

Deep-Learning-Based Rice Phenological Stage Recognition

Jiale Qin ^{1,2}, Tianci Hu ³, Jianghao Yuan ⁴, Qingzhi Liu ⁵, Wensheng Wang ², Jie Liu ⁶, Leifeng Guo ² and Guozhu Song ^{1,*}

¹ School of Software, Shanxi Agricultural University, Jinzhong 030801, China; z20213613@stu.edu.cn

² Agricultural Information Institute of Chinese Academy of Agricultural Sciences, Beijing 100081, China; guoleifeng@caas.cn

³ College of Computer and Information Engineering, Xinjiang Agricultural University, Urumqi 830052, China

⁴ College of Information Science and Technology, Hebei Agricultural University, Baoding 071001, China

⁵ Information Technology Group, Wageningen University & Research, 6708 PB Wageningen, The Netherlands

⁶ Institute of Artificial Intelligence, Harbin Institute of Technology, Harbin 150080, China

* Correspondence: songgz@sxau.edu.cn

Abstract: Crop phenology is an important attribute of crops, not only reflecting the growth and development of crops, but also affecting crop yield. By observing the phenological stages, agricultural production losses can be reduced and corresponding systems and plans can be formulated according to their changes, having guiding significance for agricultural production activities. Traditionally, crop phenological stages are determined mainly by manual analysis of remote sensing data collected by UAVs, which is time-consuming, labor-intensive, and may lead to data loss. To cope with this problem, this paper proposes a deep-learning-based method for rice phenological stage recognition. Firstly, we use a weather station equipped with RGB cameras to collect image data of the whole life cycle of rice and build a dataset. Secondly, we use object detection technology to clean the dataset and divide it into six subsets. Finally, we use ResNet-50 as the backbone network to extract spatial feature information from image data and achieve accurate recognition of six rice phenological stages, including seedling, tillering, booting jointing, heading flowering, grain filling, and maturity. Compared with the existing solutions, our method guarantees long-term, continuous, and accurate phenology monitoring. The experimental results show that our method can achieve an accuracy of around 87.33%, providing a new research direction for crop phenological stage recognition.

Keywords: phenology; ResNet; YOLOv5; weather stations; deep learning



Citation: Qin, J.; Hu, T.; Yuan, J.; Liu, Q.; Wang, W.; Liu, J.; Guo, L.; Song, G. Deep-Learning-Based Rice Phenological Stage Recognition. *Remote Sens.* **2023**, *15*, 2891. <https://doi.org/10.3390/rs15112891>

Academic Editors: Francesco Pirotti and Mitsunori Yoshimura

Received: 13 April 2023

Revised: 25 May 2023

Accepted: 30 May 2023

Published: 1 June 2023



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

1. Introduction

Identifying the phenological stages of crops is a critical aspect of monitoring crop conditions. This process helps field managers schedule and carry out various activities such as irrigation and fertilization at the right time. Moreover, it provides essential references for estimating the growth stage of crop yield [1]. Rice is one of the three significant staple crops globally, with Asia hosting 90% of its planting area. Nearly 60% of the Chinese population depends on rice as their primary food source. Unfortunately, accurately identifying and monitoring the phenological stages of crops remains a challenging and time-consuming task [2,3]. One of the key reasons is that rice has different characteristics and precautions at different growth stages [4]. For example, during the seeding state, rice with 3 leaves 1 heart to 4 leaves 1 heart is highly prone to slow growth and rotten seedlings if the water level and temperature are too high or too low, resulting in large areas of missing plants or fewer plants. During the tillering stage, rice enters a rapid growth period, requiring more nutrients at this time. If nutrients cannot be supplied in time during this growth stage, it will reduce or stop tillering. During the booting jointing (BJ) stage, the development speed continues to increase. The spike stage begins to differentiate, and this is the key period to determine the number of grains per spike and the consolidation of the effective number

of spikes per acre. In production, timely mid-tillage fertilization should be carried out to ensure the number of spikes in production. During the heading flowering period, when the rice ears emerge from the leaf sheaths and the white rice flowers begin to germinate, the daily temperature and various diseases should be monitored. In the monitoring process, if abnormal growth is found, artificial intervention should be carried out to ensure normal flowering pollination. During the grain filling (GF) stage, photosynthesis of rice leaves and carbon hydrate transport to grain rice occur, and filling begins. At this time, there is a close relationship between the leaf nitrogen content and photosynthesis ability. An appropriate nitrogen supply can increase leaf area photosynthesis, prevent early senescence, and improve root vitality. In production, to ensure a normal filling process, the root external topdressing method often uses supplement phosphorus potassium fertilizer, among others. In the maturity stage, rice needs to prepare for harvest sunning. The reason for this is that, after maturity, rice grains become heavier. If they cannot be laid down to be sunned in time, they fall off very easily, leading to a yield reduction [5]. Based on the examples above, it can be found that long-term phenological stage recognition can reflect environmental change patterns [6]. Based on the recognition results, farmers can formulate corresponding system plans according to their changes in morphology during crop growth, which could provide timely help and guidance for farm owners to cope with emergencies.

The current research on phenological stage identification mainly uses satellite remote sensing [7], drone remote sensing [8], and vegetation canopy images to analyze and monitor phenological stages. The methods of phenological stage identification vary according to different methods and types of data collection. Cruz-Sanabria collected data through multi-spectral instrument sensors on Sentinel-2 satellite and used the random forest technique to identify and analyze sugarcane crop phenological stages [9]; L. Chu et al. [10] analyzed the moderate resolution imaging spectroradiometer (MODIS) time series data, used the double Gaussian model and maximum curvature method to extract phenology, and proposed a mechanism for winter wheat discrimination and phenology detection in the Yellow River Delta region; Brazilian scholar Tiago Boechel recorded meteorological data and constructed a phenological stage identification model for apple trees based on the fuzzy time series prediction mechanism [11]; Boschetti et al. [12] used normalized vegetation index (NDVI) time series data provided by moderate resolution imaging spectroradiometer to estimate key phenological information of Italian rice; Chao Zhang obtained time series canopy reflectance image data of winter rapeseed by drone, fitted the time series vegetation index with three mathematical functions—symmetric Gaussian function (AGF), Fourier function, and double logistic function—extracted phenological stage information, and constructed an identification model [13]. However, there are still many problems in these methods of monitoring crop growth. In practical operation, remote sensing data have low spatial resolution, and the index extracted from images often cannot effectively reflect the real phenological information [14], such as linking soybean's phenological stage related to yield with time series data, which is difficult to extract from feature points because of a lack of obvious features [15]; artificial observation of phenology is objective and accurate, but it is costly and difficult to achieve under complex topographic conditions [16]. In order to make up for the insufficient resolution of remote sensing data and to solve the limitations of artificial field observation, using a digital camera for the near-surface remote sensing method to monitor vegetation has gradually become a new means for people to monitor vegetation phenology changes [17]. Digital camera technology has outstanding advantages, such as low-cost equipment, certain spatial scale fusion ability, automatic and continuous acquisition of image data under field conditions, and timely and accurate acquisition of vegetation community canopy status, among others, meaning it has become an effective means for monitoring vegetation community phenology. It has been applied in many ecosystems around the world. Sa I et al. [18] used the high-performance target detection model to achieve efficient recognition of sweet pepper through RGB images and infrared images. Shin Nagai installed a downward camera on a signal tower to collect leaf forest image information and extracted RGB color information from it, thus completing contin-

uous monitoring of vegetation canopy phenology [19]. Yahui Guo collected daily RGB images covering summer corn's entire growth period and determined four phenological dates: six-leaf stage, tasseling, silking, and maturity stage. They extracted indices based on Phenocam's images and completed corn's phenological extraction [20]. Teng Jiakun et al. [21] used a digital camera to capture continuous stable RGB picture data, extracted the image's RGB brightness value, accurately calculated the index, and identified key time nodes during acacia's growth defoliation process.

Given the challenges of accurately identifying mixed dense rice plants within complex field environments, it is crucial to develop a recognition method that offers higher accuracy and enables continuous observation of rice phenological periods. This study aims to address this issue by focusing on rice plants in a large field setting and proposing a deep-learning-based approach for recognizing RGB images of rice to determine its phenological stages. Specifically, this study emphasizes the detection of rice germination using deep learning techniques and incorporates rice growth cycle data collection to design an experimental study.

2. Materials and Methods

2.1. Data Acquisition

The experimental field was in Suibin County's 290 Agricultural Demonstration Zone, Hegang City, Heilongjiang Province, China (131.98326807°E, 47.59170584°N). The rice variety used was Longjing 1624 and the data were collected between 7 May and 19 September 2022. The camera used Huawei's Hysis main control chip and Sony's 8-megapixel COMS, with a maximum resolution of 3840×2160 , and has an RGB visible and RGN multispectral distortion-free lens with the same viewing angle, enabling the acquisition of high-quality and distortion-free image data from the same position and field of view. In the experiments, four RGB cameras in the rice field collected data at 8:00, 12:00, 14:00, and 16:00. The cameras were deployed at a height of 2.4 m with a 90° field of view angle. The shooting area measured 4.4 m in length and 2.5 m in width, and the primary shooting method used was vertical overhead under natural light conditions, as shown in Figure 1. The experiment used artificial visual assessment and field observation to record the rice's phenological stages, with an observation frequency of once every three days. The experimental data included the whole growth cycle of rice: seedling stage, transition stage of seedling to tillering, tillering stage, transition stage of tillering to booting jointing, booting jointing stage, transition stage of booting jointing to heading flowering, heading flowering stage, transition stage of heading flowering to grain filling, grain filling stage, transition stage of grain filling to maturity, and maturity stage, totaling 11 stages.

2.2. Dataset Construction

The original images were cropped and converted to 256 pixels by 256 pixels in image pre-processing, for a total of 25,385 images. There are 1865 images of the seedling stage, 600 images of the transition stage of seedling to tillering (S-T), 3480 images of the tillering stage, 600 images of the transition stage of tillering to booting jointing (T-B), 4800 images of the booting jointing stage, 600 images of the transition stage of booting jointing to heading flowering (B-H), 1800 images of the heading flowering stage, 600 images of the transition stage of heading flowering to grain filling (H-G), 4800 images of the grain filling stage, 720 images of the transition stage of grain filling to maturity (G-M), and 5520 images of the maturity stage. The images are named in a uniform format, with JPG image type. Some example rice phenology images are shown in Figure 2. The timing of each phenological stage is shown in Figure 3.

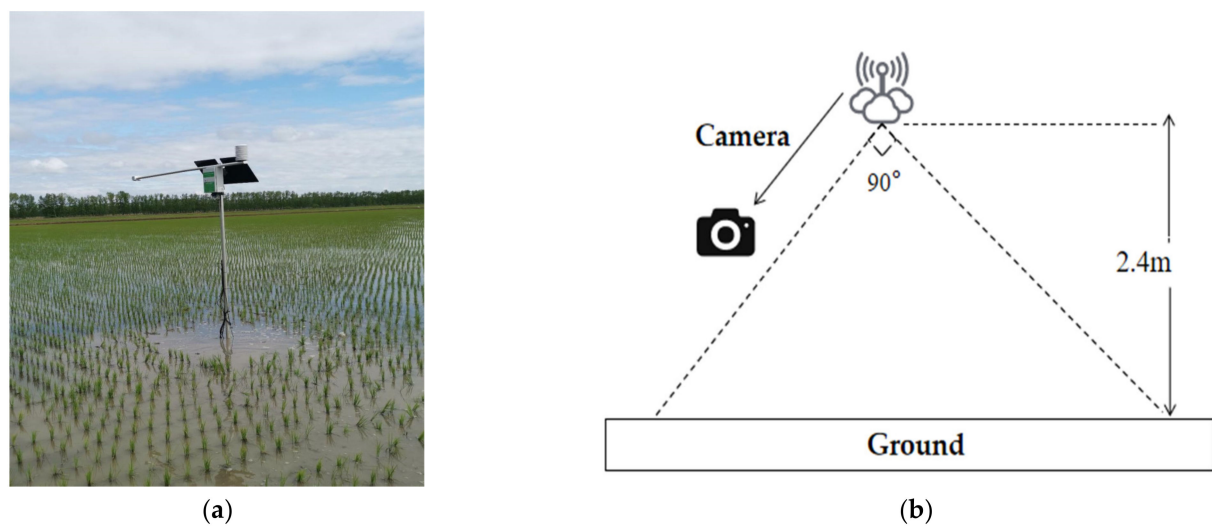


Figure 1. Schematic diagram of weather station installation. (a) Schematic diagram of the field rice weather station installation scene; (b) ground image acquisition.

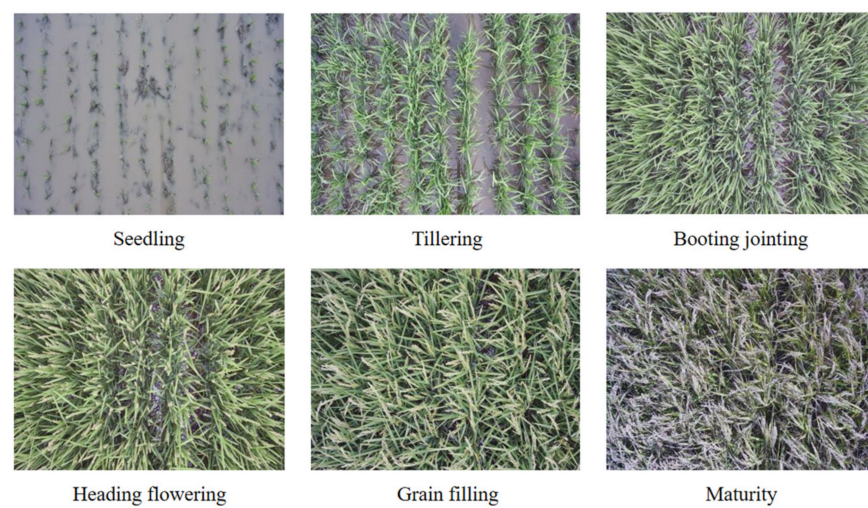


Figure 2. Rice phenology image data.

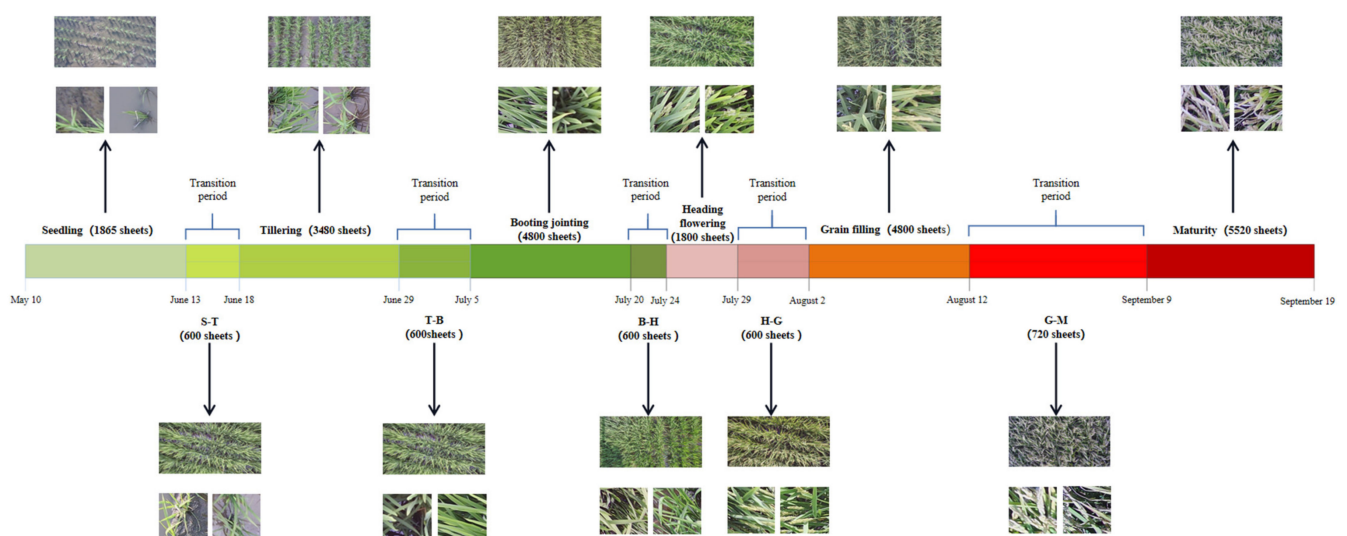


Figure 3. Timing of each phenological stage.

2.2.1. Dataset Annotation

Data annotation is a crucial process for enabling the deep learning model to learn efficiently. Normally, manual data annotation can conveniently and accurately convert the essential elements of an image into machine-readable information. Thus, before constructing a deep learning recognition model, it is necessary to annotate the data manually.

Currently, manual methods are commonly employed to select and sort data, which is time-consuming and may involve errors in data cleaning. To address this issue, a deep learning target detection method is employed in this experiment for cleaning images without rice seedlings from the original dataset, facilitating the screening of valid data during the seedling stage. Specifically, 3000 seedling stage data were cleaned using Makesense software for annotating rice seedling plants and leaf images, resulting in the elimination of invalid images, as illustrated in Figure 4. In this research, a total of 900 images were selected for dataset partitioning, including the training set, test set, and validation set, with a ratio of 8:1:1.

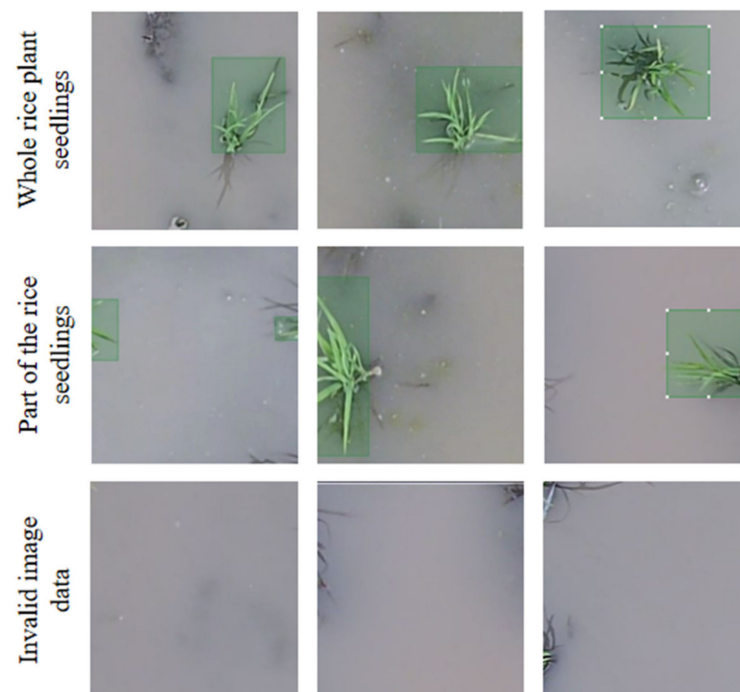


Figure 4. Images of different types of rice seedling data.

2.2.2. Data Enhancement

In the environment, it is easy to have light overexposure, fogging, and shadow occlusion. The acquired image data of rice seedling stage have a low image resolution and excessive interference factors can easily lead to low generalization ability of the model. Therefore, data enhancement methods are introduced to enrich seedling stage features and background information. The Mosaic data augmentation method [22] randomly reads four images from the training set and performs operations such as stitching, scaling, translation, rotation, and so on, and then enhances the information expression on H, S, and V color domains, as shown in Figure 5.

Mosaic image enhancement can disguisedly increase the batch size by stitching four images together, which could reduce the variance of BN when the BN operation is performed. In addition, the enhancement can improve the quality of a single image and thus the accuracy of recognition. The enhanced image contains more scenes and multi-scale information. Using the fused images to train the model could indirectly increase the sample number and speed up model convergence. Therefore, Mosaic was used to enhance the rice phenological stage seedling image dataset.

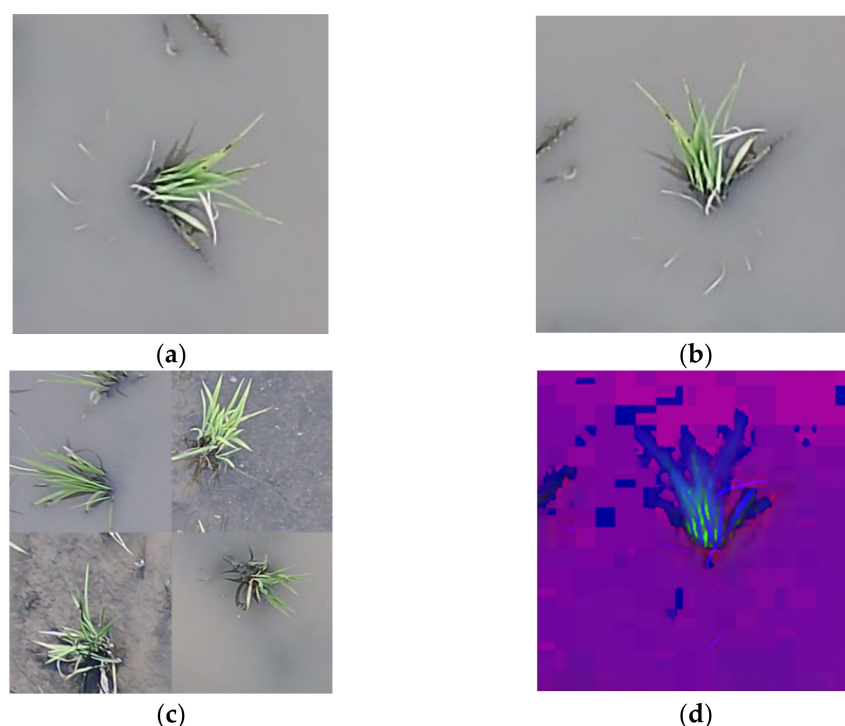


Figure 5. Original data and image enhancement effect: (a) original image data; (b) rotation; (c) splicing; and (d) color conversion.

2.3. Experimental Environment and Configuration

The experimental equipment configuration used in this research is shown in Table 1. The processor is an Intel(R) Core (TM) i9-10900 K CPU from Intel Corporation, California, USA, with a frequency of 3.70 GHz. the graphics card is an NVIDIA GeForce RTX 3090 and the RAM is 64 G. The operating system is Windows 11. The development language is python 3.9.7. The deep learning framework used is Tensorflow.

Table 1. Experimental equipment configuration.

Configuration Name	Parameters
Processor	Intel Xeon Intel(R) Core (TM) i9-10900 K CPU @ 3.70 GHz
Video cards	NVIDIA GeForce RTX 3090
Memory	64 G
Development languages	Python 3.9.7
Deep learning frameworks	Tensorflow

2.4. Model Evaluation Indicators

For classification problems, the different combinations of the true class and the predicted class of classification results can be divided into four cases, including true positive (TP), false positive (FP), true negative (TN), and false negative (FN). $TP + FP + TN + FN$ equals the total number of samples. The confusion matrix results are classified as shown in Table 2.

Table 2. Classification of confusion matrix results.

Classification Result Confusion Matrix		
Ture Situation	Predicted Label	
Ture label	Positive	Counter
Positive	TP (True positive)	FN (False negative)
Counter	FP (False positive)	TN (True negative)

TP is the true case, where the mapping indicates a positive prediction, which is the result. FP is the false positive case, where the mapping indicates a negative prediction, but the result is a positive prediction. FN is the false negative case, where the mapping indicates a positive prediction, but the result is a negative prediction. TN is the true negative case, where the mapping indicates a negative prediction, which is the result. The accuracy and completeness are defined as Equations (1)–(3), respectively.

$$\text{Accuracy rate : Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Check all rate (recall) : Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Correct Rate (Accuracy) : Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

2.5. Experimental Procedure

2.5.1. Experimental Steps

The detailed experimental process, aims, and requirements are shown in Figure 6. In this process, the first step is to set up and debug the camera to complete the data collection task. Second, the deep learning object detection algorithm is used to clean up the data and eliminate invalid data. When preprocessing the data, the original data must be enhanced. Then, deep learning image classification technology is used to input the image dataset into various network models and perform parameter tuning and optimization to build a rice phenological stage recognition model. Finally, the experimental results are compared and the recognition model with higher comprehensive performance is selected to complete the phenological stage classification task.

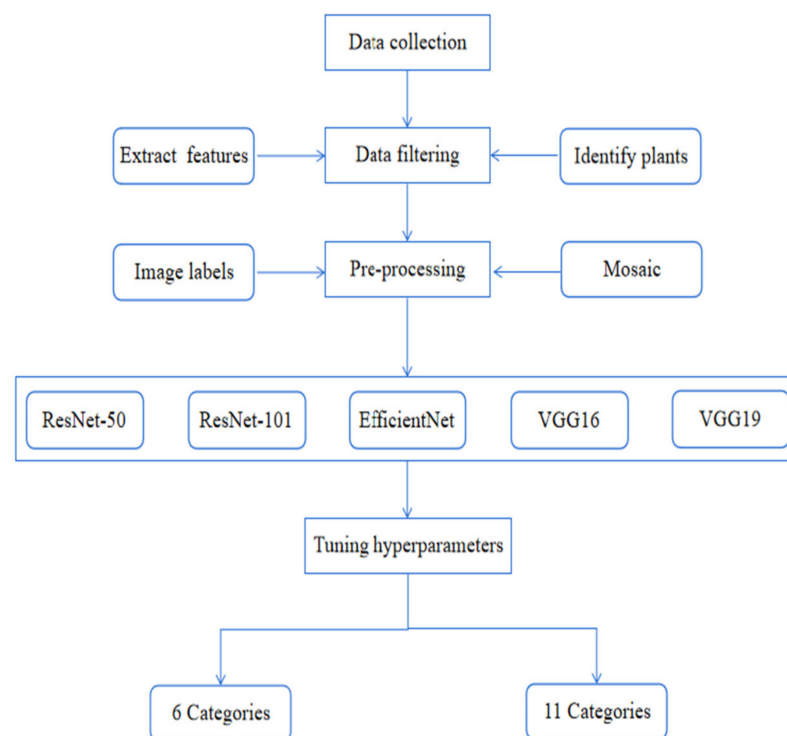


Figure 6. Experimental flow.

2.5.2. Model Training and Super-Reference Configuration

This study utilized the ability of target detection to detect multiple classes of objects in an image to identify rice seedlings at the seedling stage in a large field. Three detection

models, YOLOv5, YOLOv6, and YOLOv7, were selected before training the recognition models. During the training process, we ensured that the network structures of the three models were consistent with the training parameters. The raw data were fed into the YOLO model for filtering and adjustment. The model was trained using stochastic gradient descent with 128 batches, Adam's optimization algorithm, a learning rate (lr) of 0.01, and a "gradual decay" scheduler for the learning rate. Each time, the model reduced the learning rate by a factor of 0.5. In the training, we used 600 epochs for training, saved the model weight file every 10 epochs, and iterated 120 times per epoch. The training has a total of 72,000 iterations, with a training time of 2 h 32 min. The feature extraction process of the YOLOv5 model is shown in Figure 7.

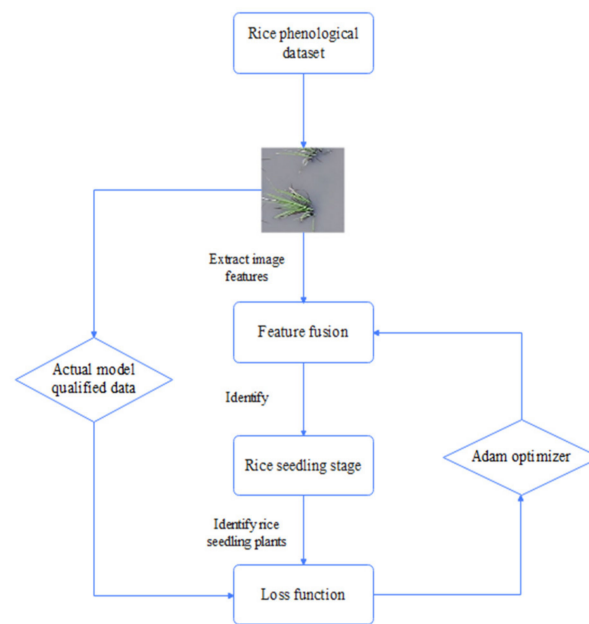


Figure 7. YOLOv5 model feature extraction.

In this study, image data of rice growth throughout the entire growth cycle were gathered. The collected data were divided into 11 stages based on the experimental design and data acquisition methods discussed in Section 2.1. To evaluate the model's performance, 2539 samples were randomly selected as the test set, maintaining a ratio of 9:1. The remaining 22,846 images were divided into training and validation sets using the same 9:1 ratio. In the experiments, several deep learning frameworks, including ResNet-50, ResNet-101, EfficientNet, VGG16, and VGG19, were used for comparative analysis. The fine-tuning process involved adjusting variables such as the scaling factor and offset of the BN layer and the weights of the classifier to achieve the highest accuracy and performance. For optimization, we employed the Adam optimization method with an initial learning rate of 0.0001. The learning rate scheduler has stepwise decay, where the learning rate is reduced by a factor of 0.97 at each step. We used softmax as the activation function in the fully connected layer. The model parameters were trained using the training set, while the validation set was used to determine the network structure as well as the model parameters. Finally, the optimal model was selected by comparing their accuracy.

3. Results

3.1. Phenological Identification Result

In this study, a model was constructed and identified for the purpose of identifying rice phenological stages. The variation in accuracy and loss rate of the network during the construction process was also demonstrated. Figure 8 illustrates that the loss function starts to converge after approximately 90 iterations, and the validation set loss function approaches that of the training set. This indicates that this model gradually stabilizes in the

training process. Then, the test set is used to verify the effect of the model. The variation in accuracy in the training process is shown in Figure 8. It can be observed that the accuracy gradually stabilizes after 90 iterations and finally converges to an accuracy of 87.33%.

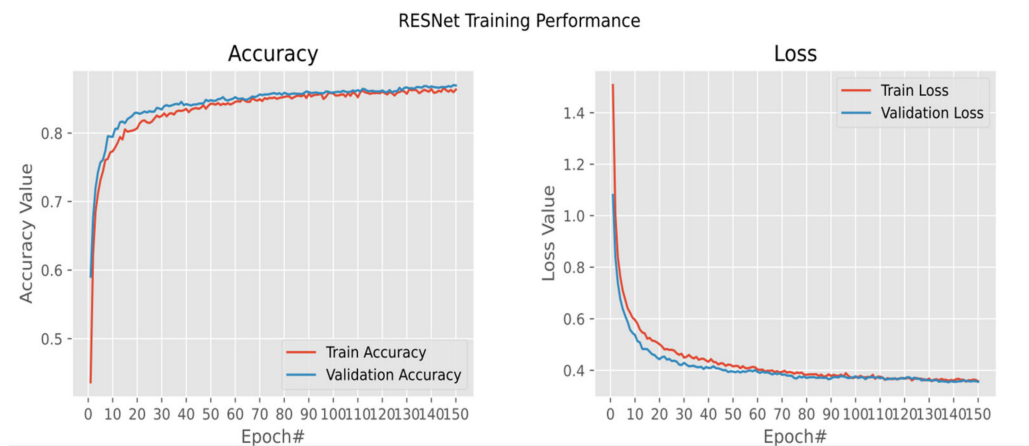


Figure 8. Images of the variation in the network’s loss function values with the number of iterations in the training and validation sets.

The training results of different deep learning models are shown in Table 3. The experimental results show that the ResNet-50 recognition model not only can achieve high overall accuracy, but also has a low loss value on the test set and the shortest single iteration time. This model balances recognition accuracy and speed, and thus has the best comprehensive performance.

Table 3. Comparison of model accuracy.

Models	Test Training Set Accuracy	Top 1	Top 5	Test Set Loss Values	Duration per Epoch
ResNet-50	87.01%	84.45%	97.06%	0.3764	286 s
ResNet-101	87.50%	84.93%	97.06%	0.3307	305 s
EfficientNet	65.77%	63.48%	70.32%	0.7231	361 s
VGG16	77.98%	75.69%	96.97%	0.6141	307 s
VGG19	76.51%	74.38%	95.63%	0.6744	302 s

3.2. Data Cleaning Results

We used the YOLO model to identify rice seedlings and constructed a complete model that accurately distinguishes water from seedlings in the rice field. The specific recognition is shown in Figure 9, where images containing rice seedlings were selected to compile a new dataset of seedling stage images. This dataset can be re-entered into the ResNet50 network model for further experiments.

To verify the effectiveness of cleaning the seedling image data, we conducted experiments using ResNet-50 on the cleaned dataset and the original dataset. The experimental results are shown in Table 4.

After running the model on the rice seedling images, there were a total of 3000 original images, that is, 1865 output images by the model and 1135 deleted images. The images of successful identification outputs were reorganized and summarized by the model to build a high-quality dataset of rice seedling images.

After comparing the trained models to different datasets, the dataset with the highest accuracy was selected as the input dataset. The dataset comparison was conducted from multiple perspectives. The model training results are shown in Table 4. Comparing different models to evaluation indicators shows that Yolov5 has better comprehensive performance. It cleaned data with 0.32% higher accuracy than the original data and 1.15% higher in Top

1. The single iteration time is shortened by 6 s. This experiment proves that the method used in this study can improve its training accuracy while shortening the model training time and build a better recognition model.

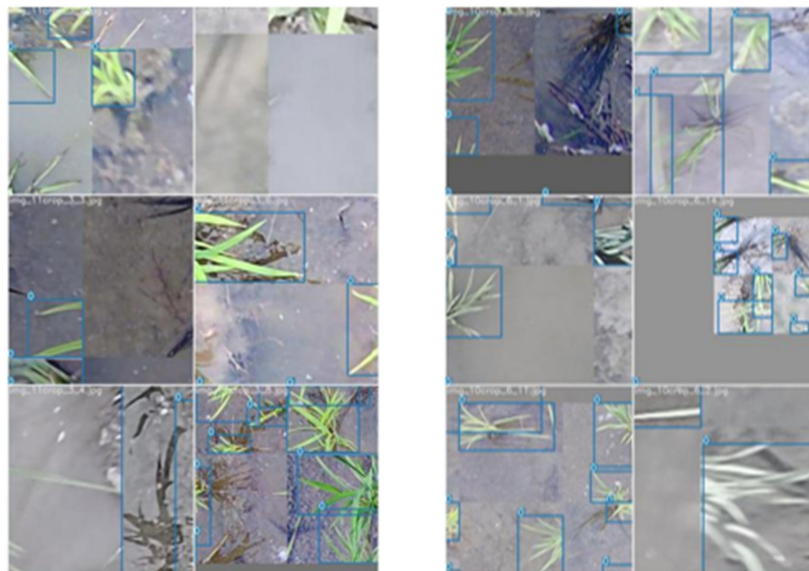


Figure 9. Graph of training results.

Table 4. Comparison of accuracy for different datasets.

Datasets	Model Testing Training Set Accuracy	Top 1	Top 5	Test Set Loss Values	Duration per Epoch
Yolov5 data cleaning	87.33%	85.60%	97.06%	0.3866	280 s
Yolov6 data cleaning	86.51%	83.22%	95.31%	0.3648	282 s
Yolov7 data cleaning	86.86%	83.94%	96.78%	0.3521	279 s
Raw data	87.01%	84.45%	97.06%	0.3823	286 s

3.3. Results of the Different Taxonomic Classifications for the Identification of Phenology

Based on the ResNet-50 convolutional neural network model, we analyzed two types of models that only contain phenological stages (6 classes) and contain transition stages (11 classes). Figure 10 shows the normalized confusion matrix of the two models on the test set. Each column in the figure represents the number of true classes of the model. Each row represents the number of predicted classes of the model. The number of rows indicates the number of corresponding model classes.

The accuracy comparison results of each phenological stage are shown in Figure 11. By comparing the accuracy of the model composed of transition period image data to that of the model without a transition period, it can be observed that rice in the 11-category model has low classification accuracy in four phenological stages: seedling stage, transition period of T–B, heading flowering stage, and transition period of H–G. The main reason is that the time duration of the adjacent two growth stages are short and, therefore, the image similarity is relatively high. Meanwhile, it can be observed that the accuracy of the other categories is high. Further, it can be found that the volatility is large and the difference in accuracy between each category of the model is large. In the six-class recognition model, the accuracy of the grain filling stage and maturity stage reached 93.03% and 96.07%, respectively. The accuracy of the other four classes under the same model meets the crop

recognition accuracy requirements in field production, which is more suitable for rice phenological stage recognition.

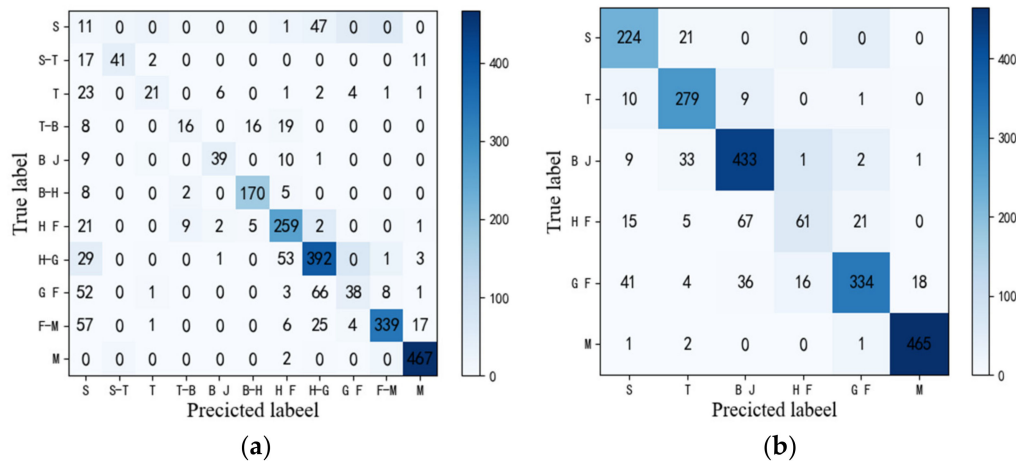


Figure 10. Confusion matrix for the two models: (a) contains transitional periods (11 categories); (b) contains only single phenological periods (6 categories).

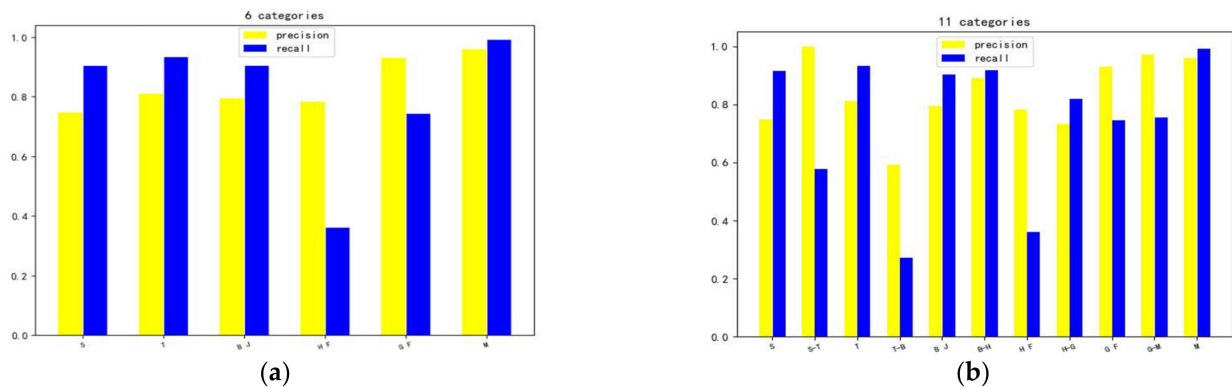


Figure 11. Comparison of accuracy by phenological stage: (a) 6 categories and (b) 11 categories.

4. Discussion

The characteristics of rice vary during different growth and development periods, and the corresponding fertilization, chemical application, and irrigation plans can be adjusted according to the different periods. The key to the effectiveness of the rice phenological stage recognition is the accuracy of the recognition. With the widespread use of image acquisition equipment, a large amount of data are available for analysis and experimentation [23]. We conducted experiments on data collected from weather stations equipped with RGB cameras to demonstrate that the experimental research method can accurately determine the entire growth and development period of rice for practical production needs.

This study opted to construct a rice recognition model using the ResNet network, known for its capability to train neural networks with an increased number of layers while avoiding issues such as gradient disappearance or gradient explosion. For image classification using ResNet, which mainly involves efficiency as well as accuracy issues, Zhong et al. [24] divides and processes images with different complexity and selects the network structure according to the average complexity of the training images. Through experiments, they proved that this method can reduce the labor cost and time cost of the algorithm under the condition that the deep learning algorithm is guaranteed. Lin et al. [25] train the model before the acquired hyperspectral images (HSIs) are labeled and introduce the active learning process to initialize significant samples on the HSI data. They propose to construct and connect higher-level features for source and target HSI data

to further overcome cross-domain differences. Through experiments, they verified that this method can not only reduce redundant data, but also improve its model recognition accuracy. Sarwinda et al. [26] takes the data images and inputs them into different network models. Through comparison, we found that ResNet-18 has better performance and can reach more than 87% classification accuracy and more than 83% sensitivity. This confirms that the model has strong generalization ability and has better applicability in solving this kind of image classification problem. As can be seen in Table 3, we verified that the use of the ResNet-50 network structure can achieve better recognition of the object period by comparing different network models. Not only can it display high accuracy, but also its inference speed can be up to 81 s faster than other network models.

In order to make the model more comprehensive, we conducted exploratory experiments, mainly using the target detection model to filter the seedling stage data. For seedling images, we compared no screening with screening using the YOLO model. In some current studies, using YOLO for single target detection is very effective. Zhang et al. [27] designed a lightweight Yolov4 apple detection model that incorporates an attention mechanism module to remove non-target rows of apples in dwarf rootstock standardized densely planted apple orchards with a final mAP improvement of 3.45% (to 95.72%). Zhao et al. [28] proposed an improved model of Yolov5s for disease detection of crops, which was compared with the Yolov3 and Yolov4 models, and the mAP and recall of the improved model were 95.92% and 87.89%, respectively. In this paper, several versions of YOLO were used to detect the seedling stage data, and better results were obtained. From Tables 3 and 4, it can be concluded that the overall accuracy of classification of the filtered images is higher than that of the unscreened images. The accuracy of the test training set is improved from 87.01% to 87.33%, and the accuracy of Top 1 is improved from 84.45% to 85.60%. The training time of the model was shortened from 286 s to 280 s per iteration with the same learning rate, batch size, and number of fully connected layers. Accuracy rates of 74.67%, 81.10%, 79.45%, 78.21%, 93.04%, and 96.07% for the seedling, tillering, booting jointing, heading flowering, grain filling, and maturity stages, respectively, were achieved and the model was able to identify the crop phenological periods with high accuracy.

In the field of computer vision, data are one of the key factors for training models. Different datasets have different effects on the performance of the model. Han et al. [29] recorded and photographed the rice phenology continuously and at different angles using handheld cameras. This method can accomplish the detection of the phenology in a small area. Sheng et al. [30] investigated the data of various features in rice weather stage recognition from high-definition cameras and sensors on weather stations in rice fields and verified the accuracy of the image data in judging the weather stage by inputting the image data and sensor data into the model. Cheng et al. [31] pointed out that, when training the network model for recognition, the convergence of the model will not be able to continue to improve after a certain number of iterations. In the process of recognizing rice in a large field, multiple cameras are set up to change the scene to ensure the generalization ability and robustness of the model. We explored different data and recognition accuracy. We first classified the six basic periods on the phenological period and then added five transition periods for classification, and the results are shown in Figures 10 and 11. We found that adding transitional period data would lead to large fluctuations in the accuracy of identifying each period, with 96.07% accuracy in the maturity period, but less than 30% accuracy in the “H–F” period. There are three main reasons for this problem: first, images of transition periods are similar to images of other two periods, and changes in characteristics are not significant enough from the perspective of rice growth phenotype; second, the sample size of data is too small and the transition period is short; third, the resolution of images is not high enough, so it is necessary to improve the camera when detecting crops in field. Therefore, we should improve acquisition frequency and image quality during data collection when analyzing crop phenology to achieve the observation of the whole phase of crop phenology and provide high-quality data for subsequent research.

5. Conclusions

By focusing on the identification of mixed dense rice plants in complex field environments, comprehensive field data were collected and processed throughout the growth and development stages of rice. This study introduced an enhancement to the traditional phenological method by integrating the ResNet-50 algorithm with the Yolov5 algorithm, resulting in a model identification accuracy of 87.33% and a 1.15% improvement in Top 1 accuracy. The experiments convincingly demonstrate the effectiveness of this method in accurately identifying crop phenology, highlighting the potential for the developed rice phenology identification model to be sustainable and practical for long-term crop monitoring.

However, there are some limitations of this study. First, the number of cameras was not large enough and the variety and angle of the cameras were too few. It is recommended to add more cameras and rice varieties and to change the camera angles. If the monitored area is too large, it is recommended to install a weather station at the center of every four intersecting rice fields and add cameras on it to monitor the rice fields in a large scale and at multiple angles, so that more accurate judgments can be made when identifying rice phenological periods in different fields. Second, the authors plan to add more extreme weather data, such as strong midday exposure, weak exposure, cloudy rain, and other severe weather as model input data to build the rice phenological stage recognition model for the entire rice dataset collection process. These improvements will enhance the practicality of the results of their study and lay a solid foundation for the next part to be deployed in the system platform, which in turn will facilitate remote observation and diagnosis in the context of precision agriculture.

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

Funding: This research was supported by the National Key R&D Program of China (2021ZD0110901), the Basic Research Program of Shanxi Province (202103021224173) and the Science and Technology Innovation Program of AII-CAAS (CAAS-ASTIP-2023-AII).

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

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

References

1. Feng, H.; Li, Z.; He, P.; Jin, X.; Yang, G.; Yu, H.; Yang, F. Simulation of Winter Wheat Phenology in Beijing Area with DSSAT-CERES Model. In *Computer and Computing Technologies in Agriculture IX*; Springer International Publishing: Cham, Switzerland, 2016.
2. Gao, F.; Zhang, X. Mapping Crop Phenology in Near Real-Time Using Satellite Remote Sensing: Challenges and Opportunities. *J. Remote Sens.* **2021**, *2021*, 8379391. [[CrossRef](#)]
3. Zhong, L.; Hu, L.; Yu, L.; Gong, P.; Biging, G.S. Automated mapping of soybean and corn using phenology. *Isprs J. Photogramm. Remote Sens.* **2016**, *119*, 151–164. [[CrossRef](#)]
4. da Silva, J.T.; Paniz, F.P.; Sanchez, F.E.S.; Pedron, T.; Torres, D.P.; da Rocha Concenço, F.I.G.; Parfitt, J.M.B.; Batista, B.L. Selected soil water tensions at phenological phases and mineral content of trace elements in rice grains—mitigating arsenic by water management. *Agric. Water Manag.* **2020**, *228*, 105884. [[CrossRef](#)]
5. Bueno, C.S.; Lafarge, T. Higher crop performance of rice hybrids than of elite inbreds in the tropics: 1. Hybrids accumulate more biomass during each phenological phase. *Field Crop. Res.* **2009**, *112*, 229–237. [[CrossRef](#)]
6. He, L.; Asseng, S.; Zhao, G.; Wu, D.; Yang, X.; Zhuang, W.; Jin, N.; Yu, Q. Impacts of recent climate warming, cultivar changes, and crop management on winter wheat phenology across the Loess Plateau of China. *Agric. For. Meteorol.* **2015**, *200*, 135–143. [[CrossRef](#)]
7. Valipour, M.; Dietrich, J. Developing ensemble mean models of satellite remote sensing, climate reanalysis, and land surface models. *Theor. Appl. Climatol.* **2022**, *150*, 909–926. [[CrossRef](#)]
8. Ganguly, S.; Friedl, M.A.; Tan, B.; Zhang, X.; Verma, M. Land surface phenology from MODIS: Characterization of the Collection 5 global land cover dynamics product. *Remote Sens. Environ.* **2010**, *114*, 1805–1816. [[CrossRef](#)]

9. Cruz-Sanabria, H.; Sanches, M.G.; Caicedo, J.P.R.; Avila-George, H. Identification of phenological stages of sugarcane cultivation using Sentinel-2 images. In Proceedings of the 2020 9th International Conference On Software Process Improvement (CIMPS), Sinaloa, Mexico, 21–23 October 2020.
10. Chu, L.; Liu, Q.-S.; Huang, C.; Liu, G.-H. Monitoring of winter wheat distribution and phenological phases based on MODIS time-series: A case study in the Yellow River Delta, China. *Agric. Sci. China* **2016**, *15*, 2403–2416. [\[CrossRef\]](#)
11. Boechel, T.; Policarpo, L.M.; de Oliveiar Ramos, G.; da Rosa Righi, R. Fuzzy time series for predicting phenological stages of apple trees. In Proceedings of the 36th Annual ACM Symposium on Applied Computing, New York, NY, USA, 22–26 April 2021; pp. 934–941.
12. Boschetti, M.; Stroppiana, D.; Brivio, P.A.; Bocchi, S. Multi-year monitoring of rice crop phenology through time series analysis of MODIS images. *Int. J. Remote Sens.* **2009**, *30*, 4643–4662. [\[CrossRef\]](#)
13. Zhang, C.; Xie, Z.; Shang, J.; Liu, J.; Dong, T.; Tang, M.; Feng, S.; Cai, H. Detecting winter canola (*Brassica napus*) phenological stages using an improved shape-model method based on time-series UAV spectral data. *Crop J.* **2022**, *10*, 1353–1362. [\[CrossRef\]](#)
14. Pan, Y.; Li, L.; Zhang, J.; Liang, S.; Zhu, X.; Sulla-Menashe, D. Winter wheat area estimation from MODIS-EVI time series data using the Crop Proportion Phenology Index. *Remote Sens. Environ.* **2012**, *119*, 232–242. [\[CrossRef\]](#)
15. Zeng, L.; Wardlow, B.D.; Wang, R.; Shan, J.; Tadesse, T.; Hayes, M.J.; Li, D. A hybrid approach for detecting corn and soybean phenology with time-series MODIS data. *Remote Sens. Environ.* **2016**, *181*, 237–250. [\[CrossRef\]](#)
16. Richardson, A.D.; Hollinger, D.Y.; Dail, D.B.; Lee, J.T.; Munger, J.W.; O’keefe, J. Influence of spring phenology on seasonal and annual carbon balance in two contrasting New England forests. *Tree Physiol.* **2009**, *29*, 321–331. [\[CrossRef\]](#)
17. Adamsen, F.J.; Pinter, P.J., Jr.; Barnes, E.M.; LaMorte, R.L.; Wall, G.W.; Leavitt, S.W.; Kimball, B.A. Measuring Wheat Senescence with a Digital Camera. *Crop Sci.* **1999**, *39*, 719–724. [\[CrossRef\]](#)
18. Sa, I.; Ge, Z.; Dayoub, F.; Upcroft, B.; Perez, T.; McCool, C. DeepFruits: A Fruit Detection System Using Deep Neural Networks. *Sensors* **2016**, *16*, 1222. [\[CrossRef\]](#)
19. Nagai, S.; Saitoh, T.M.; Noh, N.J.; Yoon, T.K.; Kobayashi, H.; Suzuki, R.; Nasahara, K.N.; Son, Y.; Muraoka, H. Utility of information in photographs taken upwards from the floor of closed-canopy deciduous broadleaved and closed-canopy evergreen coniferous forests for continuous observation of canopy phenology. *Ecol. Inform.* **2013**, *18*, 10–19. [\[CrossRef\]](#)
20. Guo, Y.; Chen, S.; Wang, H.; de Beurs, K. Comparison of Multi-Methods for Identifying Maize Phenology Using PhenoCams. *Remote Sens.* **2022**, *14*, 244. [\[CrossRef\]](#)
21. Jia-Kun, T.; Yu, L.; Ming-Tao, D. Study on the applicable indices for monitoring seasonal changes of *Acacia sylvestris* based on RGB images. *Remote Sens. Technol. Appl.* **2018**, *33*, 476–485.
22. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
23. Sun, Y.; Wang, H.Q.; Xia, Z.Y.; Ma, J.H.; Lv, M.Z. Tobacco-disease Image Recognition via Multiple-Attention Classification Network. In Proceedings of the 4th International Conference on Data Mining, Communications and Information Technology (DMCIT 2020), Shaanxi, China, 21–24 May 2020.
24. Zhong, S.; Jia, C.; Chen, K.; Dai, P. A novel steganalysis method with deep learning for different texture complexity images. *Multimed. Tools Appl.* **2019**, *78*, 8017–8039. [\[CrossRef\]](#)
25. Lin, J.; Zhao, L.; Li, S.; Ward, R.; Wang, Z.J. Active-Learning-Incorporated Deep Transfer Learning for Hyperspectral Image Classification. *JSTARS* **2018**, *11*, 4048–4062. [\[CrossRef\]](#)
26. Sarwinda, D.; Paradisa, R.H.; Bustamam, A.; Anggia, P. Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer. *Procedia Comput. Sci.* **2021**, *179*, 423–431. [\[CrossRef\]](#)
27. Zhang, C.; Kang, F.; Wang, Y. An Improved Apple Object Detection Method Based on Lightweight YOLOv4 in Complex Backgrounds. *Remote Sens.* **2022**, *14*, 4150. [\[CrossRef\]](#)
28. Yun, Z.; Yang, Y.; Xu, X.; Sun, C. Precision detection of crop diseases based on improved YOLOv5 model. *Front. Plant Sci.* **2023**, *13*, 1066835.
29. Han, J.; Shi, L.; Yang, Q.; Huang, K.; Zha, Y.; Yu, J. Real-time detection of rice phenology through convolutional neural network using handheld camera images. *Precis. Agric.* **2020**, *22*, 154–178. [\[CrossRef\]](#)
30. Sheng, R.T.; Huang, Y.-H.; Chan, P.-C.; Bhat, S.A.; Wu, Y.-C.; Huang, N.-F. Rice Growth Stage Classification via RF-Based Machine Learning and Image Processing. *Agriculture* **2022**, *12*, 2137. [\[CrossRef\]](#)
31. Cheng, M.; Yuan, H.; Wang, Q.; Cai, Z.; Liu, Y.; Zhang, Y. Application of deep learning in sheep behaviors recognition and influence analysis of training data characteristics on the recognition effect. *Comput. Electron. Agric.* **2022**, *198*, 107010. [\[CrossRef\]](#)

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