeled images; (**b**) strawberry mask images.

To improve the network model training effect and model generalization ability, this study used the mirror flip method to increase the spatial diversity of strawberry images [26]; and brightness adjustment, adaptive contrast enhancement and Kmeans clustering were used to increase the diversity of strawberry image samples, as shown in Figure 4. A total of 6000 sample images were enhanced and the dataset was divided into training set (4200 images), test set (1200 images) and validation set (600 images) according to the ratio of 7:2:1.

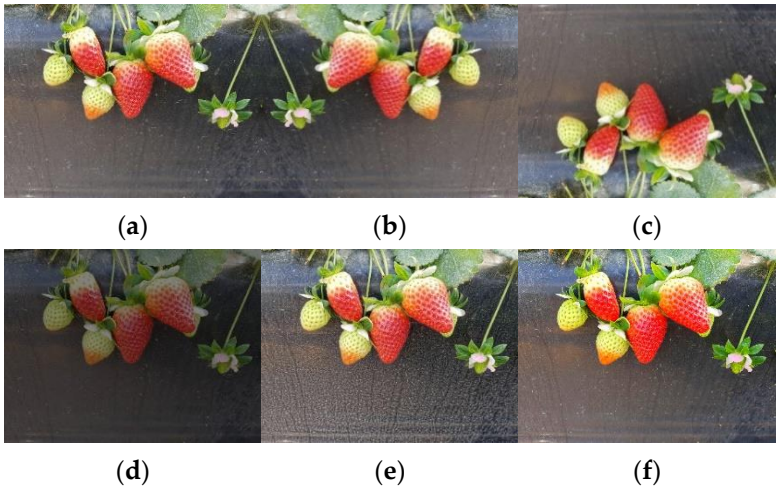


Figure 4. Plot of sample effects produced by different data enhancement methods. (**a**) Original image; (**b**) horizontal mirror; (**c**) vertical mirror; (**d**) diminished brightness; (**e**) contrast enhancement; (**f**) color enhancement.

2.2. Improved DeepLabV3+ Strawberry Image Segmentation Model

DeepLabV3+ [27] is a classical semantic segmentation network containing 2 parts: encoder and decoder. The encoder consists of a backbone feature extraction network and an atrous spatial pyramid pooling (ASPP) [28] structure, while the decoder obtains low-level features from the backbone feature extraction network and upsamples them to obtain pixel-by-pixel classification results of the same size as the input image. The backbone feature extraction network is an Xception [29] network, which is based on InceptionV3 [30] and uses depth wise separable convolution to replace the multi-size convolutional kernel feature response operation in InceptionV3, significantly reducing the number of model parameters, lowering the computational cost of the model, and improving the operational efficiency of the model. The ASPP structure consists of one 1×1 convolution in parallel, three 3×3 null convolutions and one global average pooling operation, where the expansion rates of the three 3×3 convolution operations are 6, 12, and 18, respectively. The structure is capable of multi-scale sampling of the feature map using null convolution operations with different sampling rates, expanding the perception of the convolution kernel, avoiding the loss of image detail features, and enhancing the adaptability to multi-scale targets. In the decoder part, the low-level features output from the backbone network are spliced and fused with the high-level features output from the encoder part, and then 1×1 convolution and upsampling are performed to obtain a classification result mask image with the same size as the input image.

The DeepLabV3+ network model was applied to strawberry image segmentation, and its model structure and segmentation effect are shown in Figure 5. Although the model can segment the strawberry image regions well, there are mis-segmentation cases when classifying strawberry images with different maturities, and the overall segmentation accuracy is low. Since the backbone network Xception and ASPP modules in DeepLabV3+ model adopt depthwise separable convolution and Dilated Convolutions with different expansion rates (6, 12, 18), respectively, although it improves the overall operational efficiency of the model, it reduces the extraction ability of target features and affects the association between local features of the target, thus producing the phenomenon of target semantic segmentation void; when DeepLabV3+ model deepens the number of network layers in the forward propagation process, the representation of obscured strawberry fruit features becomes weaker and weaker, leading to the disappearance of strawberry detail features in the whole network propagation process and the phenomenon of missegmentation.

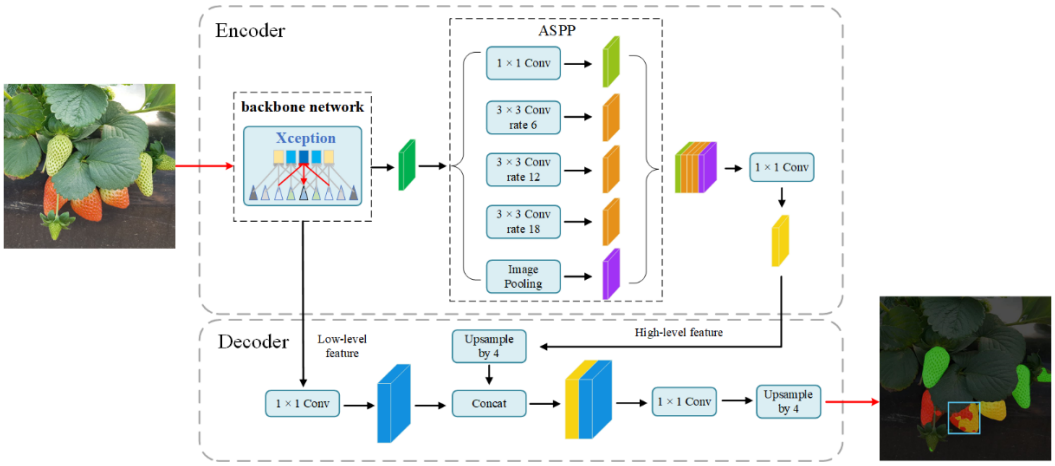


Figure 5. DeepLabV3+ structure and segmentation effect.

To address the above problems and considering the deployment of a strawberry image segmentation model in strawberry picking robots, the segmentation accuracy of the model needs to be improved without losing model operation efficiency. Therefore, this paper introduces the attention mechanism into the backbone network and ASPP module of the DeepLabV3+ model, respectively, and proposes an improved DeepLabV3+ strawberry image segmentation method. The structure of the improved DeepLabV3+ network is shown in Figure 6.

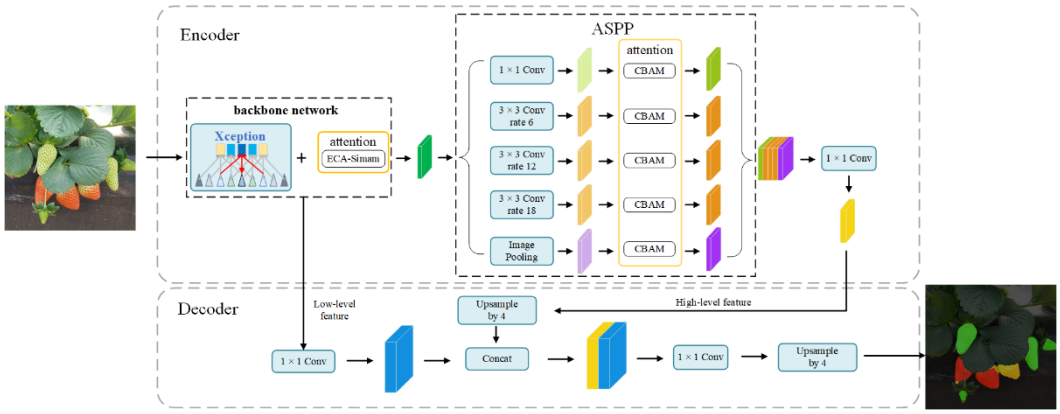


Figure 6. Improved DeepLabV3+ network structure diagram.

Compared with the classical DeepLabV3+ model, the present study mainly improves the backbone network and ASPP module in the DeepLabV3+ model. In the backbone network of the DeepLabV3+ model, after each depth-separable convolution, the ECA-SimAM module was introduced to strengthen the correlation between the feature information extracted by the network in the spatial domain and the channel domain, which improved the feature extraction ability of the backbone network and the accuracy of the model image segmentation. With the ASPP module of DeepLabV3+, after each atrous convolution, the CBAM module was introduced to adjust the weight share of feature channels to reduce the interference of environmental factors on strawberry image detection. The results showed an improvement in image segmentation accuracy.

2.2.1. Dual-Attention Mechanism to Optimize the Backbone Network

In this paper, the dual-attention mechanism is introduced into the DeepLabV3+ backbone network Xception to optimize the feature extraction ability of the backbone network, mine the important information in the feature map with fewer parameter calculations, and

adjust the proportion of important information weights in the neural network propagation process so as to enhance the feature extraction ability of the backbone network.

The network structure of the dual-attention mechanism is shown in Figure 7. The SimAM [31] module mines the importance of each neuron of the feature map by an energy function; no additional parameters are required to derive 3D attention weights for the feature map. Therefore, the ECA [32] module was combined with the SimAM module to construct an ECA-SimAM serial structure to improve the feature extraction capability of the ECA module without increasing the number of extra parameters for this model. The structure compresses the feature map F into a one-dimensional feature vector by global average pooling, then multiplies the convolved one-dimensional feature vector with the original input feature map F to get the feature map F′; then, it mines the importance of each neuron in the feature map F′ by a set of energy functions to derive a set of 3D attention weights for the feature map F′, and finally multiplies the obtained weights with the feature map F′ to get the final output feature map F′′. The energy function equation is as follows:

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1}\sum_{i=1}^{M-1}(-1-(\omega_t x_i + b_t))^2 + (1-(\omega_t t + b_t))^2 + \lambda\omega_t^2 \quad (1)$$

where t and x_i denote the target neurons and other neurons of the input features, i denotes the index on the spatial dimension, M denotes the number of neurons on the channel, and ω_t and b_t denote the weights and biases of the neurons when they are transformed, respectively.

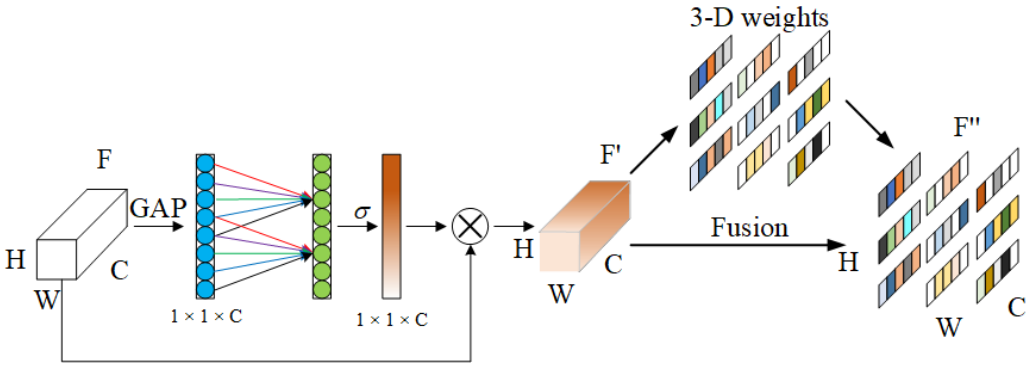


Figure 7. ECA-SimAM attention module.

The dual-attention mechanism optimizes the backbone network as shown in Figure 8, where the black font indicates the original structure of the backbone Xception network and the red font indicates the access attention module. Xception consists of Entry flow, Middle flow and Exit flow, each of which makes extensive use of deep separable convolution to split the correlation between the spatial dimension and the channel dimension, reducing the number of parameters needed for convolution calculations, reducing the complexity of parameter calculations, making the model more lightweight and losing some detection accuracy. Therefore, after each depth-separable calculation of the Xception network, the ECA-SimAM module is introduced to strengthen the correlation of the network extracted feature information in the spatial and channel domains, thus enhancing the feature extraction capability of the Xception network and providing model image segmentation accuracy.

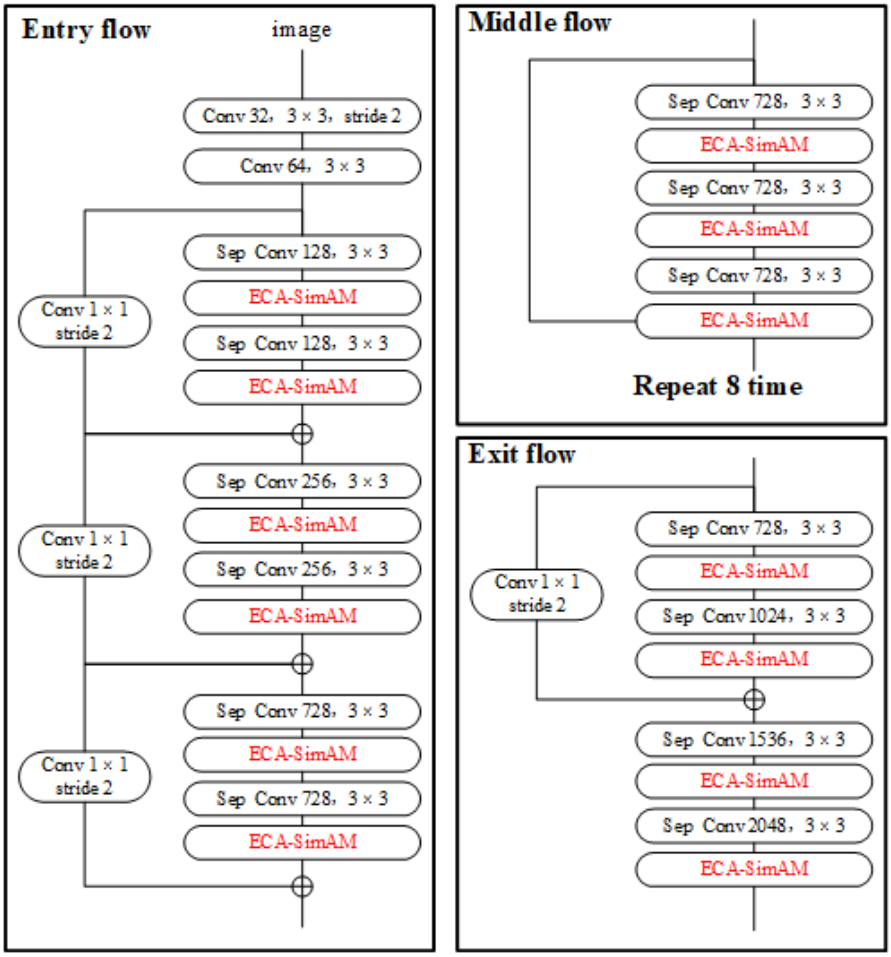


Figure 8. Dual-attention mechanism to optimize the backbone network.

2.2.2. Convolutional Attention Mechanism to Enhance ASPP

The ASPP module of the DeepLabV3+ model uses the null convolution with different expansion rates (6, 12, 18) to process the feature maps output from the backbone network in parallel to extracting multi-scale target information, but too large expansion rates will prevent the network from extracting image edge feature information well and also affect the association between local features of the target, making the expression of the obscured strawberry image weaker and reducing the strawberry image segmentation accuracy. Therefore, this paper integrates the CBAM [33] module into the ASPP module, reduces the interference of environmental factors (fruit adhesion, branch and leaf occlusion, and fruit stacking) through the attention mechanism, adjusts the weight share of the feature channels, solves the problem of weak feature expression of the occluded strawberry images, and improves the accuracy of model image segmentation. Among them, the CBAM module is an attention module combining both channel and spatial dimensions, and its network structure is shown in Figure 9; in the channel dimension, the input feature map F is pooled by the maximum and average values to get two sets of one-dimensional vectors; then, the multilayer perceptron network (MLP) is sequentially downscaled and upscaled from the channel dimension, the one-dimensional vector Mc(F) is obtained by summing up the elements, and finally Mc(F) is multiplied with the input The feature map F' is obtained by multiplying Mc(F) with the input feature map F; in the spatial dimension, the feature map F' is obtained by average pooling and maximum pooling to obtain 2 feature maps, then the feature map Mc(F') is obtained by descending the 3×3 convolution kernel, and finally the output feature map F'' is obtained by multiplying it with the feature map F'.

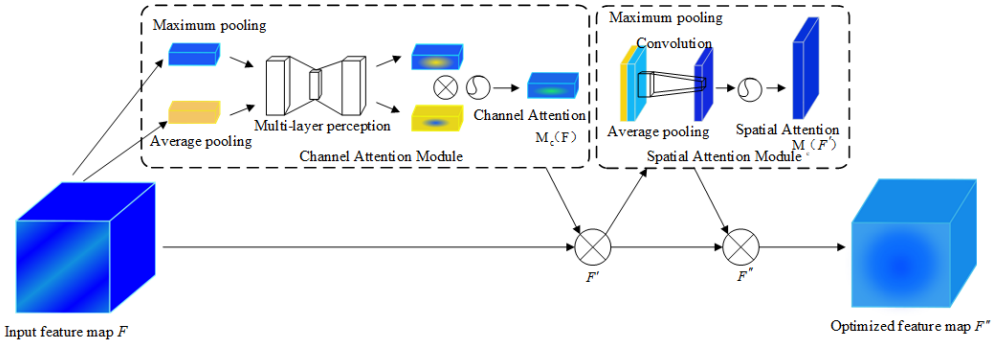


Figure 9. CBAM Attention Module.

2.3. Model Evaluation Metrics

In this study, strawberry ripeness was classified into mature (s > 75%), near-mature (25% < s < 75%) and immature (s < 25%) based on the strawberry fruits red coloring area (s). The Mean Intersection over Union (e_{mIoU}) and the Mean Pixel Accuracy (e_{mPA}) [34] were selected to evaluate the segmentation accuracy of the strawberry image segmentation model, and the Frames Per Second (FPS) was selected to evaluate the detection efficiency of the strawberry image segmentation model, where e_{mIoU} can evaluate both missed and false detection, e_{mPA} mainly evaluates the missed detection of the algorithm, and FPS indicates the number of input image frames per second. The specific formulas of e_{mIoU} and e_{mPA} are listed as:

$$e_{mIoU} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}} \tag{2}$$

$$e_{mPA} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \tag{3}$$

where k is the number of categories; p_{ij} is the pixel marked as class i, but the prediction result is class j; p_{ii} is the pixel marked as class i, and the prediction result is also class i.

3. Results

3.1. Confirmation of Test Environment and Parameter Settings

The test environment in this paper was divided into the hardware environment and the software environment, as shown in Table 1.

Table 1. Test environment configuration.

Environment	Configuration	Parameters
Hardware environment	CPU	4-core Intel(R) Xeon(R) Silver 4110 @ 2.10 GHz
	GPU	RTX 2080 Ti
	Memory	16 GB
Software environment	System	Ubuntu 16.04
	deep learning framework	Pytorch 1.10.0
	programming environment	python 3.8
	GPU parallel computing architecture	Cuba 10.1
	GPU acceleration library	Cudnn 7.6.5.

The model uses pre-trained weights from the PASCAL VOC dataset with a freeze training strategy. First, the backbone network parameters are frozen for training 1000 times, with 8 images in each batch, and the initial learning rate is set to 0.0003; then the network

training is unfrozen for 2000 times, with 2 images in each batch, and the initial learning rate is set to 0.00001, for a total of 3000 iterations.

3.2. Analysis of Strawberry Image Segmentation Models

3.2.1. Ablation Experiments

In this paper, we mainly introduce the attention mechanism to improve the classical DeepLabV3+ network, optimize the backbone network, and enhance the ASPP module, respectively, and to improve the model segmentation accuracy without losing the model real-time detection efficiency. The model performance evaluation is shown in Table 2, where "√" indicates that the marked module is used for the experiment and no "√" indicates that this module is not added.

Table 2. Ablation experiments were performed by adding an attention module.

DeeplabV3+	SimAM Optimize Backbone Network	ECA Optimize Backbone Network	CBAM Enhanced ASPP	Network Model Parameters	MIoU	MPA	FPS
√				209.70M	74.04%	85.45%	8.52
√	√			209.71M	75.12%	86.40%	8.06
√		√		209.74M	76.33%	86.33%	8.18
√			√	209.77M	75.92%	86.59%	8.37
√	√	√		209.75M	80.78%	88.55%	7.71
√	√		√	209.77M	77.52%	87.44%	7.83
√		√	√	209.80M	78.63%	87.74%	8.01
√	√	√	√	209.81M	**83.05%**	**90.90%**	7.67

Note: Bold is the optimal result and also the network proposed in this paper.

In the ablation experiment, compared with the classical DeepLabV3+, DeepLabV3+ with the SimAM module, DeepLabV3+ with the ECA module and DeepLabV3+ with the ECA-SimAM module, the MIoU was improved by 1.08%, 2.29% and 6.74%, respectively, and the MPA was improved by 0.95%, 0.88%, and 3.1%, respectively. The experimental results show that the DeepLabV3+ model with the ECA-SimAM module added has the highest MIoU and MPA, and the dual-attention mechanism has stronger feature extraction capability than the single-attention mechanism.

To further analyze the improvement of the model image segmentation accuracy, different attention mechanisms were introduced for experiments, respectively. The experimental results show that the MIoU and MPA of DeepLabV3+ with the introduction of the ECA-SimAM module and CBAM module are the highest, which are 83.05% and 90.90%, respectively, and compared with the classic DeepLabV3+ model; MIou and MPA are improved by 9.01% and 5.62%, respectively, compared with the classical DeepLabV3+ model. Meanwhile, the FPS of the model running in real time is 7.67, which is only 0.85 lower compared with the classical DeepLabV3+ model, and does not affect the efficiency of the model running in real time. Overall, the model proposed in this paper can improve segmentation accuracy without losing real-time detection efficiency.

3.2.2. Performance Analysis of Different Models

To verify the superiority of the model in this paper for the segmentation accuracy of strawberry images with different maturity, U-Net [35], PSPNet [36], HRNet [37] and DeepLabV3+ models were selected for comparison tests, and the evaluation indexes of the segmentation accuracy of different models were obtained as shown in Table 3.

Table 3. Segmentation accuracy of different segmentation models in strawberry images.

Models	Intersection over Union			e_{mIoU}	Pixel Accuracy			e_{mPA}
	Immature	Near-Mature	Mature		Immature	Near-Mature	Mature	
U-Net	74.14%	70.87%	74.65%	73.22%	83.71%	84.92%	78.83%	82.48%
PSPNet	72.66%	72.84%	75.02%	73.50%	82.60%	89.18%	79.59%	83.79%
HRNet	77.99%	71.32%	77.21%	75.51%	84.22%	83.92%	87.97%	85.37%
DeepLabV3+	76.95%	69.91%	75.26%	74.04%	86.54%	88.21%	81.59%	85.45%
Proposed	80.59%	82.23%	86.34%	**83.05%**	90.40%	90.52%	91.79%	**90.90%**

Note: Bold is the optimal result and also the network proposed in this paper.

As shown in the Table 3, the e_{mIoU} and the e_{mPA} in this paper are the highest, which are 83.05% and 90.90%, respectively. Compared with the U-Net, PSPNet, HRNet and DeepLabV3+ models, the e_{mIoU} of the model in this paper is improved by 9.83%, 9.55%, 7.51% and 9.01%, respectively; the e_{mPA} is improved by 8.42%, 7.11%, 5.53 and 5.45%, respectively. Overall, the performance of the model proposed in this paper is optimal.

3.2.3. Different Model Segmentation Results

The visualization results of the model in this paper with different image segmentation models are shown in Figure 10. In the visualization results, the red area is the ripe strawberry mask image, the yellow area is the semi-ripe strawberry mask image, and the green area is the unripe strawberry mask image; the blue boxed area in the figure indicates the mis-segmentation phenomenon of the model.

From Figure 10a, it can be seen that all models can accurately segment the strawberry images when the fruits are in the unobstructed situation, and only U-Net is inaccurate in segmentation of the details. From Figure 10b, it can be seen that U-Net segmentation is poor when there are adhesions between fruits, and it is difficult to accurately segment the strawberry images with different ripeness; PSPNet, HRNet, DeepLabV3+ and the models in this paper can accurately segment the strawberry images with different ripeness.

From Figure 10c, it can be seen that U-Net, PSPNet, HRNet, and DeepLabV3+ all showed missegmentation when the fruit was obscured by leaves. In sample 4, U-Net could not accurately segment the strawberry images, and PSPNet, HRNet and DeepLabV3+ had difficulty in accurately extracting the detailed features of strawberries of different maturity and failed to accurately segment the strawberry images of different maturity; in sample 5, U-Net and PSPNet incorrectly classified the ripeness of strawberries, and HRNet and DeepLabV3+ accurately classified the ripeness of strawberries, but failed to completely segment the strawberry images in the unobscured area, while the model in this paper can not only accurately classify the ripeness of strawberries but also completely segment the strawberry images.

From Figure 10d, we can see that U-Net, PSPNet, HRNet, and DeepLabV3+ all show mis-segmentation when the fruits are stacked, while the model in this paper can accurately segment the strawberry images.

In summary, compared with the classical image segmentation model, the model in this paper can effectively reduce the interference of environmental factors, extract more strawberry image feature information, accurately segment strawberry images with different maturity, and has stronger robustness.

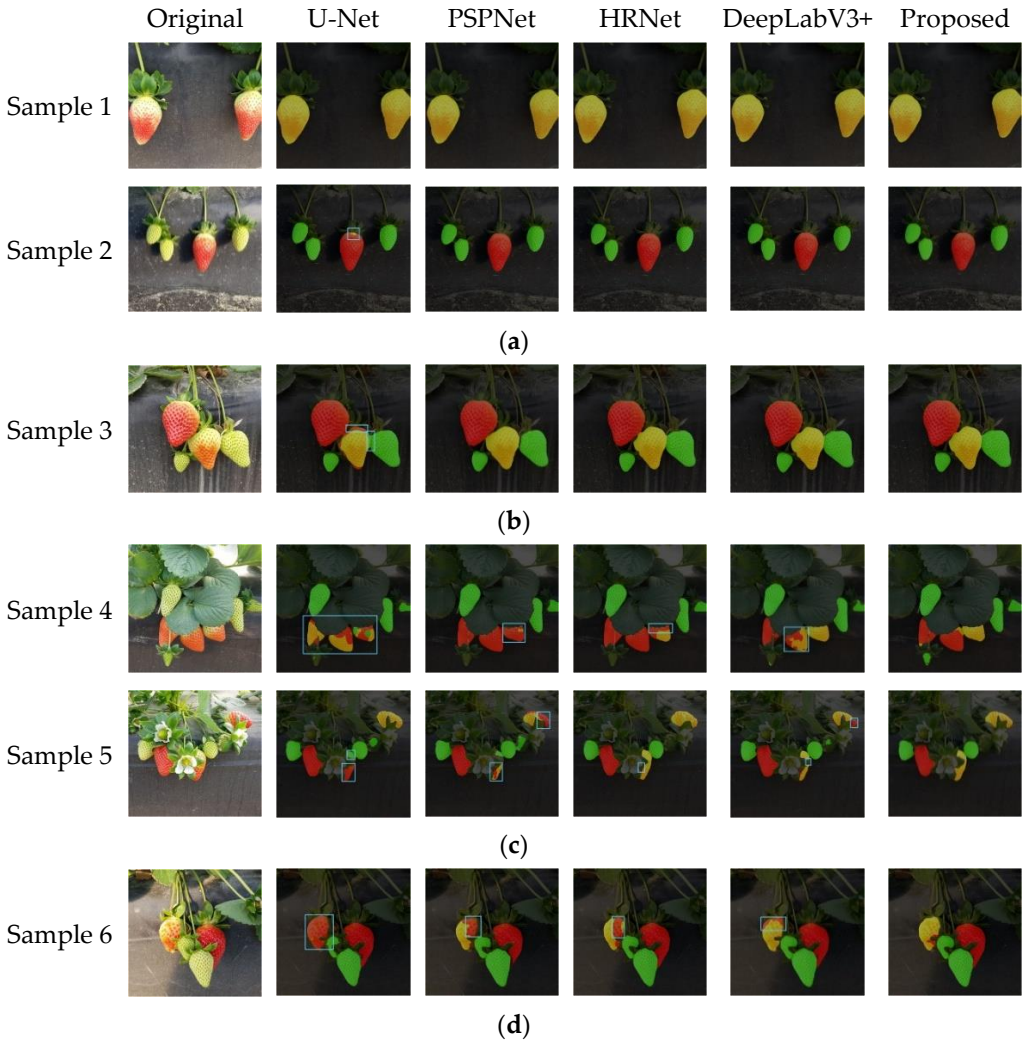


Figure 10. The effect of different strawberry image segmentation models. (**a**) No shade; (**b**) fruit adhesion; (**c**) leaf shading; (**d**) fruit stacking.

4. Discussion

Emerging in smart agriculture research, image segmentation of strawberries at different maturity levels has attracted much interest, which has focused on designing many image segmentation methods to reduce the interference of environmental factors and improve the segmentation accuracy of images. The ultimate goal of these studies was also to deploy the methods in strawberry picking robots to provide a basis for their accurate operation. Therefore, the research on image segmentation methods for strawberries should improve the image segmentation accuracy without losing the real-time operation efficiency of the model. In this paper, we propose an improved DeepLabV3+ method for strawberry image segmentation with different maturity levels, which is based on a DeepLabV3+ network and allows the neural network to ignore irrelevant feature information and focus on important information through the attention mechanism so as to improve the model image segmentation accuracy with less computational effort. The method can improve strawberry image segmentation accuracy on the basis of satisfying real-time operation of strawberry picking robots.

In the future work, we will conduct experiments on different locations and different varieties of strawberries, analyze the commonality and characteristics of different locations and different varieties of strawberries, and build a more universal strawberry image dataset to make the designed image segmentation model more accurate. At the same time, considering the deployment of the model in strawberry picking robots, the model should

be further designed to be lightweight, with reduced internal architecture and components and reduced computational cost, without losing its recognition accuracy, so as to be better adapted to strawberry picking robots.

5. Conclusions

In this paper, we propose a method to improve DeepLabV3+ for strawberry image segmentation with different maturity levels. The multi-attention mechanism is incorporated into the DeepLabV3+ model to increase the weight of strawberry image feature information and decrease the weight of environmental background feature information so that the network model pays more attention to the feature information with larger weight and ignores the feature information with smaller weight during the training process, effectively reducing the interference of environmental factors (dense distribution of fruits, leaf occlusion and fruit stacking) and improving the accuracy of strawberry image segmentation. The experimental results show that the proposed method can accurately segment strawberry images with different maturity, the average pixel accuracy and average intersection ratio of the model are 90.9% and 83.05%, respectively, and the FPS is 7.67.

The method is based on deep learning to extract digital information of strawberries, which accurately segmented images of strawberries with different ripeness. This method could be applied to other fruits and vegetables (tomatoes, cucumbers, etc.) after re-collecting their sample data, building the training data set and training the model to complete the image segmentation and ripeness judgment. This work provides a real case study of utilizing machine learning in digital agriculture practice.

Author Contributions: C.C., J.T. and P.Z. conceived and designed the manuscript. C.C., J.T., P.Z., Y.Y. and J.Z. analyzed the data and wrote the paper. Y.Y. and J.Z. reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Department of Science and Technology of Jilin Province, (20200402116NC).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data, models, and code generated or used during the study appear in the submitted article.

Acknowledgments: We are grateful to Changchun Guoxin Modern Agriculture Com-pany for providing the experimental site.

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

References

1. Shu, R.; Jiao, J.; Zang, C.J.; Liu, S.J.; Sun, Y.N.; Yue, L.X. The current situation of China's strawberry industry and development suggestions. *China Fruit Veg.* **2019**, *39*, 57–59. [CrossRef]
2. Liu, J.Z.; Wu, S. Research progress of mechanized technology and equipment for the whole production of strawberry. *Trans-Actions Chin. Soc. Agric. Mach.* **2021**, *52*, 1–16. [CrossRef]
3. Wang, H.; Peng, J.; Xie, C.; Bao, Y.; He, Y. Fruit Quality Evaluation Using Spectroscopy Technology: A Review. *Sensors* **2015**, *15*, 11889–11927. [CrossRef] [PubMed]
4. Raj, R.; Cosgun, A.; Kulić, D. Strawberry Water Content Estimation and Ripeness Classification Using Hyperspectral Sensing. *Agronomy* **2022**, *12*, 425. [CrossRef]
5. Su, Z.; Zhang, C.; Yan, T.; Zhu, J.; Zeng, Y.; Lu, X.; Gao, P.; Feng, L.; He, L.; Fan, L. Application of Hyperspectral Imaging for Maturity and Soluble Solids Content Determination of Strawberry With Deep Learning Approaches. *Front. Plant Sci.* **2021**, *12*, 1897. [CrossRef] [PubMed]
6. Guo, Y.; Liu, Y.; Oerlemans, A.; Lao, S.; Wu, S.; Lew, M.S. Deep learning for visual understanding: A review. *Neurocomputing* **2016**, *187*, 27–48. [CrossRef]
7. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
8. Zheng, T.X.; Jiang, M.Z.; Feng, M.C. A review of vision-based target recognition and localization methods for picking robots. *Chin. J. Sci. Instrum.* **2021**, *42*, 28–51. [CrossRef]

9. Chen, S.W.; Shivakumar, S.S.; Dcunha, S.; Das, J.; Okon, E.; Qu, C.; Taylor, C.J.; Kumar, V. Counting Apples and Oranges With Deep Learning: A Data-Driven Approach. *IEEE Robot. Autom. Lett.* **2017**, *2*, 781–788. [CrossRef]
10. Liu, X.; Chen, S.W.; Aditya, S.; Sivakumar, N.; Dcunha, S.; Qu, C.; Taylor, C.J.; Das, J.; Kumar, V. Robust Fruit Counting: Combining Deep Learning, Tracking, and Structure from Motion. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 1045–1052. [CrossRef]
11. Chen, S.; Song, Y.; Su, J.; Fang, Y.; Shen, L.; Mi, Z.; Su, B. Segmentation of field grape bunches via an improved pyramid scene parsing network. *Int. J. Agric. Biol. Eng.* **2021**, *14*, 185–194. [CrossRef]
12. Chen, J.; Han, M.N.; Lian, Y.; Zhang, S. Image segmentation of rice seeds containing heterogeneous rice based on U-Net model. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 174–180. [CrossRef]
13. Majeed, Y.; Zhang, J.; Zhang, X.; Fu, L.; Karkee, M.; Zhang, Q.; Whiting, M.D. Apple Tree Trunk and Branch Segmentation for Automatic Trellis Training Using Convolutional Neural Network Based Semantic Segmentation. *IFAC-PapersOnline* **2018**, *51*, 75–80. [CrossRef]
14. Li, J.; Tang, Y.; Zou, X.; Lin, G.; Wang, H. Detection of Fruit-Bearing Branches and Localization of Litchi Clusters for Vision-Based Harvesting Robots. *IEEE Access* **2020**, *8*, 117746–117758. [CrossRef]
15. Peng, H.; Xue, C.; Shao, Y.; Chen, K.; Xiong, J.; Xie, Z.; Zhang, L. Semantic Segmentation of Litchi Branches Using DeepLabV3+ Model. *IEEE Access* **2020**, *8*, 164546–164555. [CrossRef]
16. Wang, C.; Du, P.; Wu, H.; Li, J.; Zhao, C.; Zhu, H. A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net. *Comput. Electron. Agric.* **2021**, *189*, 106373. [CrossRef]
17. Pérez-Borrero, I.; Marín-Santos, D.; Gegúndez-Arias, M.E.; Cortés-Ancos, E. A fast and accurate deep learning method for strawberry instance segmentation. *Comput. Electron. Agric.* **2020**, *178*, 105736. [CrossRef]
18. Perez-Borrero, I.; Marin-Santos, D.; Vasallo-Vazquez, M.J.; Gegundez-Arias, M.E. A new deep-learning strawberry instance segmentation methodology based on a fully convolutional neural network. *Neural Comput. Appl.* **2021**, *33*, 15059–15071. [CrossRef]
19. Afzaal, U.; Bhattarai, B.; Pandeya, Y.R.; Lee, J. An Instance Segmentation Model for Strawberry Diseases Based on Mask R-CNN. *Sensors* **2021**, *21*, 6565. [CrossRef] [PubMed]
20. Long, J.H.; Zhao, C.J.; Lin, S.; Guo, W.Z.; Wen, C.W.; Zhang, Y. Improved Mask R-CNN for tomato fruit segmentation with different ripeness in greenhouse environment. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 100–108. [CrossRef]
21. Ilyas, T.; Khan, A.; Umraiz, M.; Jeong, Y.; Kim, H. Multi-Scale Context Aggregation for Strawberry Fruit Recognition and Disease Phenotyping. *IEEE Access* **2021**, *9*, 124491–124504. [CrossRef]
22. Yu, Y.; Zhang, K.; Yang, L.; Zhang, D. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Comput. Electron. Agric.* **2019**, *163*, 104846. [CrossRef]
23. Jia, Z.W.; Yao, S.; Zhang, R.Y.; Wang, R.B.; Zhang, J. Fruit recognition of greenhouse strawberries based on improved U-Net network. *J. Shanxi Agric. Univ. (Nat. Sci. Ed.)* **2022**, *42*, 120–128. [CrossRef]
24. Ge, Y.; Xiong, Y.; From, P.J. Instance Segmentation and Localization of Strawberries in Farm Conditions for Automatic Fruit Harvesting. *IFAC-PapersOnline* **2019**, *52*, 294–299. [CrossRef]
25. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [CrossRef]
26. Zhu, F.; Zheng, Z. Image and convolutional neural network based growth potential assessment of Phalaenopsis seedlings. *Trans. Chin. Soc. Agric. Eng.* **2020**, *36*, 185–194. [CrossRef]
27. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmenta-tion. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
28. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587.
29. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807. [CrossRef]
30. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
31. Yang, L.; Zhang, R.Y.; Li, L.; Xie, X. Simam: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual, 18–24 July 2021; pp. 11863–11874.
32. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 11531–11539. [CrossRef]
33. 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. 3–9.
34. Luo, H.; Zhang, Y. A review of image semantic segmentation based on deep networks. *Acta Electron. Sin.* **2019**, *47*, 2211–2220. [CrossRef]
35. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 234–241. [CrossRef]

36. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890. Available online: https://jiaya.me/papers/PSPNet_cvpr17.pdf (accessed on 1 June 2022).
37. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep High-Resolution Representation Learning for Human Pose Estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 5686–5696. [CrossRef]

agronomy

MDPI

Article

Development of Deep Learning Methodology for Maize Seed Variety Recognition Based on Improved Swin Transformer

Chunguang Bi [1,2], Nan Hu [1], Yiqiang Zou [1], Shuo Zhang [1], Suzhen Xu [1] and Helong Yu [1,2,*]

[1] College of Information Technology, Jilin Agricultural University, Changchun 130118, China
[2] Institute for the Smart Agriculture, Jilin Agricultural University, Changchun 130118, China
* Correspondence: yuhelong@jlau.edu.cn; Tel.: +86-135-0082-8956

Abstract: In order to solve the problems of high subjectivity, frequent error occurrence and easy damage of traditional corn seed identification methods, this paper combines deep learning with machine vision and the utilization of the basis of the Swin Transformer to improve maize seed recognition. The study was focused on feature attention and multi-scale feature fusion learning. Firstly, input the seed image into the network to obtain shallow features and deep features; secondly, a feature attention layer was introduced to give weights to different stages of features to strengthen and suppress; and finally, the shallow features and deep features were fused to construct multi-scale fusion features of corn seed images, and the seed images are divided into 19 varieties through a classifier. The experimental results showed that the average precision, recall and F1 values of the MFSwin Transformer model on the test set were 96.53%, 96.46%, and 96.47%, respectively, and the parameter memory is 12.83 M. Compared to other models, the MFSwin Transformer model achieved the highest classification accuracy results. Therefore, the neural network proposed in this paper can classify corn seeds accurately and efficiently, could meet the high-precision classification requirements of corn seed images, and provide a reference tool for seed identification.

Keywords: corn seeds; image identification; multi-scale feature fusion; deep learning; machine vision

Citation: Bi, C.; Hu, N.; Zou, Y.; Zhang, S.; Xu, S.; Yu, H. Development of Deep Learning Methodology for Maize Seed Variety Recognition Based on Improved Swin Transformer. *Agronomy* **2022**, *12*, 1843. https://doi.org/10.3390/agronomy12081843

Academic Editor: Wen-Hao Su

Received: 3 July 2022
Accepted: 1 August 2022
Published: 4 August 2022

1. Introduction

In many countries, corn is not only one of the important food crops, but it also plays an important role in feed production, industrial raw materials, and bioenergy [1,2]. Improving maize production technology can effectively promote the high-quality development of agriculture and drive the growth of the national economy. In the process of agricultural production, seeds are the most basic means of production and the "chip" of agriculture [3,4]. At present, in order to meet the needs of grower recognition of varieties and the different regional environment planting needs, breeders have developed a considerable number of maize varieties. Although these varieties meet the market demand, the risk of mixing seeds is raised. For example, in the process of seed production, processing, storage and transportation, the quality of seed has not been completely effective in detection and control; or in the process of selling seeds, some individuals or enterprises use low-quality corn varieties to fake high-quality corn varieties to make huge profits [5]. All these behaviors will eventually lead to the low purity of seeds. However, the poor purity of maize seeds would potentially reduce the germination rate, seedling rate and disease resistance of maize, evenly affecting the yield of crops and the income of farmers [6]. The accurate identification of seed varieties plays an important role in reducing seed mixing, improving seed purity and ensuring market circulation order. Therefore, it is urgent to explore a nondestructive, efficient and environmental friendly identification method to select high-quality maize seed varieties, provide pure maize seeds for the market and consumers, enhance the international competitiveness of one country's agricultural products and ensure the sustainable development of agricultural production.

The traditional identification methods of seed purity include the morphological inspection method, field planting inspection method, chemical identification method and electrophoresis technology inspection method [7–10], however, the above methods are time-consuming, cumbersome in their identification processes and need professional personnel, so they are not suitable for mass analysis and nondestructive testing of seeds. In recent years, hyperspectral imaging technology [11–14] has been widely studied in seed purity detection, which has the characteristics of combining image and spectral information. For example, Wang [15] combined hyperspectral spectral data with image texture features and used least squares support vector machine to classify different corn seed varieties. Based on the two data fusion methods, the classification accuracy of 88.889% is obtained, and the classification effect is better. The results show that hyperspectral imaging technology has the potential of online and real-time variety classification. Xia [16] et al. collected hyperspectral images (400–1000 nm) of corn seeds of 17 varieties, and extracted 14 features including spectral features and imaging features from the hyperspectral images, and a LS-SVM classification model based on MLDA wavelength selection algorithm is proposed, which achieves high classification accuracy. This method can be effectively used for seed classification and identification. Zhang [17] combined hyperspectral imaging with deep convolutional neural network (DCNN) to classify the endosperm side average spectra extracted from four different varieties of maize seeds, and compared DCNN, K nearest neighbor (KNN), and Support Vector Machine (SVM) performance of three networks. In most cases, DCNN outperforms the other two networks in all aspects, achieving 93.3% accuracy. But hyper-spectrometer equipment is expensive, which is not conducive to practical application. With the rapid development of machine vision technology, more and more researchers rely on extracting seed characteristics, such as color, size, texture, and so on, and combined with classifier to classify seed varieties. For example, Wang [18] and others extracted the color and geometric features of corn seeds, reduced the dimension of the extracted features by using the principal component analysis method, and used them as the input of BP neural network for variety recognition. The comprehensive recognition rate of this method can reach more than 97%, which shows that it is feasible to use seed characterization for variety recognition and identification. Kantip [19] et al. proposed a method combining the color and texture features of corn seeds with support vector machine classifier to classify more than ten categories of seed defects. In tens of thousands of sample images, the accuracy of normal seed category is 95.6%, and the accuracy of defective seed category is 80.6%. This method provides useful information for the future development of seed quality identification. However, in the process of artificial feature selection and extraction, the operation steps are cumbersome, the time cost is high, and the recognition accuracy is greatly affected by the error of feature extraction. In summary, these seed inspection methods are difficult to meet the need for real-time seed classification during seed processing. Therefore, there is a need to investigate a nondestructive, efficient and end-to-end method for classifying maize seed varieties.

Deep learning [20–22] has continuously made breakthroughs in image fields such as image classification [23–25], object detection [26–28] and semantic segmentation [29–31] by virtue of its strong learning ability and wide coverage. Solve the tedious feature extraction, and the features extracted by deep learning have more generalization ability. The seed variety classification model based on convolutional neural network (CNN) has also become the focus of many scholars' research [32,33]. However, maize seeds of different varieties are highly similar in appearance, and the spatial perception of CNN is localized and cannot model the long-distance dependencies within the images, which has limited ability to extract similar features and cannot achieve accurate classification of maize seeds.

In order to extract the characteristic information of corn seeds more comprehensively and improve the accuracy of identification of corn seed varieties. In this study, a corn seed classification model based on the Swin Transformer is proposed. The self-attention mechanism of this model can efficiently extract image information. At the same time, the model is improved for the characteristics of maize seeds and images, so that it can effectively learn

the detailed differences between different varieties of maize seeds, to provide new ideas for seed classification to achieve refinement.

2. Materials and Methods

2.1. Image Acquisition and Preprocessing

2.1.1. Data Source and Acquisition

In this study, 19 different maize seeds were selected, including ChunQiu313, FuLai135, HeYu160, HuiHuang103, HuiHuang109, HX658, JH115, JH122, JinHui579, JiuDan21, JLAUY205, JLAUY1825, JLAUY1859, JX108, JX317, LuYu711, TianFeng8, TongDan168, XinGong379 (Figure 1). In the process of sample selection, seeds with plump grains and uniform shape were manually screened as experimental samples, and the varieties to which they belonged were identified by experts. Then the image acquisition work is carried out. To ensure the randomness of the data, the canon 70D camera is used to capture the corn seed image under the indoor natural light. The acquisition equipment is shown in Figure 2.

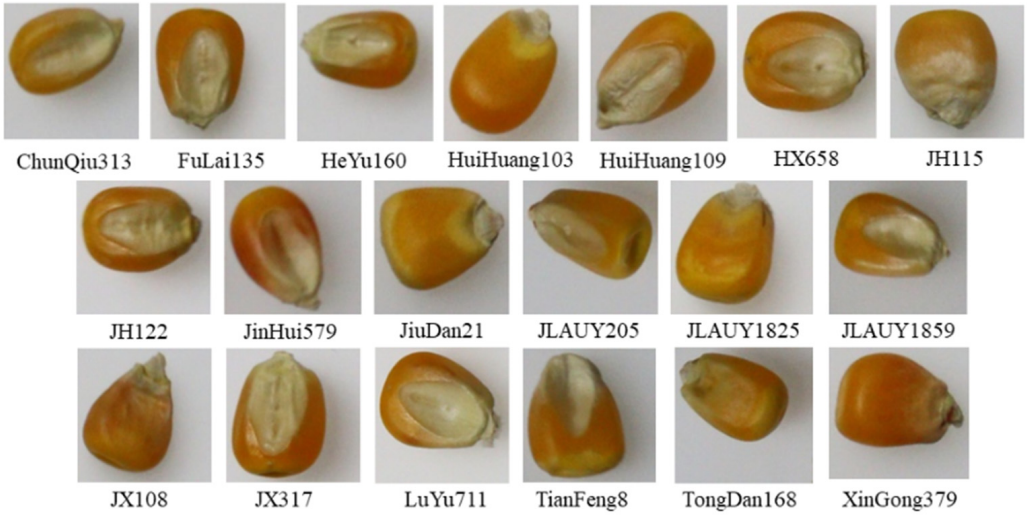


Figure 1. Image of maize seed varieties.

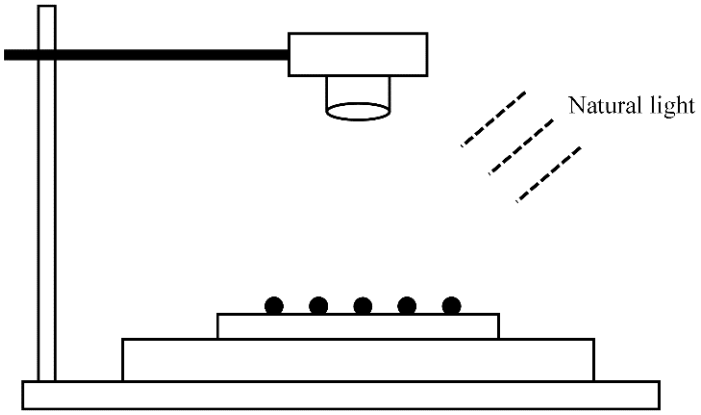


Figure 2. Corn seed image acquisition system.

2.1.2. Image Preprocessing

The research on Maize Variety identification is the research on the authenticity and purity of maize seeds. Because the seed purity is composed of the authenticity of a single seed, the single seed identification method is used to identify the seed varieties [34]. The image containing multiple corn seeds needs to be cut (Figure 3). Firstly, the image is denoised by Gray processing and Gaussian filter [35], and then Thresholding [36]. The edge of the corn seed is obtained by the contour extraction algorithm, and then the coordinates of the center

point and four vertices of the corn seed are obtained by the minimum circumscribed matrix method. Finally, the samples completely containing corn seeds were obtained by cutting, and a total of 6423 original images were obtained.

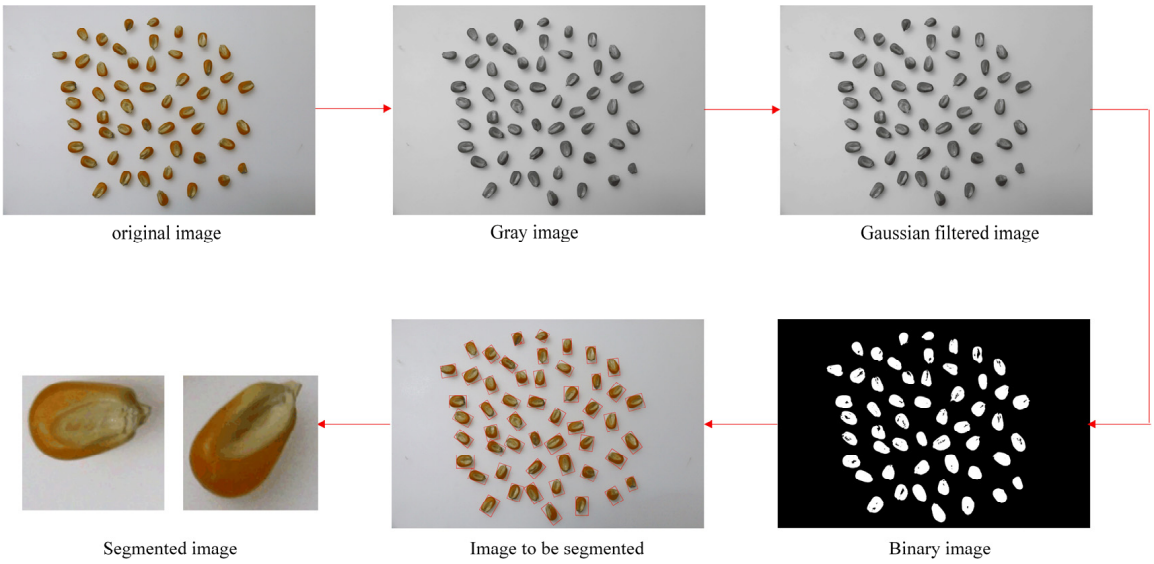


Figure 3. Corn seed image cutting processing.

The sample distribution of the dataset is shown in Figure 4, where 0 is ChunQiu313, 1 is FuLai135, 2 is HeYu160, 3 is HuiHuang103, 4 is HuiHuang109, 5 is HX658, 6 is JH115, 7 is JH122, 8 is JinHui579, 9 is JiuDan21, 10 is JLAUY205, 11 is JLAUY1825, 12 is JLAUY1859, 13 is JX108, 14 is JX317, 15 is LuYu711, 16 is TianFeng8, 17 is TongDan168, and 18 is XinGong379.

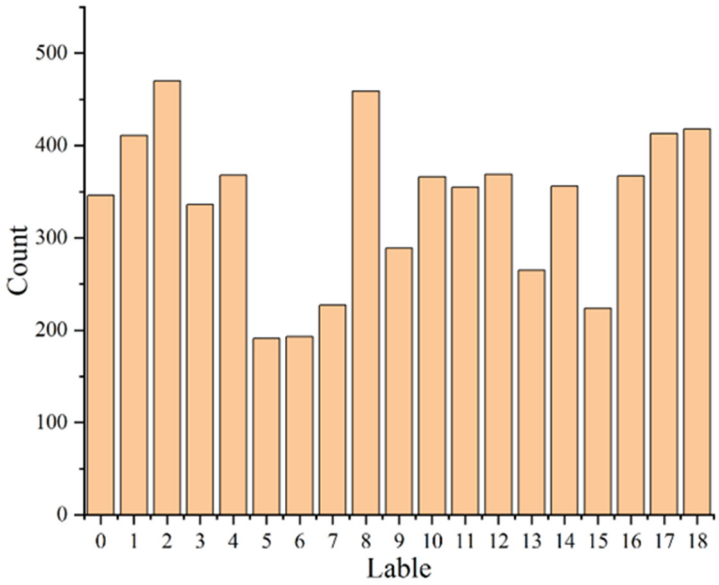


Figure 4. Data distribution and labeling.

It can be seen from Figure 4 that the total number of samples in the data set is insufficient and the distribution of various samples is uneven, which will affect the classification effect of the model [37,38]. To improve the generalization ability and robustness of the model, data enhancement methods such as vertical rotation, random brightness and salt and pepper noise are used to increase the number of data sets.

Figure 5 shows an example of image enhancement. Taking LuYu711 as an example, 224 original images are enhanced to 896 by random rotation and vertical flip, and then the

images are expanded from 896 to 1563 by Gaussian blur, random brightness and salt and pepper noise.

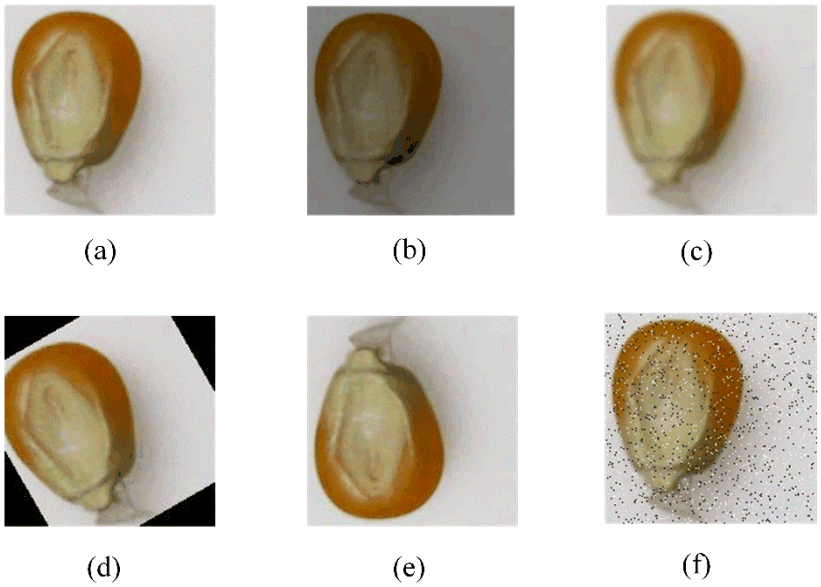


Figure 5. Image enhancement effect: (**a**) original image (**b**) Random brightness (**c**) Gaussian blur (**d**) Random rotation (**e**) Vertical rotation (**f**) Salt and pepper noise.

As shown in Table 1, all kinds of samples in the expanded data set are basically balanced, with a total of 32,500. Then, 80% of the data set is randomly selected as the training set and 20% as the test set. This paper adopts the 5-fold cross verification method [39], that is, 20,790 pictures are extracted from the training set as training data and 5197 pictures as verification data each time, and the verification data is not repeated.

Table 1. Profile of samples.

Seed Category	Number of Original Samples	Number of Enhanced Samples	Label
ChunQiu313	346	1730	0
FuLai135	411	1631	1
HeYu160	470	1882	2
HuiHuang103	336	1680	3
HuiHuang109	368	1843	4
HX658	191	1329	5
JH115	193	1351	6
JH122	227	1580	7
JinHui579	459	1836	8
JiuDan21	289	1728	9
JLAUY205	366	1830	10
JLAUY1825	355	1774	11
JLAUY1859	369	1845	12
JX108	265	1847	13
JX317	356	1773	14
LuYu711	224	1563	15
TianFeng8	367	1835	16
TongDan168	413	1703	17
XinGong379	418	1740	18
Total	6423	32,500	/

2.2. Model Building

2.2.1. Transformer Model

The transformer model [40] was first applied in the field of natural language processing. It adopts encoder decoder architecture. Both are stacked by multi head attention layer, and feed forward network connection layer. Skip, connect and layer normalization layers are added behind each sub-layer. The core concept of the transformer is the multi head attention mechanism [41], which is connected by multiple self-attention. The structure is shown in Figure 6. The attention mechanism [42] actually wants the computer to imitate the human visual system and efficiently pay attention to the more critical information in the task goal. Self-attention first needs to set three trainable weight matrices Q, K and V; the input vector is multiplied by Q matrix, K matrix and V matrix to obtain the query vector, key vector and value vector, and then use the Scaled dot-product attention to calculate the attention weight,

$$self\_attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

where, Q, K and V represent query, key and value matrices, respectively; The vector dimension of each key is d_k,The dimension of the value is d_v. Finally, connect multiple self-attention mechanisms to form a multi head attention mechanism, and its calculation formula is as follows:

$$\begin{gathered} MultiHead(Q, K, V) = Concat(head_1, \ldots, head_h)W^o \\ head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \end{gathered} \tag{2}$$

where, W^o is the weight matrix, W_i^Q, W_i^K, W_i^V is the conversion matrix of Q, K and V, respectively; Concat means to merge the attention information of all headers.

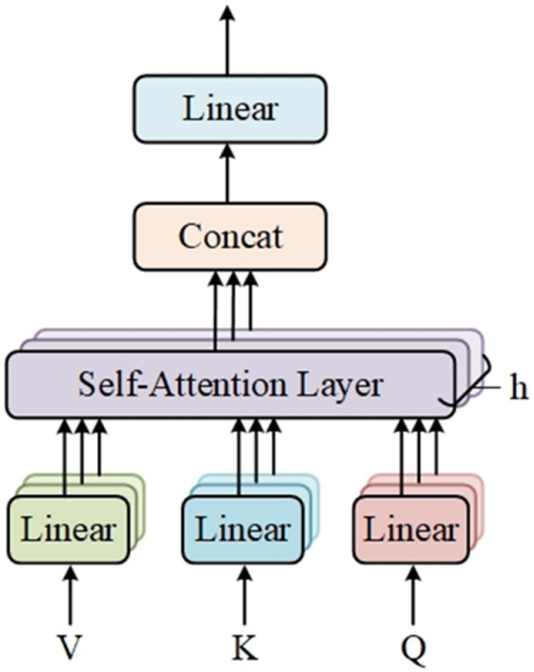


Figure 6. Structure of multi head attention mechanism.

With the continuous advancement of research, some scholars have migrated transformer technology to the field of computer vision [43,44]. Compared with convolutional neural network (CNN), the Transformer's self-attention is not limited by local interaction, so that the model can not only be parallelized training, but also mine long-distance dependencies and extract more powerful features. Vision Transformer (ViT) [45,46] is the first work of transformer to replace standard convolution in machine vision. But the ViT model

still has shortcomings such as being unable to model the spatial information of pictures, requiring large computing resources, and low learning efficiency. When the amount of data is not large enough, its advantages cannot be reflected.

2.2.2. Swin Transformer Model

Aiming at some problems of Transformer model in CV field, Microsoft proposed an improved model-Swin Transformer model [47,48]. It adopts a hierarchical construction method, which not only has the advantages of CNN processing large-size images, but also has the advantages of using moving windows to establish long-range dependencies, which solves the problems of computational complexity and lack of information interaction between groups. The Swin-T network structure is shown in Figure 7.

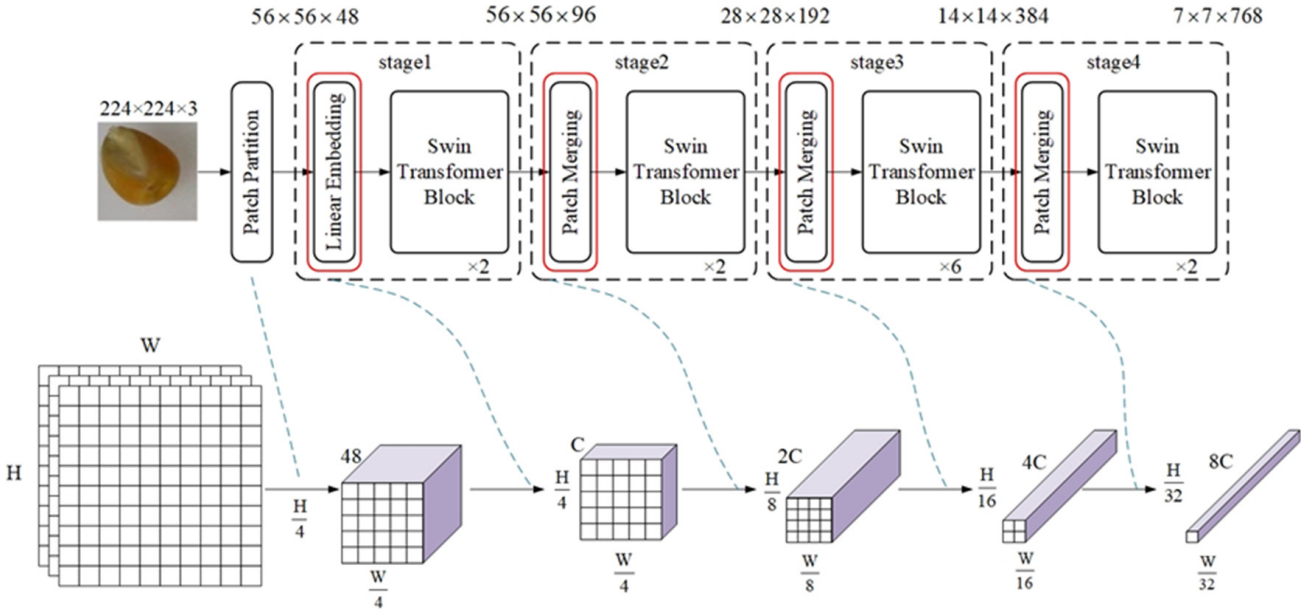


Figure 7. Swin Transformer model structure diagram.

Swin-T consists of four stages, and each stage reduces the resolution of the input feature map to expand the receptive field layer by layer. Similar to ViT, the image is first input into the Patch Partition module to divide the image into non-overlapping image blocks, each divided image block is regarded as a token, and the flattening operation is performed in the channel direction. The Linear Embedding module then uses linear variation to map it into a vector of dimension C (Figure 8). Each Swin Transformer module consists of two Blocks.

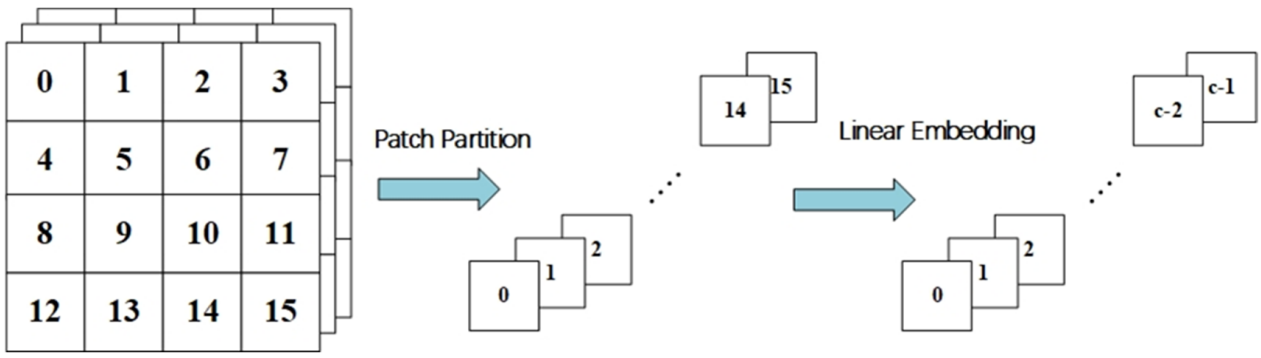


Figure 8. Patch Embedding module.

2.2.3. Swin Transformer Block

The core part of Swin-T is the Swin Transformer block, the detailed structure is shown in Figure 9. This module is a cascade of two multi-head attention modules, consisting of Windowed Multi-Head Self-Attention (W-MSA), Shifted Windowed Multi-Head Self-Attention (SW-MSA) and Multilayer Perceptron (MLP) [49].The LayerNorm layer is used before each MSA module and each MLP to make the training more stable and connected by residual after each module. Expressed as:

$$\begin{aligned} \hat{X}^l &= W - MSA\left(\mathrm{LN}\left(X^{l-1}\right)\right) + X^{l-1} \\ X^l &= MLP\left(LN\left(\hat{X}^l\right)\right) + \hat{X}^l \\ \hat{X}^{l+1} &= SW - MSA\left(LN\left(X^l\right)\right) + X^l \\ X^{l+1} &= MLP\left(LN\left(\hat{X}^{l+1}\right)\right) + \hat{X}^{l+1} \end{aligned} \tag{3}$$

where, X^l and $\hat{X}^l$ denote the output features of the two self-attention modules and the MLP module in block L, respectively.

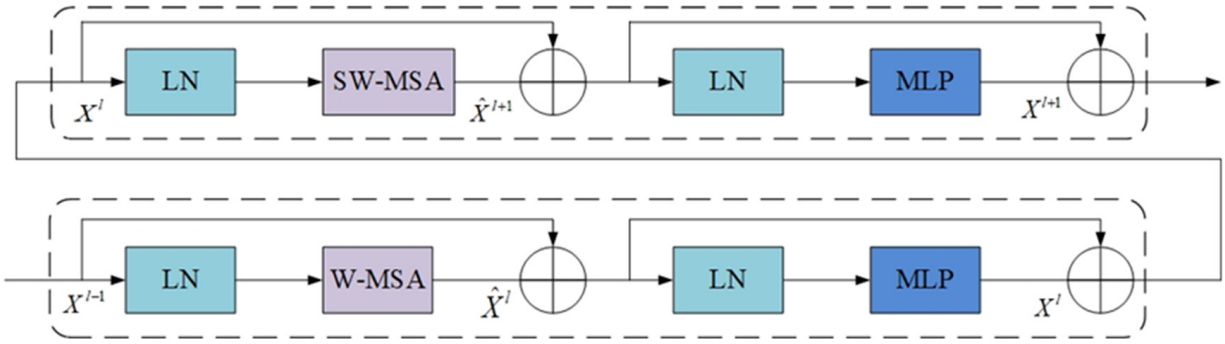


Figure 9. Swin Transformer module network structure.

2.2.4. W-MSA Module and SW-MSA Module

Different from the multi-head self-attention used in the ViT model, the window multi-head self-attention (W-MSA) in the Swin-T model first divides the input image into non-overlapping windows, and then the pixels in each window can only be the inner product is performed with other pixels in the window to obtain information, so that the computational complexity of W-MSA is linearly related to the image size. In this way, the computational complexity of the network can be greatly reduced, thereby improving the computational efficiency of the network. The computational complexity of *MSA* and W-*MSA* are:

$$\begin{aligned} \Omega(\mathrm{MSA}) &= 4\mathrm{hwC}^2 + 2(\mathrm{hw})^2C \\ \Omega(W - MSA) &= 4hwC^2 + 2M^2hwC \end{aligned} \tag{4}$$

Among them, the former $h \times w$ has quadratic complexity, while the latter has linear complexity when M is fixed.

Although the computational complexity problem has been solved, the information interaction between windows can't be carried out, resulting in the inability to obtain more globally accurate information, which will affect the accuracy of the network. To achieve information exchange between different windows, Shifted Window Multi-Head Self-Attention (SW-MSA) [50] is introduced. Each loop image is simultaneously shifted to the left and up by a certain window, then the cyan and red regions in Figure 10c are shifted to the right and below of the image, respectively. The SW-MSA mechanism can complete the pixel self-attention calculation of the offset window, thereby indirectly increasing the receptive field of the network and improving the efficiency of information utilization. The specific operation is shown in Figure 10.

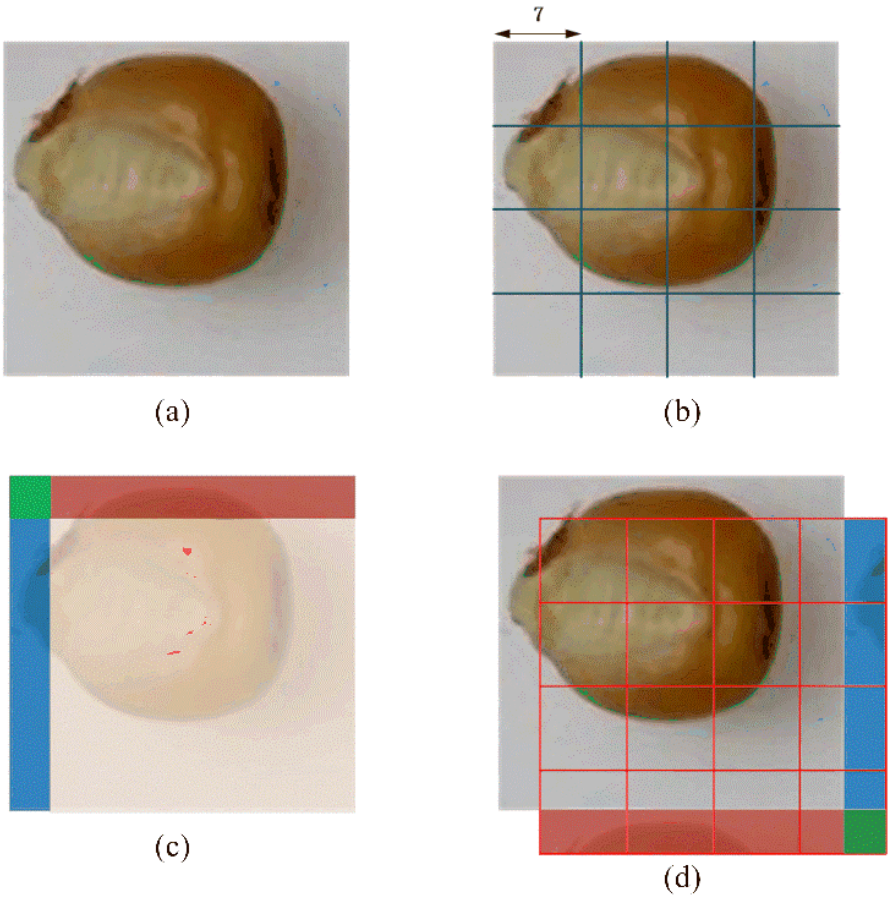


Figure 10. (**a**) Input image (**b**) Window segmentation of input image through W-MSA (**c**) Operation of moving window (**d**) Different window segmentation methods through SW-MSA.

2.3. MFSwin-T Model Design

Swin-T can automatically learn from images and find discriminative and representative features and obtain final classification results by training the model. However, considering that the characters of corn seeds of different varieties are not obvious and difficult to distinguish, and there are differences between different varieties not only in shape and outline, but also in texture and details. The Swin-T model usually only utilizes the high-order features of the last stage, and the high-order features reflect more abstract semantic features, such as contour, shape and other features. The low-level semantic information is usually in the first few layers of the deep network, which can reflect the subtle changes in texture and color of seeds. Shallow networks contain more features and have the ability to extract key subtle features [51]. Although the shallow network has a strong ability to represent detailed information, its receptive field is small and lacks generalization of the overall seed image. In response to this problem, this paper considers the fusion of shallow network and deep network features, so as to learn a more comprehensive and effective feature representation, and extract corn seed information more completely [52]. Therefore, this paper establishes a network MFSwin-Transformer based on multi-scale feature fusion to classify 19 different varieties of maize seeds. The specific structure is shown in the Figure 11.

As can be seen from the figure, MFSwin-T model is composed of a backbone network, multi-stage feature fusion module and classification and recognition layer.

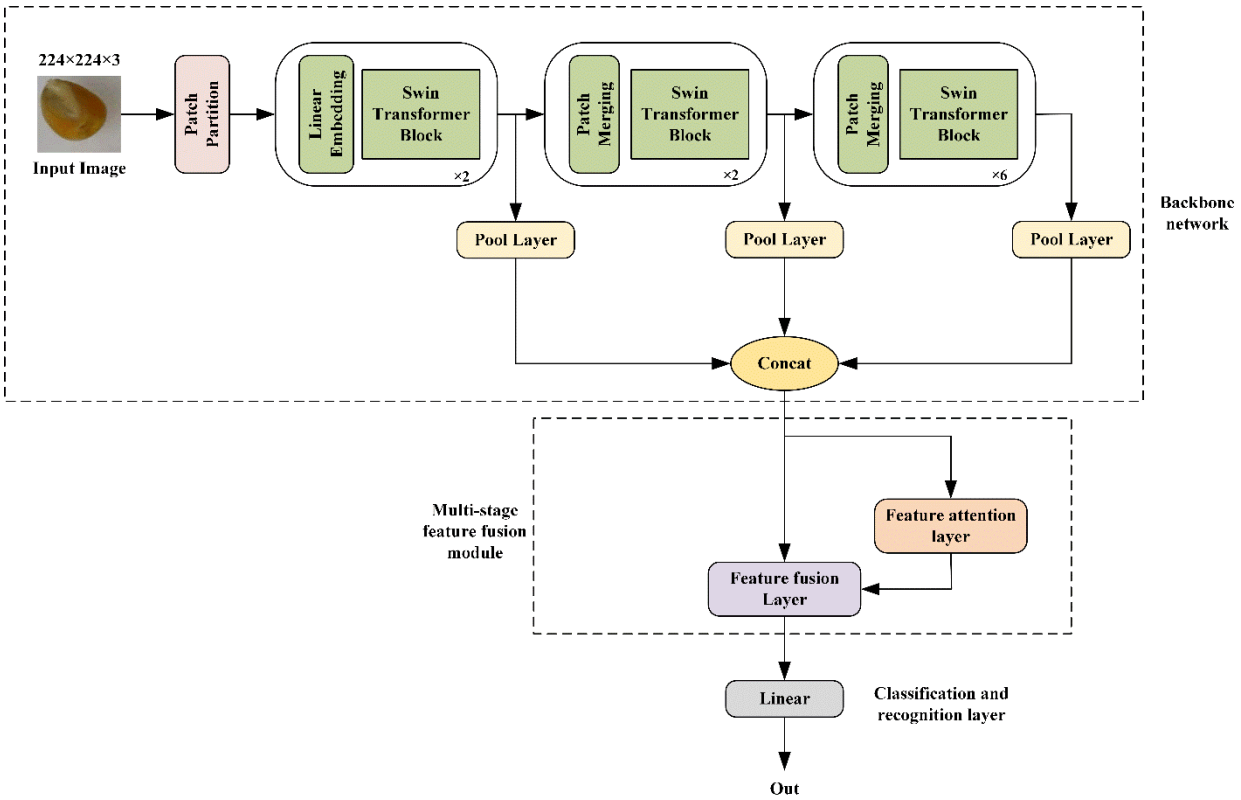


Figure 11. MFSwin- T network structure diagram.

2.3.1. Backbone Network

The original Swin-T backbone network includes four stages. However, due to the increase in multi-head attention and the number of channels, it still occupies a large amount of computing resources and cannot meet the needs of real-time identification of corn seeds on resource-constrained devices. To avoid too many parameters, the increase in computational complexity leads to overfitting, so this paper cuts off stage 4 in the backbone network of the Swin-T model. First, input the image $I \in R^{C\times H\times W}$ into the network, and obtain the characteristics of each stage $F_1 \in R^{C_1\times H_1\times W_1}$, where: C, H and W represent the input height, width and number of channels of the image, respectively; C_1, H_1 and W_1 represent the height, width and number of channels of features at each stage, respectively. Specifically, the input image $I \in R^{224\times 224\times 3}$ is sent to the backbone network, the output size of stage1 is $F_1 \in R^{56\times 56\times 96}$, the output of stage2 is $F_2 \in R^{28\times 28\times 192}$, and the output of stage3 is $F_3 \in R^{14\times 14\times 384}$. The Pool Layer is added after each stage, which consists of Linear, GELU, LayerNorm, and Average Pool. It mainly performs downsampling operation through the adaptive average pooling layer to unify the dimension of the feature vector, so that the dimensions of the features in different stages are the same. The unified dimensions are $F_{1,2,3} \in R^{1\times 1\times 384}$. Finally, the Concat method is used to splice the feature vectors obtained in different stages.

$$F_4 = Concat(F_1, F_2, F_3) \tag{5}$$

Among them, $F_4 \in R^{1\times 3\times 384}$ represents the dimension obtained by splicing the feature vectors of the three stages.

2.3.2. Multi-Stage Feature Fusion Module

The multi-stage feature fusion module takes the output of the backbone network feature extraction unit as input. To enable the model to selectively emphasize key features,

this paper designs a feature attention layer, which consists of Linear and Softmax (Figure 12). The specific calculation is as follows:

$$w = Softmax(Linear(Mean(F_4))) \tag{6}$$

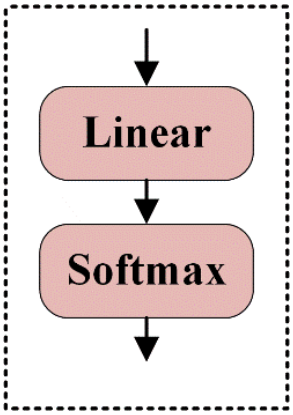


Figure 12. Feature attention layer.

Among them, $w \in R^{1\times3\times1}$, w represents the weight, Linear is a linear transformationMean ():$R^{1\times3\times384}$ -> $R^{1\times1\times384}$, represents the average operation of three eigenvectors. Softmax represents the normalized exponential function. Softmax's formula is as follows:

$$Softmax(z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \tag{7}$$

where z is a vector, z_i and z_j is one of the elements.

The stitched feature vectors are input into the feature attention layer, which gives different weights to the feature vectors in different stages to obtain w_1, w_2, w_3, Make the important features be assigned large weight to achieve enhancement, while other features are assigned small weight to achieve autonomous inhibition, so as to learn the key characteristics of different maize seed varieties. After the feature weight is obtained, it is passed into the feature fusion layer structure as shown in Figure 13, and the feature weight is multiplied by the feature vector element by element, and finally the multi-scale fusion feature F_final is obtained. The specific calculation is as follows:

$$F\_final = BL(Sum((BL(F_4) \times w))) \tag{8}$$

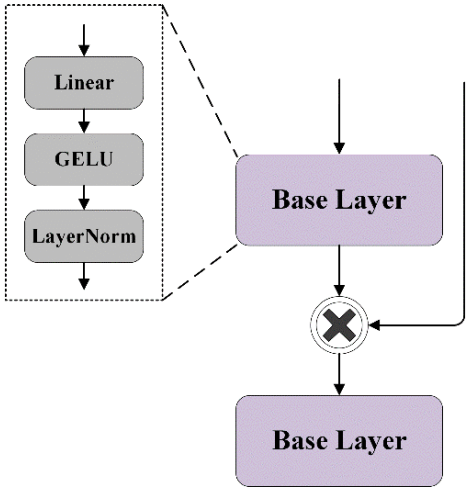


Figure 13. Feature fusion layer.

Among them, $F\_final \in R^{1\times1\times384}$, Sum ():$R^{1\times3\times384}$ -> $R^{1\times1\times384}$, BL is composed of Linear, GELU and LayerNorm.

2.3.3. Classification and Recognition Layer

Finally, F_final is input into the SoftMax classifier for classification and recognition of 19 different varieties of corn seeds. Its specific calculation is as follows:

$$Out = Softmax(Linear(F\_final)) \quad (9)$$

Among them, $F\_final \in R^{1\times19}$, represents the category of corn seeds.

3. Results

3.1. Experimental Environment and Hyperparameters

The classification model and comparative experiments proposed in this paper are all carried out in the Windows 10 operating environment, and the experimental models are implemented using the Pytorch deep learning framework. The specific experimental environment parameters are shown in Table 2.

Table 2. Experimental environment configuration.

Accessories	Operating System	Development Framework	Development Language	CUDA	GPU	RAM
Parameter	Windows10	Pytorch1.11.0	Python3.8.3	11.4	GeForce RTX 3080	32 G

In order to train the best model, we conduct multiple pre-experiments using the original Swin transformer model, and the hyperparameters of the model are set as follows: The size of the input corn seed image is set to 224 × 224, and the learning rate is an important hyperparameter for deep learning. If the learning rate is set too small, the convergence of the network will be too slow and the training time will be prolonged; if the learning rate is set too large, it will lead to the shock cannot converge. Therefore, during the pre-experiment, the learning rate was divided into four groups, 0.1, 0.01, 0.001 and 0.0001, respectively, and the batch size was set to 8, 16, 32 and 64. Through the comparative experiment, the learning rate is finally set to 0.0001 and the batch size is set to 32. At the same time, in the pre-experiment process, it was found that the model can converge when the number of Epochs is less than 50, so the number of experimental training iterations Epochs is set to 50. At the same time, the optimizer of the network selects the Adaptive Moment (Adam), and selects the Cross-Entropy loss function. The hyperparameters of the specific model are shown in Table 3.

Table 3. Parameters for model training.

Parameter.	Parameter Value
Epoch	50
Learning rate	0.0001
Batch-size	32
Optimizer	Adam

3.2. Evaluation Indicators

In the field of machine learning, confusion matrices are often used to compare the results of model classification in supervised learning. Each column of the matrix represents the predicted class, and each row represents the actual class. Taking the binary classification problem as an example, define the actual result as positive and the predicted result as positive, denoted as TP; if the actual result is negative, the predicted result is positive, denoted as FP; if the actual result is positive, the predicted result is negative, denoted as FN; The actual result is negative, the predicted result is negative, denoted as TN. The specific structure of the confusion matrix is shown in Table 4.

Table 4. Confusion matrix of binary classification problem.

Confusion Matrix		Actual Results	
		Positive	**Negative**
Forecast Results	Positive	TP	FP
	Negative	FN	TN

For the corn seed data set built in this paper, the confusion matrix is used to calculate the Accuracy, Precision, Recall and F1 score of each network as the index to evaluate the effect of the model on corn seed variety recognition. In a multi-class classification task, treat each breed individually as "positive" and all other breeds as "negative". To measure the performance of the entire network, the average precision and recall of 19 different varieties were calculated. Usually, Precision and Recall are a pair of contradictory indicators, so this paper uses the F1 score to calculate the weighted average of Precision and Recall. The higher the F1 score, the higher the model prediction accuracy and the better the performance. The calculation process of each evaluation index is shown in Table 5:

Table 5. Calculation formulas of each indicator.

Index	Formula	Significance
Accuracy	$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$	The number of correct predictions by the model, as a proportion of the total sample size
m-Precision	$m - Precision = \frac{TP}{TP+FP}$	Among the samples that the model predicts as positive, the model predicts the correct proportion of samples
Precision	$Precision = avg(\sum_{m=1}^{19} m - Precision)$	Average accuracy of 19 varieties
m-Recall	$m - Recall = \frac{TP}{TP+FN}$	In the samples whose true value is positive, the model predicts the correct proportion of samples
Recall	$Recall = \sum_{m=1}^{19} (m - Recall)$	Average recall rate of 19 varieties
F1-Score	$F1 = 2 \times \frac{Recall \times Precision}{Recall+Precision}$	F1 score is the harmonic average of accuracy and recall

At the same time, in order to measure the performance of the network, the parameter memory requirements and FLOPs are also used as the evaluation indicators of the model. The parameter memory demand is determined by the number of parameters. On the premise of meeting the task requirements, the smaller the parameter memory, the less the consumption of computer memory resources and the higher the applicability. FLOPs refer to the number of floating point calculations, that is, the time complexity of the model. The lower the FLOPs, the less computation the model needs and the shorter the network execution time. FPS refers to the number of corn seed images processed per second, which can measure the recognition speed of the model. The larger the FPS, the faster the inference speed of the classification model.

3.3. Analysis of Experimental Results and Evaluation of Network Performance

3.3.1. Deep Feature Extraction Experiment

This experiment uses the Swin Transformer model. First, input the corn seed image into the network to get 7×7 deep features with 768 channels. Then, through the classification layer, the probability of 19 categories is predicted. The deep feature extraction experiment is the Baseline experiment in this paper. Finally, the average accuracy on the corn seed image dataset is 92.9%.

3.3.2. Multi-Scale Feature Fusion Experiment

This experiment extracts multi-scale features of corn seed images. To prevent overfitting due to too many parameters, we cut off Stage 4 in the Swin Transformer model. First, the corn seed image is input into the network, and the adaptive average pooling method

is used to sample the features of the three stages, and the features are uniformly mapped to the channel dimension of 384. Then the features of the above three stages are spliced and fused on average to obtain multi-scale fusion features. Finally, the classifier is used for seed classification. The experimental accuracy is shown in the Table 6. The average accuracy on the corn seed image dataset is 94.8%, which is 1.9% higher than that of the Baseline experiment. The experimental results show that the fusion of shallow network features and deep network features is effective, the shallow feature resolution is higher, and it has a strong ability to represent seed detail information.

Table 6. Indicators of ablation experiments.

Evaluation Indicators	Method		
	Baseline	**Multiscale Features**	**Multi-Scale Features + Feature Attention**
Average Accuracy (%)	92.91	94.77	96.47
Average Precision (%)	93.36	94.26	96.53
Average Recall (%)	92.86	94.38	96.46
Average F1-Score (%)	93.00	94.27	96.47

3.3.3. Multi-Scale Feature Fusion Experiment with Feature Attention Layer

This experiment adds the feature attention layer based on the above experiment. Three feature weights can be obtained by passing the above average fused features through the attention layer, and then the feature weights are multiplied with the features of each stage to obtain the final multi-scale fused features. The experimental accuracy is shown in the Table 6. Compared with the 4.3.2 experiment, the accuracy is increased by 1.6%. The experimental results show that adding the feature attention layer can effectively integrate the features of different stages and improve the classification ability of the model for corn seed varieties.

3.4. Comparative Experiment and Analysis

To verify the effectiveness and superiority of the improved model proposed in this paper, the MFSwin-Transformer model is compared with some representative convolutional neural networks [53–55] under the same dataset, experimental environment and the same network parameter configuration. Comparative experiments are carried out, including AlexNet [56–58], Vgg16 [59–61], ResNet50 [62–64], Visio-Transformer, Swin-Transformer models. Table 7 lists the main parameters of some comparison networks.

Table 7. Main parameters of classical convolutional neural network.

AlexNet	**VGG16**	**ResNet50**
Layer1: 11×11, 96; 3×3 Maxpool	Layer1: $\begin{bmatrix} 3 \times 3, & 64 \\ 3 \times 3, & 64 \end{bmatrix}$; 2×2 Maxpool	Layer1: 7×7, 64; 3×3 Maxpool
Layer2: 5×5, 256; 3×3 Maxpool	Layer2: $\begin{bmatrix} 3 \times 3, & 128 \\ 3 \times 3, & 128 \end{bmatrix}$; 2×2 Maxpool	Layer2: $\begin{bmatrix} 1 \times 1, & 64 \\ 3 \times 3, & 64 \\ 1 \times 1, & 256 \end{bmatrix} \times 3$
Layer3: 3×3, 384	Layer3: $\begin{bmatrix} 3 \times 3, & 256 \\ 3 \times 3, & 256 \\ 3 \times 3, & 256 \end{bmatrix}$; 2×2 Maxpool	Layer3: $\begin{bmatrix} 1 \times 1, & 128 \\ 3 \times 3, & 128 \\ 1 \times 1, & 512 \end{bmatrix} \times 4$
Layer4: 3×3, 384	Layer4: $\begin{bmatrix} 3 \times 3, & 512 \\ 3 \times 3, & 512 \\ 3 \times 3, & 512 \end{bmatrix}$; 2×2 Maxpool	Layer4: $\begin{bmatrix} 1 \times 1, & 256 \\ 3 \times 3, & 256 \\ 1 \times 1, & 1024 \end{bmatrix} \times 6$
Layer5: 3×3, 256; 3×3 Maxpool	Layer5: $\begin{bmatrix} 3 \times 3, & 512 \\ 3 \times 3, & 512 \\ 3 \times 3, & 512 \end{bmatrix}$; 2×2 Maxpool	Layer5: $\begin{bmatrix} 1 \times 1, & 512 \\ 3 \times 3, & 512 \\ 1 \times 1, & 2048 \end{bmatrix} \times 3$
FC-1: 4096	FC-1: 4096	FC-1: 19
FC-2: 4096	FC-2: 4096	Classifier: Softmax
FC-3: 19	FC-3: 19	
Classifier: Softmax	Classifier: Softmax	

Figure 14 shows the relationship between the training loss value of each model and the iteration rounds. The loss curve represents the deviation between the predicted value and the real value of the network with the increase in iteration rounds. The smaller the loss value, the stronger the classification ability of the model and the smaller the probability of prediction error. It can be clearly seen from the figure that in the initial stage of training, the loss values of all networks are continuously decreasing, and eventually become stable without large fluctuations. Among them, the convergence speed of the MFSwin Transformer model is significantly better than other models. After the 13th epoch, the loss value of the MFSwin Transformer remained below 0.15, while the other models dropped below 0.15 after the 20th epoch. From the perspective of loss convergence, the training effect of the MFSwin Transformer model is optimal.

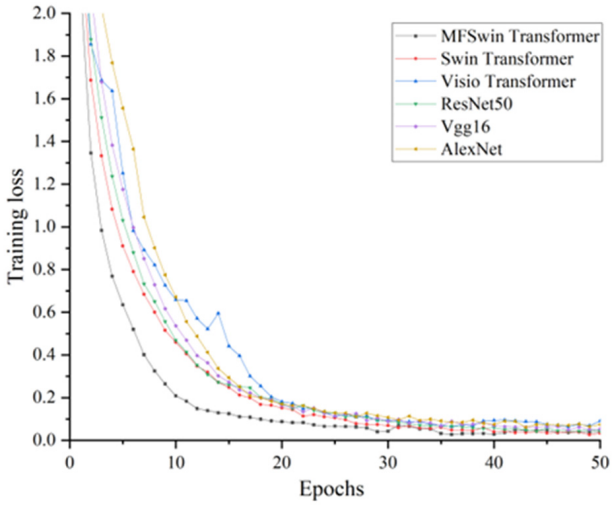


Figure 14. Relationship between model training loss and iteration rounds.

Figure 15 shows the relationship between the validation accuracy of each model and the number of iterations. The verification accuracy curve can describe the fluctuation of the recognition accuracy of the network as the number of iterations increases. As can be seen from the figure, the recognition accuracy of the MFSwin Transformer network has been continuously improved during the verification process. Compared with other comparative networks, the curve of the recognition accuracy of the MFSwin-T network is relatively smooth and stable, and there is no overfitting situation. The accuracy stabilizes above 90% after 13 epochs, while the other models only reach around 80% at the 13th epoch. It shows that the network has good recognition accuracy and generalization ability.

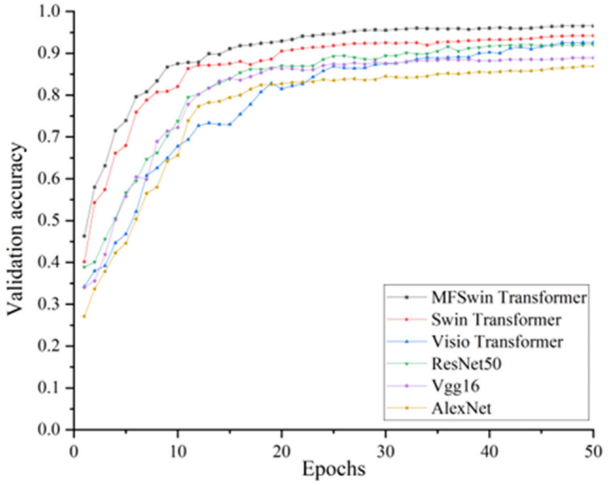


Figure 15. Relationship between model verification accuracy and iteration rounds.

Table 8 lists the classification performance of six different models, AlexNet, Vgg16, ResNet50, Swin Transformer, Visio Transformer, and MFSwin Transformer. Among them, the average accuracy, precision, recall, and F1-Score of the MFSwin-T model reached 96.47%, 96.54%, 96.46%, and 96.48%, respectively. It has better performance and higher accuracy than other models.

Table 8. Classification and comparison results of different networks.

Network	MFSwin-T	Swin-T	ViT	ResNet50	AlexNet	Vgg16
Average Accuracy (%)	96.47	92.91	92.00	91.29	88.75	85.95
Average Precision (%)	96.53	93.36	92.23	91.58	88.87	86.06
Average Recall (%)	96.46	92.86	92.06	91.31	88.82	86.06
Average F1-Score (%)	96.47	93.00	92.06	91.35	88.76	85.97

In addition, MFSwin-T has lower parameter memory and model complexity than other network models. As shown in Table 9, the computational complexity and parameter amount of the six models are shown. Under the same experimental environment and configuration, when the input corn seed image size is 224 × 224, the computational complexity of the MFSwin-T model is 3.783 G, and the parameter memory is 12.83 M. This is a reduction of 0.568 G and 14.66 M, respectively, compared to the network before the improvement. Meanwhile, compared with other networks, MFSwin Transformer has lower computational complexity and number of model parameters, where Vgg16 and Vit models have more than five times the computational complexity and seven times the number of parameters. And the image processing speed of MFSwin Transformer is also advantageous among all comparison models. It can be seen that MFSwin-T not only has faster network performance, but also can more accurately identify different varieties of corn seeds. These comparisons are enough to show the feasibility and superiority of the MFSwin-T model.

Table 9. Performance comparison results of different networks.

Network	FLOPs (G)	Params (M)	FPS
MFSwin-T	3.783	12.83	341.18
Swin-T	4.351	27.49	306.62
Vit	16.848	86.21	240.09
ResNet50	4.1095	23.55	304.65
AlexNet	0.711	57.08	365.76
Vgg16	15.480	134.34	294.80

In order to evaluate the performance of the network more comprehensively, this paper draws a comparison and analysis diagram of the confusion matrix to reflect the actual recognition of each variety by the MFSwin Transformer network. As shown in Figure 16, due to the similar characteristics between seeds of different varieties, some varieties have large errors, such as JX317 and TianFeng8, which are prone to misclassification. However, it has higher accuracy and more advantages than manual recognition. But in general, the confusion matrix and classification results prove that the model proposed in this paper has good recognition ability for 19 kinds of corn seed categories, and the seeds of most varieties can be correctly recognized, indicating that the model can lay a good foundation for the recognition of highly similar seeds.

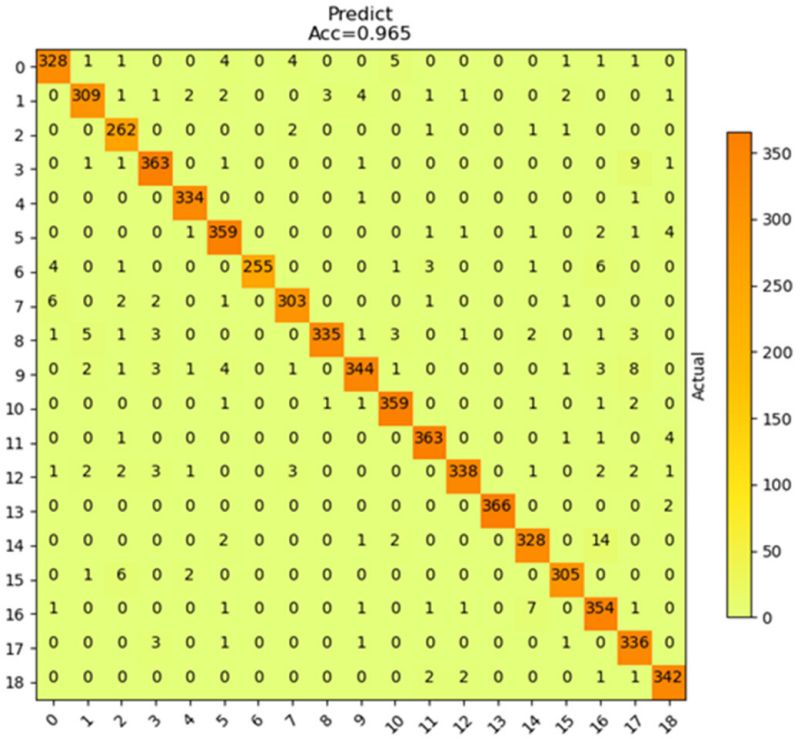


Figure 16. MFSwin-T Confusion matrix.

4. Discussion

In this work, we propose a non-destructive and efficient model for maize seed variety identification. First, we established a maize seed dataset with 19 varieties, and segmented the multi-seed images into images containing only single seeds by methods such as Gaussian filtering and contour extraction. In order to improve the generalization ability of the model, the data is augmented. According to the characteristics of different maize seed varieties with small differences in characteristics and details that cannot be ignored, this paper designs a maize seed variety recognition model MFSwin Transformer based on feature attention and multi-scale feature fusion. The model has high recognition accuracy and less parameters and has strong practicability. Maize seed varieties have the characteristics of small inter-class differences and large intra-class differences. Although a deeper network can obtain the outline and shape of seeds, it is easy to ignore some detailed features. Therefore, feature attention and multi-scale feature fusion are introduced to combine deep features and shallow features, so that the model can extract richer seed features. At the same time, to avoid overfitting caused by too many parameters, we cut off stage 4 in the backbone network. It can be seen from the ablation experiments that the improved method designed in this paper can improve the classification accuracy of the model and reduce the number of parameters and computational complexity of the model. However, it can be seen from the confusion matrix of the MFSwin Transformer model that some varieties are still classified incorrectly, which also reflects the need to further improve the ability of the model to extract fine features, so that it can provide better performance and accuracy under highly similar seed datasets. Because only low-cost digital cameras are needed to collect images, this method can be widely used and popularized in smart agriculture.

By comparing the performance of different models, the model proposed in this paper is better than other classic network models in terms of performance and recognition effect. Using deep learning to automatically extract image features is not only more effective but also avoids complex feature extraction in traditional recognition methods. However, this study only considers the classification effect of corn seed samples in the same year, and corn seeds of different years will have certain errors in color. In the follow-up study, the effects

of different years, climatic and environmental factors, and geographical factors on seed classification can be compared.

5. Conclusions

In this paper, a non-destructive identification method is proposed, which can automatically classify different varieties of corn seeds from images, thereby overcoming the issue of large errors and low efficiency from the traditional methods. The main contributions of this study are:

(1) Collect the image containing multiple corn seeds, denoise and segment the image by using Gray processing, Gaussian filter and minimum circumscribed matrix, and get the image containing only a single corn seed. Finally, a dataset of maize seeds containing 19 different varieties was constructed, containing a total of 32,500 images. (2) According to the highly similar characteristics of maize seeds among different varieties, this paper integrates the shallow and deep features, and proposes a neural network maize seed image classification model MFSwin transformer based on feature attention and multi-scale feature fusion. The models and methods proposed in this paper provide a new idea for seed classification. (3) For the maize seed variety identification task, the proposed model (MFSwin Transformer) in this paper achieves relatively high classification results, and its accuracy, recall, and F1-score are substantially improved compared with the original model Swin Transformer, with an average recognition accuracy of 96%. Meanwhile, the number of parameters is only half of the original model, and the computational complexity of the model has been reduced. These results demonstrate the advantages and certain potential of the method in seed classification, which can provide high-quality maize seeds for agricultural production.

However, there are still some problems in this method that need to be solved urgently. Through experiments, it is found that the classification ability of this model between highly similar seeds still needs to be improved. Future research will continue to optimize the model so that it can be provided in more complex data sets with higher accuracy, forming an online real-time identification system capable of identifying multiple similar seeds.

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

Funding: This research was funded by the Joint Fund Project of the National Natural Science Foundation of China (U19A2061) and the Big Data Technology and Smart Agriculture Team (20200301047RQ).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All relevant data are included in the manuscript. Raw images are available on request from the corresponding author.

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

References

1. García-Lara, S.; Serna-Saldivar, S.O.J.C. Corn History and Culture. *Corn* **2019**, 1–18. [CrossRef]
2. Aimin, C.; Jiugang, Z.; Hu, Z. Preliminary exploration on current situation and development of maize production in China. *J. Agric. Sci. Technol.* **2020**, *22*, 10.
3. Costa, C.J.; Meneghello, G.E.; Jorge, M.H. The importance of physiological quality of seeds for agriculture. *Colloquim Agrar.* **2021**, *17*, 102–119. [CrossRef]
4. Queiroz, T.; Valiguzski, A.; Braga, C. Evaluation of the physiological quality of seeds of traditional varieties of maize. *Revista da Universidade Vale do Rio Verde* **2019**, *17*, 20193215435.
5. Sun, J.; Zou, Y. Analysis on the Method of Corn Seed Purity Identification. *Hans J. Agric. Sci.* **2020**, *10*, 292–298.
6. TeKrony, D.M. Seeds: The delivery system for crop science. *Crop Sci.* **2006**, *46*, 2263–2269. [CrossRef]

7. Sundaram, R.; Naveenkumar, B.; Biradar, S. Identification of informative SSR markers capable of distinguishing hybrid rice parental lines and their utilization in seed purity assessment. *Euphytica* **2008**, *163*, 215–224. [CrossRef]
8. Ye-Yun, X.; Zhan, Z.; Yi-Ping, X. Identification and purity test of super hybrid rice with SSR molecular markers. *Rice Sci.* **2005**, *12*, 7.
9. Satturu, V.; Rani, D.; Gattu, S. DNA fingerprinting for identification of rice varieties and seed genetic purity assessment. *Agric. Res.* **2018**, *7*, 379–390. [CrossRef]
10. Pallavi, H.; Gowda, R.; Shadakshari, Y. Identification of SSR markers for hybridity and seed genetic purity testing in sunflower (*Helianthus annuus* L.). *Helia* **2011**, *34*, 59–66. [CrossRef]
11. Lu, B.; Dao, P.D.; Liu, J. Recent Advances of Hyperspectral Imaging Technology and Applications in Agriculture. *Remote Sens.* **2020**, *12*, 2659. [CrossRef]
12. Wang, C.; Liu, B.; Liu, L. A review of deep learning used in the hyperspectral image analysis for agriculture. *Artif. Intell. Rev.* **2021**, *54*, 5205–5253. [CrossRef]
13. ElMasry, G.; Mandour, N.; Al-Rejaie, S. Recent applications of multispectral imaging in seed phenotyping and quality monitoring—An overview. *Sensors* **2019**, *19*, 1090. [CrossRef] [PubMed]
14. Hong, W.; Kun, W.; Jing-zhu, W. Progress in Research on Rapid and Non-Destructive Detection of Seed Quality Based on Spectroscopy and Imaging Technology. *Spectrosc. Spectr. Anal.* **2021**, *41*, 52–59.
15. Wang, L.; Sun, D.; Pu, H. Application of Hyperspectral Imaging to Discriminate the Variety of Maize Seeds. *Food Anal. Methods* **2015**, *9*, 225–234. [CrossRef]
16. Xia, C.; Yang, S.; Huang, M. Maize seed classification using hyperspectral image coupled with multi-linear discriminant analysis. *Infrared Phys. Technol.* **2019**, *103*, 103077. [CrossRef]
17. Zhang, J.; Dai, L. Corn seed variety classification based on hyperspectral reflectance imaging and deep convolutional neural network. *Food Meas. Charact.* **2020**, *15*, 484–494. [CrossRef]
18. Wang, Y.; Liu, X.; Su, Q. Maize seeds varieties identification based on multi-object feature extraction and optimized neural network. *Trans. Chin. Soc. Agric. Eng.* **2010**, *26*, 199–204.
19. Kiratiratanapruk, K.; Sinthupinyo, W. Color and texture for corn seed classification by machine vision. In Proceedings of the 2011 International symposium on intelligent signal processing and communications systems (ISPACS), Chiang Mai, Thailand, 7–9 December 2011.
20. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
21. Stevens, E.; Antiga, L.; Viehmann, T. *Deep Learning with PyTorch*; Manning Publications: Greenwich, CT, USA, 2020.
22. Wani, J.A.; Sharma, S.; Muzamil, M. Machine learning and deep learning based computational techniques in automatic agricultural diseases detection: Methodologies, applications, and challenges. *Arch. Comput. Methods Eng.* **2021**, *29*, 641–677. [CrossRef]
23. Rawat, W.; Wang, Z. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Comput.* **2017**, *29*, 2352–2449. [CrossRef]
24. Jena, B.; Saxena, S.; Nayak, G.k. Artificial intelligence-based hybrid deep learning models for image classification: The first narrative review. *Comput. Biol. Med.* **2021**, *137*, 104803. [CrossRef] [PubMed]
25. Thenmozhi, K.; Reddy, U.S. Crop pest classification based on deep convolutional neural network and transfer learning. *Comput. Electron. Agric.* **2019**, *164*, 104906. [CrossRef]
26. Zhao, Z.-Q.; Zheng, P.; Xu, S. Object detection with deep learning: A review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. [CrossRef] [PubMed]
27. Chen, Y.; Wu, Z.; Zhao, B. Weed and corn seedling detection in field based on multi feature fusion and support vector machine. *Sensors* **2020**, *21*, 212. [CrossRef] [PubMed]
28. Hu, D.; Ma, C.; Tian, Z. Rice Weed detection method on YOLOv4 convolutional neural network. In Proceedings of the 2021 International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA), Xi'an, China, 28–30 May 2021.
29. Yu, C.; Wang, J.; Peng, C. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV 2018), Munich, Germany, 8–14 September 2018; pp. 325–341.
30. Giménez-Gallego, J.; González-Teruel, J.; Jiménez-Buendía, M. Segmentation of multiple tree leaves pictures with natural backgrounds using deep learning for image-based agriculture applications. *Appl. Sci.* **2019**, *10*, 202. [CrossRef]
31. Su, D.; Kong, H.; Qiao, Y. Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. *Comput. Electron. Agric.* **2021**, *190*, 106418. [CrossRef]
32. Gulzar, Y.; Hamid, Y.; Soomro, A.B. A Convolution Neural Network-Based Seed Classification System. *Symmetry* **2020**, *12*, 2018. [CrossRef]
33. Sabanci, K.; Aslan, M.F.; Ropelewska, E. A convolutional neural network-based comparative study for pepper seed classification: Analysis of selected deep features with support vector machine. *J. Food Process Eng.* **2022**, *45*, e13955. [CrossRef]
34. Hong, P.T.T.; Hai, T.T.T.; Hoang, V.T. Comparative study on vision based rice seed varieties identification. In Proceedings of the 2015 Seventh International Conference on Knowledge and Systems Engineering (KSE), Ho Chi Minh City, Vietnam, 8–10 October 2015.
35. Buades, A.; Coll, B.; Morel, J.-M. A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.* **2005**, *4*, 490–530. [CrossRef]
36. Szostek, K.; Gronkowska-Serafin, J.; Piórkowski, A. Problems of corneal endothelial image binarization. *Schedae Inform.* **2011**, *20*, 211.
37. Buda, M.; Maki, A.; Mazurowski, M.A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw.* **2018**, *106*, 249–259. [CrossRef] [PubMed]

38. Wan, X.; Zhang, X.; Liu, L. An Improved VGG19 Transfer Learning Strip Steel Surface Defect Recognition Deep Neural Network Based on Few Samples and Imbalanced Datasets. *Appl. Sci.* **2021**, *11*, 2606. [CrossRef]
39. Lopez-del Rio, A.; Nonell-Canals, A.; Vidal, D. Evaluation of cross-validation strategies in sequence-based binding prediction using deep learning. *J. Chem. Inf. Modeling* **2019**, *59*, 1645–1657. [CrossRef]
40. Vaswani, A.; Shazeer, N.; Parmar, N. Attention is all you need. *Adv. Neural Inf. Processing Syst.* **2017**, *30*, 1–11.
41. Xi, C.; Lu, G.; Yan, J. Multimodal sentiment analysis based on multi-head attention mechanism. In Proceedings of the 4th International Conference on Machine Learning and Soft Computing, Haiphong City, Vietnam, 17–19 January 2020.
42. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
43. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A survey on visual transformer. *arXiv* **2020**, arXiv:2012.12556.
44. Khan, S.; Naseer, M.; Hayat, M. Transformers in vision: A survey. *ACM Comput. Surv.* **2022**. [CrossRef]
45. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
46. Naseer, M.M.; Ranasinghe, K.; Khan, S.H. Intriguing properties of vision transformers. *Adv. Neural Inf. Processing Syst.* **2021**, *34*, 23296–23308.
47. Liu, Z.; Lin, Y.; Cao, Y. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021.
48. Liu, Z.; Lin, Y.; Cao, Y. Video swin transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–23 June 2022.
49. Zheng, H.; Wang, G.; Li, X. Swin-MLP: A strawberry appearance quality identification method by Swin Transformer and multi-layer perceptron. *J. Food Meas. Charact.* **2022**, 1–12. [CrossRef]
50. Xu, X.; Feng, Z.; Cao, C. An Improved Swin Transformer-Based Model for Remote Sensing Object Detection and Instance Segmentation. *Remote Sens.* **2021**, *13*, 4779. [CrossRef]
51. Jiang, W.; Meng, X.; Xi, J. Multilevel Attention and Multiscale Feature Fusion Network for Author Classification of Chinese Ink-Wash Paintings. *Discret. Dyn. Nat. Soc.* **2022**, *2022*, 1–10. [CrossRef]
52. Qu, Z.; Cao, C.; Liu, L. A deeply supervised convolutional neural network for pavement crack detection with multiscale feature fusion. *IEEE Trans. Neural Networks Learn. Syst.* **2021**, 1–10. [CrossRef] [PubMed]
53. Gu, J.; Wang, Z.; Kuen, J. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [CrossRef]
54. Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a convolutional neural network. In Proceedings of the 2017 International conference on engineering and technology (ICET), Antalya, Turkey, 21–23 August 2017.
55. Kamilaris, A.; Prenafeta-Boldú, F.X. A review of the use of convolutional neural networks in agriculture. *J. Agric. Sci.* **2018**, *156*, 312–322. [CrossRef]
56. Zhu, L.; Li, Z.; Li, C. High performance vegetable classification from images based on alexnet deep learning model. *J. Agric. Biol. Eng.* **2018**, *11*, 217–223. [CrossRef]
57. Wang, L.; Sun, J.; Wu, X. Identification of crop diseases using improved convolutional neural networks. *IET Comput. Vis.* **2020**, *14*, 538–545. [CrossRef]
58. Lv, M.; Zhou, G.; He, M. Maize leaf disease identification based on feature enhancement and DMS-robust alexnet. *EEE Access* **2020**, *8*, 57952–57966. [CrossRef]
59. Albashish, D.; Al-Sayyed, R.; Abdullah, A. Deep CNN model based on VGG16 for breast cancer classification. In Proceedings of the2021 International Conference on Information Technology (ICIT), Amman, Jordan, 14–15 July 2021.
60. Zhu, H.; Yang, L.; Fei, J. Recognition of carrot appearance quality based on deep feature and support vector machine. *Comput. Electron. Agric.* **2021**, *186*, 106185. [CrossRef]
61. Ishengoma, F.S.; Rai, I.A.; Said, R.N. Identification of maize leaves infected by fall armyworms using UAV-based imagery and convolutional neural networks. *Comput. Electron. Agric.* **2021**, *184*, 106124. [CrossRef]
62. Mukti, I.Z.; Biswas, D. Transfer learning based plant diseases detection using ResNet50. In Proceedings of the 2019 4th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, 20–22 December 2019.
63. Gupta, K.; Rani, R.; Bahia, N.K. Plant-Seedling Classification Using Transfer Learning-Based Deep Convolutional Neural Networks. *Int. J. Agric. Environ. Inf. Syst.* **2020**, *11*, 25–40. [CrossRef]
64. Sethy, P.K.; Barpanda, N.K.; Rath, A.K. Deep feature based rice leaf disease identification using support vector machine. *Comput. Electron. Agric.* **2020**, *175*, 105527. [CrossRef]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
www.mdpi.com

Agronomy Editorial Office
E-mail: agronomy@mdpi.com
www.mdpi.com/journal/agronomy

Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-0817-5