

Article

Automatic Counting and Location Labeling of Rice Seedlings from Unmanned Aerial Vehicle Images

Jui-Feng Yeh *, Kuei-Mei Lin, Li-Ching Yuan and Jenq-Muh Hsu

Department of Computer Science and Information Engineering, National Chiayi University,
Chia-Yi City 60004, Taiwan; s1120311@mail.ncyu.edu.tw (K.-M.L.); s1110326@mail.ncyu.edu.tw (L.-C.Y.);
hsujm@mail.ncyu.edu.tw (J.-M.H.)

* Correspondence: ralph@mail.ncyu.edu.tw

Abstract: Traditional counting of rice seedlings in agriculture is often labor-intensive, time-consuming, and prone to errors. Therefore, agricultural automation has gradually become a prominent solution. In this paper, UAV detection, combining deep learning with unmanned aerial vehicle (UAV) sensors, contributes to precision agriculture. We propose a YOLOv4-based approach for the counting and location marking of rice seedlings from unmanned aerial vehicle (UAV) images. The detection of tiny objects is a crucial and challenging task in agricultural imagery. Therefore, we make modifications to the data augmentation and activation functions in the neural elements of the deep learning model to meet the requirements of rice seedling detection and counting. In the preprocessing stage, we segment the UAV images into different sizes for training. Mish activation is employed to enhance the accuracy of the YOLO one-stage detector. We utilize the dataset provided in the AIdea 2021 competition to evaluate the system, achieving an F1-score of 0.91. These results indicate the superiority of the proposed method over the baseline system. Furthermore, the outcomes affirm the potential for precise detection of rice seedlings in precision agriculture.

Keywords: object detection; YOLOv4; smart agriculture; unmanned aerial vehicle images; precision agriculture



Citation: Yeh, J.-F.; Lin, K.-M.; Yuan, L.-C.; Hsu, J.-M. Automatic Counting and Location Labeling of Rice Seedlings from Unmanned Aerial Vehicle Images. *Electronics* **2024**, *13*, 273. <https://doi.org/10.3390/electronics13020273>

Academic Editors: Charles Tijus, Kuei-Shu Hsu, Mahmut Reyhanoglu, Teen-Hang Meen, Po-Lei Lee and Chun-Yen Chang

Received: 6 November 2023

Revised: 30 December 2023

Accepted: 2 January 2024

Published: 8 January 2024



Copyright: © 2024 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

The deployment of artificial intelligence technologies in agricultural practices, known as precision agriculture, has become a significant topic in recent years. Compared to that of industrial areas, the complex natural environments of agriculture make it more difficult to adopt automatic technologies, especially in instance segmentation models for detecting and counting. Traditional rice cultivation encounters numerous challenges. First, with the escalating issue of rural depopulation, the adoption of automated agricultural activities to enhance efficiency has become an important concern. Second, due to the impact of extreme climates, agriculture often encounters water shortage crises. In response to United Nations Sustainable Development Goal (SDG) 6.4 “substantially increase water-use efficiency across all sectors to address water scarcity”, precision agriculture, which enables the control of irrigation water quantity based on the number of plants, has emerged as a potential savior. Finally, the manual disaster assessment of farmland is time-consuming and should be assisted by AI to improve efficiency. At the same time, novel image sensor technologies by Unmanned Aerial Vehicles (UAV) are also widely used in many fields. Unmanned Aerial Vehicles (UAVs) can capture images over large areas and multiple perspectives, enabling computer vision to be used for the detection of desired objects. Therefore, the application of UAV-based detection has become a focal point in precision agriculture research.

In recent studies, precision agriculture typically employed machine learning and computer vision for target detection and achieved commendable results. Jiang et al. [1] used the model based on DenseNet to introduce depth-wise separable convolutions to improve parameter utilization and training speed. By leveraging channel attention mechanisms to

enhance effective features, the accuracy of rice pest and disease recognition was improved by 13.8%. Ammar et al. [2] compared several convolutional neural networks, such as Faster R-CNN, EfficientDet, etc. They integrated geotagged metadata, photogrammetric concepts, and distance calibration to identify palm trees' geographical locations. This enables the automated recognition, counting, and geospatial positioning of palm trees and other vegetation.

There is much research applying machine learning to Unmanned Aerial Vehicles. However, targets in UAV images are significantly smaller compared to typical targets, leading to frequent errors during detection. Therefore, enhancing the performance of deep learning models in the detection of tiny targets is a crucial challenge. Luo et al. [3] integrated the network based on YOLOv3, collaborating with the K-means++ algorithm and soft non-maximum suppression (Soft-NMS) algorithm. They improved the performance of detecting tiny targets in UAV images. Wang et al. [4] used a Kalman filter for tracking the seedling path to support YOLOv3 and improved the performance on detecting tiny targets such as corn seedlings. Luo et al. [5] adopted the Asymmetric ResNet (ASResNet) module, Asymmetric Enhanced Feature Extraction (AEFE) module, and Asymmetric Res2Net (ASRes2Net) module to replace the backbone of YOLOv5. They employed Group Spatial Pyramid Pooling (GSPP) instead of Spatial Pyramid Pooling (SPP) and incorporated the Efficient Channel Attention (IECA) module to enhance the accuracy of the UAV image target classification task. Yang et al. [6] created a lightweight model by using the GhostNet module to replace the relevant convolution in the YOLOv5 and utilized Efficient Intersection over Union (EIoU) to accelerate convergence to improve regression performance. The model achieved the requirement to reduce parameters while simultaneously improving performance.

This study proposes a UAV image-based rice seedling labeling method for precision agriculture using YOLOv4 [7]. The approach integrates data augmentation, a target detection model based on YOLOv4, and improves activation functions. Simultaneously, we introduce a formula for assessing the accuracy of predicting the plant center point. Our approach achieves automatic counting and position labeling functionalities essential for precision agriculture. This method optimizes automatic irrigation by detecting the coordinates of rice plants in UAV images. Additionally, the plant count serves as a reference for yield assessment, facilitating UAV image-based precision agriculture applications. The dataset from the 2021 AIda "Crop Location Auto-Labeling Competition" is utilized as the experimental dataset for our system. UAV images are split into different sizes during training. Then, we use the YOLOv4 one-stage detector to identify, count, and label the exact coordinates of the rice plants in the UAV images. Finally, for the configuration of YOLO, we use the Mish activation function to increase the generalization ability and accuracy of the model. Then, we prove the feasibility and effectiveness of our method through experiments and participate in the AIda 2021 competition. To sum up, our contributions are:

1. Improving the accuracy of tiny object detection under the lack of data.
2. Instead of ReLU, the Mish activation function is used in YOLOv4 and achieves better performance.
3. Proposing a criterion for determining the hit of detecting of the rice seedling.
4. Proposing the application of YOLOv4 for smart agriculture.

The remainder of this paper is organized as follows: Section 2 contains the work on object detection in UAV images and discusses methods for improving the YOLO series. We explain the rationale behind choosing a YOLO-based method for our task, introduce the structure of the YOLOv4 one-stage detector, and describe a modification of the activation function in Section 3. In Section 4, we analyze the effects of the Mish activation function and data augmentation for image segmentation at different sizes on the experimental results. The conclusion is presented in Section 5.

2. Related Works

Precision agriculture has emerged as common practice in agricultural automation. Leveraging information technologies such as unmanned aerial vehicles (UAVs) and sensors aims to enhance production efficiency and resource utilization by optimizing crop requirements. Jiang et al. [1] proposed a rice disease identification method based on an improved DenseNet network. The approach integrated the channel attention mechanism squeeze-and-excitation, depth-wise separable convolutions, and the AdaBound algorithm to accelerate training and enhance accuracy. Ammar et al. [2] proposed a convolutional neural networks-based framework for the automated counting and geolocation of palm trees from aerial images. Liu et al. [8] developed a variable-rate spraying system. The system employed a CNN-based model to classify weeds and strawberries, thereby reducing pesticide losses. Li et al. [9] proposed a system for predicting the nutritional status of rice using UAV imagery. The system analyzes multispectral images, including normalized difference vegetation index (NDVI), normalized difference red edge (NDRE), and plant nitrogen content to predict the nutritional status of rice. Oliveira et al. [10] employed machine learning methods such as Extremely Randomized Trees and XGBoost to predict the yield of corn.

To enable the management of irrigation resources and yield assessment in precision agriculture applications, the integration of UAVs with deep learning has emerged as a trend in agricultural research. Hu et al. [11] proposed a UAV tiny object detection method based on fully convolutional one-stage object detection (FCOS). They introduced the global context module combined with feature pyramid networks (FPN) in ResNet50 as the backbone to strengthen feature representation. Complete intersection over union (CIoU) Loss is used and an adaptive feature balancing subnetwork was designed to enhance the detection performance of tiny objects. Jawaharlalnehru et al. [12] also adopted a YOLOv5-based method to detect UAV images. They increased three 3×3 dimensions to replace the final convolutional layer in the Darknet 19 backbone, aiming to preserve spatial information. To enhance the robustness of the model, they utilized their homemade image classification dataset with varying resolutions for pre-training. During the training phase, the model's input size was dynamically adjusted. Tseng et al. [13] achieved precision agriculture through the detection of rice seedlings in UAV images. They employed transfer learning with two machine learning models, EfficientDet-D0 and Faster R-CNN to implement a Support Vector Machine (SVM) classification approach based on a Histogram of Oriented Gradients (HOG). This approach reduced computational time while achieving outstanding performance. Liu et al. [14] optimized the residual block on YOLOv3 by concatenating two ResNet units. Subsequently, convolutional operations were added in the early layers of the darknet to enhance spatial information, significantly improving the performance of detecting tiny objects in UAV images. Shen et al. [15] employed YOLOv5-based approach and estimated the object scale using an inertial measurement unit (IMU) to improve the performance of the detector. Their proposed method can detect tiny objects in different scales of UAV images. Amarasingam et al. [16] proposed a system for the automated inspection of sugarcane white leaf disease using UAV images. The system integrated UAVs with the YOLOv5 model, achieving promising results in the detection of sugarcane white leaf disease. Tatini et al. [17] proposed YOLOv4-SUFF for detecting rice fields in UAV images. YOLOv4-SUFF introduced an additional layer to YOLOv4, enabling the detector to extract specific feature maps for obtaining more detailed information.

Combining the observations from the aforementioned related works, it can be noted that in the field of object detection, most tasks opt for utilizing the YOLO series methods and have achieved commendable results. Nevertheless, in addressing the challenge of tiny object detection, there is still a need to develop improved methods to enhance the performance of YOLO in detecting tiny objects. Therefore, the improvement method of YOLO has been a trend in recent years. Sun et al. [18] employed MobileNetv2 as the backbone of their YOLOv4-based lightweight model. By incorporating focal loss and a convolutional block attention module, they improved the performance of detecting

various varieties of rice panicles in UAV images. Zhang et al. [19] proposed a lightweight GhostNet-YOLOv4 to identify objects. They embedded the full convolutional attention module (C-SE) model into the backbone of YOLOv4 and replaced ordinary convolution by depth-separable convolution. In order to optimize their lightweight model, they applied an adaptive cosine annealing learning rate and a focal loss function, resulting in improved accuracy. Du et al. [20] applied secondary transfer learning to solve the limitation of insufficient datasets. They also incorporated a hard negative mining block into the YOLOv4 model to solve the problem of detecting a tiny object on a complex background. Li et al. [21] proposed a YOLOv4-based tiny objects detection improvement method PF_YOLOv4. PF_YOLOv4 added a soft thresholding module to the residual structure of the backbone network which can enhance the robustness of the model. It also used depthwise separable convolution to reduce parameters. Finally, they added the Convolutional Block Attention Module (CBAM) to strengthen the feature representation of the network. Zhang et al. [22] proposed a YOLOv5-pole UAV image detection method. They utilized the mixup data augmentation technique to improve the model's generalization ability and robustness, introducing GhostBottleneck to lighten the model and combine it with the Shuffle Attention (SA) module to focus more on tiny objects. Ultimately, the goal was to improve the performance of YOLOv5 while reducing the number of parameters. Their experiment was successfully mounted on agricultural UAVs' onboard equipment and applied to smart agriculture. Wang et al. [23] proposed a YOLOv5-based lightweight object detection model MFP-YOLO. To address the challenges of significant scale variations and complex background interference in UAV images, MFP-YOLO adopted an attention mechanism in the multi-path inverted residual module. Additionally, it utilized parallelized reverse convolutional spatial pyramid pooling to enhance multi-scale target detection. The Focal-EIoU loss function was employed to enhance training stability and detection accuracy. By utilizing lightweight decoupled heads, they also reduced the parameters in YOLOv5, thereby improving overall performance.

3. Materials and Methods

3.1. System Framework

The framework of the proposed seedling labeling system is shown in Figure 1. The framework integrates three methods: data augmentation, a YOLOv4-based object detection model, and an activation function modification. During the training phase, UAV images are fed into the model. To address the issue of limited data, we employ a data augmentation method involving image cropping. Additionally, for enhanced model generalization ability and training speed, we replaced activation functions in the YOLOv4 model with the Mish activation function.

During the detection phase, users upload the UAV images they want to detect through the interface. These UAV images are processed by the YOLOv4 one-stage detector, utilizing trained weights for object labeling. In the YOLOv4 one-stage detector, UAV images are initially fed into CSPDarknet53 [7]. Subsequently, the Neck component integrates features using SPPNet and PANet. Finally, the Head component performs Dense Prediction to predict label results. The rice seedlings' labels and total counts are displayed in the user interface, and the positions of the labeled seedlings' centroids are saved in a text file.

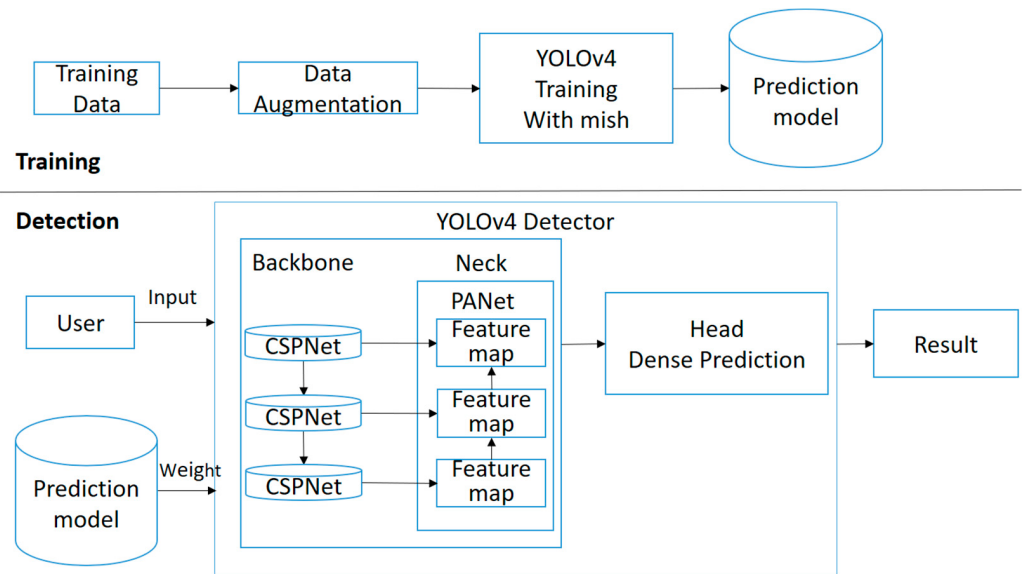


Figure 1. System framework of the proposed seedling labeling system.

3.2. YOLOv4 One-Stage Detector

The architecture of the YOLOv4 one-stage detector is illustrated in Figure 2. The detection process comprises four modules: input, backbone, neck, and head. In this paper, input refers to UAV images containing the target rice seedlings, and the head is implemented through dense prediction to predict the final labels. We will discuss the advantages of each stage, starting with the CSPDarknet53 network in the ‘Backbone’ stage.

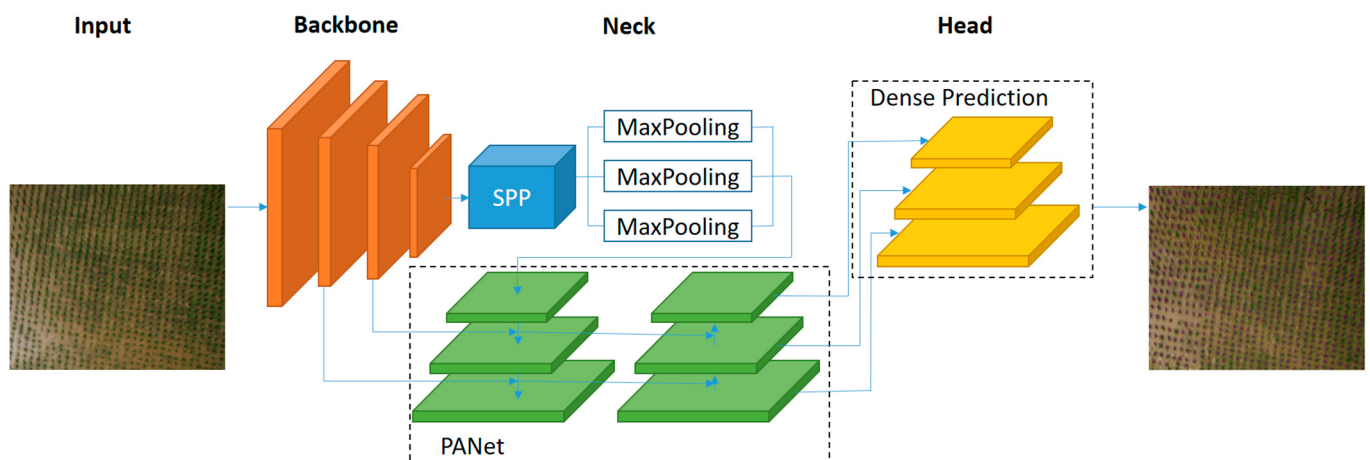


Figure 2. YOLOv4 one-stage detector architecture used in the proposed approach.

3.2.1. Backbone

The Backbone stage of the YOLOv4 One-Stage Detector is implemented using the CSPDarknet53 network. The architecture of CSPDarknet53 is shown in Figure 3 below. CSPDarknet53 exhibits excellent performance in feature extraction, a crucial aspect for accurate object detection. In comparison to CSPResNext50, which excels in image classification, CSPDarknet53 demonstrates superior results in object detection. The architecture of CSPDarknet53 introduces the concept of cross-stage partial connections. At each stage, features are divided into two parts: one part enters the computation block implemented by DenseNet, and the other part is directly sent to the next stage. Finally, these two parts are concatenated. Additionally, a transition layer is added to the backbone to improve parameter utilization. The cross-stage partially connected architecture of CSPDarknet53

improves information flow, enhances feature representation, and reduces computation time. This results in better localization and classification of objects in UAV images.

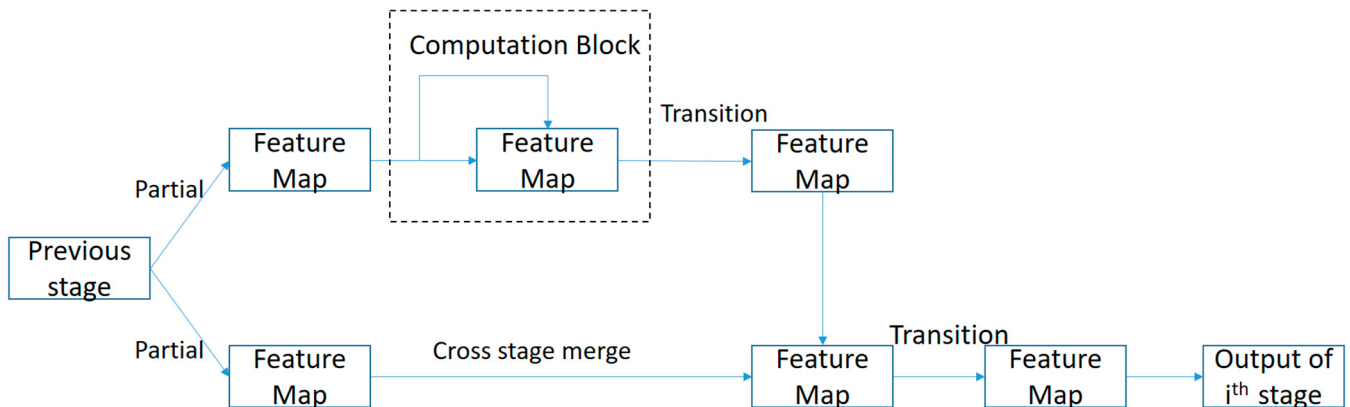


Figure 3. CSPDarknet53 architecture as the backbone of the proposed approach.

3.2.2. Spatial Pyramid Pooling in Neck

The first module in the YOLOv4 architecture's neck utilizes the spatial pyramid pooling (SPP) architecture proposed by Kaiming He et al. [24] as shown in Figure 4. SPP enables the model to accommodate feature maps of different sizes, addressing the limitation of fully-connected layers that traditionally accept feature maps of a fixed size, leading to sub-optimal performance when presented with images of different resolutions. In YOLOv4, the SPP module is enhanced by employing max-pooling with a larger k value ($k = \{1, 5, 9, 13\}$) and concatenating the output. This modification significantly increases the receptive field, contributing to improved model performance. The refined SPP mechanism in YOLOv4 is crucial for handling diverse resolutions in images, resulting in more robust and effective object detection.

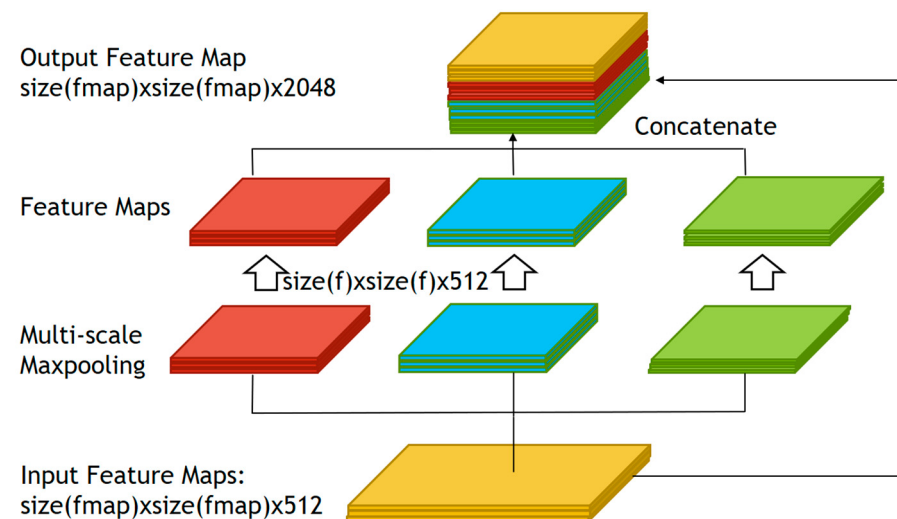


Figure 4. SPP architecture was adopted to deal with the images with multiple resolutions.

3.2.3. Path Aggregation Network in Neck

The second module in the YOLOv4 architecture's neck incorporates the Path Aggregation Network (PANet). The architecture of PANet is shown in Figure 5. In comparison to alternative methods that establish interconnections between layers, the Feature Pyramid Network (FPN) tends to traverse longer paths during information propagation. To address this, PANet is modified by introducing an additional direct path from the bottom to the top. This adjustment facilitates the efficient transmission of information from lower to higher layers, reducing the length of the transmission path. Furthermore, YOLOv4 introduces a

modification to the original PANet architecture. Unlike the original PANet, which adds the feature map to the adjacent layer, YOLOv4 adopts a concatenation approach for the added components. This adjustment aims to reduce computations when the required number of features is fixed. Concatenating adjacent layers not only streamlines computations but also enhances processing speed without significant loss of accuracy.

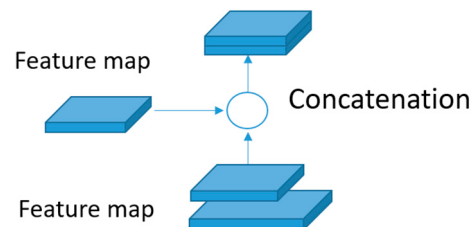


Figure 5. PANet is used in the neck in the proposed approach.

3.3. Activation Function Replacement

In this paper, we propose a modification to the YOLOv4 model by replacing some of the activation functions. The original YOLOv4 configurations predominantly employ Rectified Linear Unit (ReLU) and leaky ReLU in most layers. We conduct a comparative analysis between our modified configuration and the original settings, elucidating the advantages of each activation function employed in our proposed configuration.

3.3.1. ReLU and Leaky ReLU

The activation functions used in the original YOLOv4 are ReLU and leaky ReLU. ReLU, defined by Equation (1), stands out as one of the most widely used activation functions in deep learning models due to its simplicity and effectiveness in solving the vanishing gradient problem and enhancing neural network performance. However, ReLU suffers from the issue of “dead neurons” during training, adversely affecting learning capability and model performance. To solve this concern, leaky ReLU introduces a small slope for negative inputs. The leaky ReLU function, defined by Equation (2), behaves similarly to ReLU for positive inputs, passing them through unchanged. Conversely, for negative inputs, it multiplies the input by a small slope ‘a’, introducing a non-zero gradient that enables the flow of some information. Leaky ReLU effectively prevents the occurrence of “dead neurons”. It also improves the gradient flow during backpropagation, leading to faster convergence, and enhancing model performance. In general, Leaky ReLU helps the model learn to be more stable and robust. However, the effectiveness of Leaky ReLU depends on the correct selection of the ‘negative slope’ parameter. Poor parameter selection may even result in the opposite effect of gradient disappearance. Therefore, Leaky ReLU is not widely adopted across various models.

$$f(x) = \max(0, x) \quad (1)$$

$$f(x) = \max(ax, x) \quad (2)$$

3.3.2. Mish Function

We replaced the activation function with Mish. The Mish activation function is defined by Equation (3). It introduces non-linearity and maps the input values to the range $[-1, 1]$. The softplus function is employed to smooth out the curve, preventing it from saturating at extreme values. In comparison to ReLU, Mish exhibits a smoother and more continuous behavior. The smoothness of Mish’s activation curve contributes to stable and consistent gradient flow, avoiding abrupt transitions and saturation effects that can occur with ReLU during backpropagation. This characteristic can lead to improved convergence and optimization performance. Furthermore, Mish has demonstrated enhanced model performance

and generalization capabilities despite its larger computational cost compared to ReLU. In our task, the use of the Mish activation function results in a significant improvement [25].

$$f(x) = x \times \tanh(\text{softplus}(x)) \quad (3)$$

In particular, we have modified the activation function in the convolutional layers following the shortcut in YOLOv4 to use Mish. Additionally, Mish is employed during up-sampling in the deep layers of the network. Through this approach, especially in the up-sampling layers deep within the network, it facilitates the model in learning complex mappings more effectively. The inclusion of Mish in the convolutional layers following the shortcut is intended to enhance the model's ability to capture fine details and features of objects.

4. Experiments

4.1. Data Preparation

As mentioned earlier, the detection of plant seedlings holds significant potential in agriculture. However, there is still room for improvement in applying object detection to identify tiny objects such as rice seedlings. The prevailing method enhances the labeling performance of rice seedlings through image preprocessing and configuration adjustments in the object detection model. Our experiments on the 2021 AIda “Crop Location Auto-Labeling Competition” datasets demonstrate the enhanced performance of our method compared to existing approaches. A description of the dataset is provided in Table 1.

Table 1. Description of the dataset provided from “Crop Location Auto-Labeling Competition”.

Resolution	Training Set	Testing Set	Total
3000 × 2000	25	30	55
2304 × 1728	19	20	39
Total	44	50	94

This paper utilizes the full-color rice UAV imagery dataset provided by the AIda competition, consisting of 25 rice images with a resolution of 3000 × 2000 and 19 images with a resolution of 2304 × 1728. Due to the large size of the images in the dataset making them impractical for model training, the paper employed a strategy to split the images into smaller, more manageable sizes. Recognizing the inadequacy of the provided training data and the high resolution of the data provided, we augmented the dataset by cropping the images. Starting from the top-left coordinates of the images, we divided the images into non-overlapping segments of sizes 224 × 224 and 896 × 896. Ultimately, the dataset was expanded to a total of 4162 images.

In the initial phase, considering the commonly used default input size of 224 × 224 in backbone networks, our first approach was to employ 224 × 224 as the segmentation size for data augmentation. Additionally, we took into account the computational resources of the hardware, the variation in drone altitude for each image affecting the apparent size of rice plants, and the row spacing in rice cultivation. Since the competition dataset only includes fully visible rice plants for counting, through observations it was noted that using the dimensions of 896 × 896 possesses the of covering a substantial number of plants while preserving the integrity of complete rice plants. This enhances the recognition ability of the model.

4.2. Evaluation Metrics

Since the detection of the rice seedling is not a pixel-to-pixel problem, we must first define a criterion to hit the center of rice seedling by a bounding box. In determining the accuracy of the detection about the rice seedling, we utilize the criterion defined by Equation (4):

$$\text{Euclidean Distance} = D(|GT(x, y) - P(x, y)|) < \varepsilon \quad (4)$$

Here, $GT(x, y)$ represents the correct plant center coordinates in the dataset, $P(x, y)$ denotes the plant center coordinates predicted by the model, and ε signifies a predefined tolerance threshold. Through experimentation, we observed that the average bounding box size for rice plants was often around 60×120 . Consequently, we set ε to 90, representing the length from the center to the corner of a bounding box of size 60×120 . This threshold serves as the criterion for determining whether the predicted center point falls within an acceptable margin of error relative to the ground truth.

The evaluation metrics used in the competition is F1-score. The detailed calculation method for precision, recall, and F1-score is outlined in Equations (5) to (7). Precision and recall are derived through the Euclidean distance between the rice seedling locations predicted by the model and the corresponding ground truth.

$$Precision = \frac{\text{Total points correctly predicted as seedlings (TP)}}{\text{Total points predicted as seedlings (TP + FP)}} \quad (5)$$

$$Recall = \frac{\text{Total points correctly predicted as seedlings (TP)}}{\text{Total points annotated as seedlings in the data (TP + FN)}} \quad (6)$$

$$F1\text{-Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

And to evaluate the seedling counting function, we refer to the accuracy definition used by Wu et al. [26] as Equation (8):

$$Accuracy = \left(1 - \frac{1}{n} \sum_{i=1}^n \frac{|Predicted_n - Groundtruth_n|}{Groundtruth_n} \right) \quad (8)$$

Here, $Groundtruth_n$ and $Predicted_n$ represent the “Total points annotated as seedlings in the data” and the “Total points correctly predicted as seedlings by the model”, respectively, in the n th image.

4.3. Data Augmentations

From Table 1 above, it can be found that the amount of training data provided by the competition is relatively small. To address this limitation, we augmented the training data by isotopically cropping the images, thereby enhancing the learning of the target. After the cropping process, we divided the original dataset, expanding it to 3930 images. The example of data augmentation is shown in Figure 6.

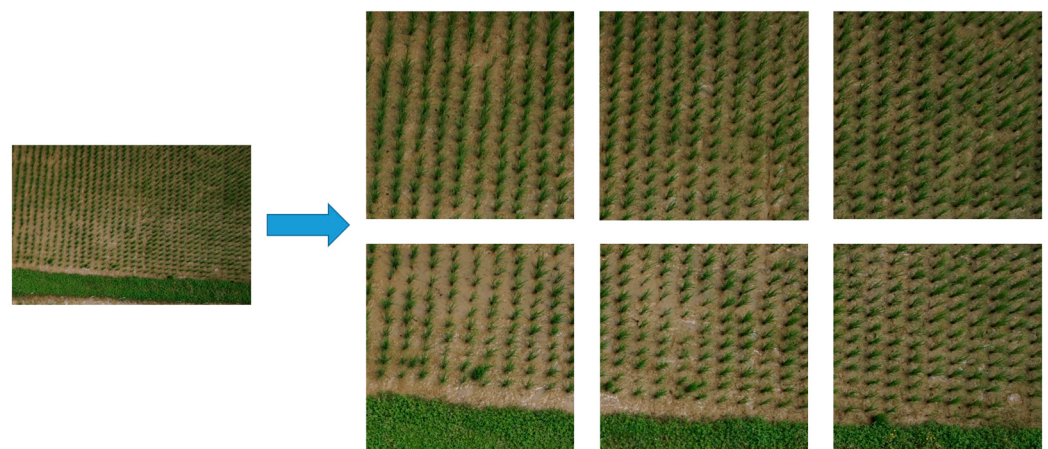


Figure 6. An example of data augmentation adopted in the proposed approach.

4.4. Experimental Results

To evaluate the performance of the developed facial expression recognition system, we equipped a computer with an Intel Core i7-6700 CPU @ 3.40 GHz (8 cores) processor,

an NVIDIA GeForce RTX 2080 Ti GPU, and 32 GB of RAM as the hardware platform for conducting our experiments.

Our experimental design is divided into three parts. First, we use the initial YOLOv4 as a baseline for comparison to assess the effectiveness of the proposed method (ablation study). Next, we compare our plant counting approach with others. Finally, we compare with state-of-the-art methods using average precision as the evaluation metric.

The experimental results of the ablation study are presented in Table 2. Initially, we trained the model using the original data and the original YOLOv4 configurations as our baseline. Subsequently, we conducted experiments by augmenting the data through various sizes of image cropping and adjusting the configurations for comparative analysis.

Table 2. Ablation study experimental results.

Experiments	Data Augmentation	Configurations	F1-Score	Acc
Model 1 (baseline)	Original dataset	Original YOLOv4	0.63	0.68
Model 2	Splits into 224×224	Width:1024 Height:1024	0.65	0.73
Model 3	Splits into 224×224 +Original dataset	Width:1024 Height:1024	0.69	0.78
Model 4	Splits into 896×896	Width:1024 Height:1024	0.72	0.85
Model 5	Splits into 896×896 +Original dataset	Width:1024 Height:1024	0.84	0.91
Model 6	Splits into 896×896 +Original dataset	Width:1024 Height:1024 +Mish	0.91	0.97

Upon examining Table 2, it is evident that this method enhances the F1-score by 0.28 compared to the baseline model using the original dataset and configuration. In the data augmentation experiments, we can observe that employing larger sizes for image segmentation yields better results. This is attributed to the high density of rice seedlings in UAV images; using excessively small sizes to segment images may inadvertently cut through rice plant boundaries, negatively impacting learning. Conversely, larger sizes improve the ability to distinguish and exclude other objects. Furthermore, it is observed that despite the lack of original data, the performance of using small-sized cuttings is still superior to using the original dataset. Irrespective of the cutting size, combining the cut data with the original dataset achieves better performance. Finally, the model using Mish as the activation function has achieved a significant improvement in F1-score. In the course of our model development, the training phase spanned a total duration of 20 h, wherein the model underwent iterative learning processes. Subsequently, during the testing phase, the annotation of each image demonstrated an impressive efficiency, averaging a mere 200 milliseconds per image. This noteworthy performance underscores the efficacy of our proposed methodology in achieving swift and accurate results in real-world applications.

Table 3 provides a reference for accuracy, where it is observed that our approach has improved the accuracy by 0.29 compared to the baseline YOLOv4. When compared to the work of Wu et al. [26], our method outperforms theirs in terms of accuracy. Additionally, in comparison to Wu et al., our model is capable of annotating the coordinates of the targets.

Table 3. Comparing other models with counting accuracy.

Experiments	Main Method	Position Labeling	Accuracy
Model 1 (baseline)	Yolov4	+	0.68
Wu et al. [26]	Regress the density map	–	0.81
Model 6	Modified YOLOv4 with Mish	+	0.97

+ / – means the approach with/without the position labeling capability separately.

Table 4 presents average precision as the evaluation metric and compares it with the one proposed by Tatini et al. [17]. It can be observed that our approach has achieved a 4.96% improvement in average precision compared to Tatini et al.

Table 4. Comparing other models with average precision.

Experiments	Main Method	Recall	Average Precision
Model 1 (baseline)	YOLOv4	0.58	84.64%
Tatini et al. (2022) [17]	YOLOv4-SUFF	–	86.78%
Model 6	YOLOv4 with Mish	0.90	91.74%

– means the approach without the recall.

5. Conclusions

This study proposes a UAV image-based rice seedling labeling method for precision agriculture using YOLOv4. By segmenting the dataset, we increase the quantity of images for a limited amount of training data, thereby improving the recognition performance for tiny objects in UAV images. The activation function Mish replaces the activation function used in YOLOv4, thereby enabling YOLOv4 to overcome limitations in detecting tiny objects. Furthermore, we propose a criterion for determining the detection accuracy of rice seedlings, aiming to evaluate the system. In the experiments, compared to the original YOLOv4 model, the F1-Score achieved was 0.91, showing an improvement of 0.28 over the baseline YOLOv4. There were also enhancements in accuracy and precision, with increases of 0.29 and 7.1%, respectively. This not only validates the effectiveness of our system but also establishes it as a state-of-the-art model. Beyond the agricultural domain, our approach is expected to apply to other applications such as crowd counting [27]. Therefore, in the future, we plan to extend our methodology to develop applications in various domains.

Author Contributions: Conceptualization, K.-M.L. and J.-M.H.; Methodology, J.-F.Y. and L.-C.Y.; Software, L.-C.Y.; Writing—original draft, L.-C.Y.; Writing—review & editing, J.-F.Y.; Project administration, J.-F.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Science and Technology Council grant number 111-2221-E-415-012-MY3.

Data Availability Statement: Data is contained within the article.

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

References

1. Jiang, M.; Feng, C.; Fang, X.; Huang, Q.; Zhang, C.; Shi, X. Rice Disease Identification Method Based on Attention Mechanism and Deep Dense Network. *Electronics* **2023**, *12*, 508. [\[CrossRef\]](#)
2. Ammar, A.; Koubaa, A.; Benjdira, B. Deep-Learning-Based Automated Palm Tree Counting and Geolocation in Large Farms from Aerial Geotagged Images. *Agronomy* **2021**, *11*, 1458. [\[CrossRef\]](#)
3. Luo, X.; Tian, X.; Zhang, H.; Hou, W.; Leng, G.; Xu, W.; Jia, H.; He, X.; Wang, M.; Zhang, J. Fast Automatic Vehicle Detection in UAV Images Using Convolutional Neural Networks. *Remote Sens.* **2020**, *12*, 1994. [\[CrossRef\]](#)
4. Wang, L.; Xiang, L.; Tang, L.; Jiang, H. A Convolutional Neural Network-Based Method for Corn Stand Counting in the Field. *Sensors* **2021**, *21*, 507. [\[CrossRef\]](#)
5. Luo, X.; Wu, Y.; Wang, F. Target Detection Method of UAV Aerial Imagery Based on Improved YOLOv5. *Remote Sens.* **2022**, *14*, 5063. [\[CrossRef\]](#)

6. Yang, R.; Zhang, J.; Shang, X.; Li, W. Lightweight Small Target Detection Algorithm with Multi-Feature Fusion. *Electronics* **2023**, *12*, 2739. [[CrossRef](#)]
7. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
8. Liu, J.; Abbas, I.; Noor, R.S. Development of deep learning-based variable rate agrochemical spraying system for targeted weeds control in strawberry crop. *Agronomy* **2021**, *11*, 1480. [[CrossRef](#)]
9. Li, G.S.; Wu, D.H.; Su, Y.C.; Kuo, B.J.; Yang, M.D.; Lai, M.H.; Lu, H.Y.; Yang, C.Y. Prediction of plant nutrition state of rice under water-saving cultivation and panicle fertilization application decision making. *Agronomy* **2021**, *11*, 1626. [[CrossRef](#)]
10. Oliveira, M.F.D.; Ortiz, B.V.; Morata, G.T.; Jiménez, A.F.; Rolim, G.D.S.; Silva, R.P.D. Training Machine Learning Algorithms Using Remote Sensing and Topographic Indices for Corn Yield Prediction. *Remote Sens.* **2022**, *14*, 6171. [[CrossRef](#)]
11. Hu, Q.; Li, L.; Duan, J.; Gao, M.; Liu, G.; Wang, Z.; Huang, D. Object Detection Algorithm of UAV Aerial Photography Image Based on Anchor-Free Algorithms. *Electronics* **2023**, *12*, 1339. [[CrossRef](#)]
12. Jawaharlalnehru, A.; Sambandham, T.; Sekar, V.; Ravikumar, D.; Loganathan, V.; Kannadasan, R.; Khan, A.A.; Wechtaisong, C.; Haq, M.A.; Alhussen, A. Target Object Detection from Unmanned Aerial Vehicle (UAV) Images Based on Improved YOLO Algorithm. *Electronics* **2022**, *11*, 2343. [[CrossRef](#)]
13. Tseng, H.H.; Yang, M.D.; Saminathan, R.; Hsu, Y.C.; Yang, C.Y.; Wu, D.H. Rice Seedling Detection in UAV Images Using Transfer Learning and Machine Learning. *Remote Sens.* **2022**, *14*, 2837. [[CrossRef](#)]
14. Liu, M.; Wang, X.; Zhou, A.; Fu, X.; Ma, Y.; Piao, C. UAV-YOLO: Small Object Detection on Unmanned Aerial Vehicle Perspective. *Sensors* **2020**, *20*, 2238. [[CrossRef](#)]
15. Shen, H.; Lin, D.; Song, T. Object Detection Deployed on UAVs for Oblique Images by Fusing IMU Information. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [[CrossRef](#)]
16. Amarasingam, N.; Gonzalez, F.; Salgadoe, A.S.A.; Sandino, J.; Powell, K. Detection of White Leaf Disease in Sugarcane Crops Using UAV-Derived RGB Imagery with Existing Deep Learning Models. *Remote Sens.* **2022**, *14*, 6137. [[CrossRef](#)]
17. Tatini, N.B.; Lu, G.Y.; Tan, T.H.; Alkhaleefah, M.; Koo, V.C.; Chan, Y.K.; Chang, Y.L. Yolov4 Based Rice Fields Classification from High-Resolution Images Taken by Drones. In Proceedings of the IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 5043–5046.
18. Sun, B.; Zhou, W.; Zhu, S.; Huang, S.; Yu, X.; Wu, Z.; Lei, X.; Yin, D.; Xia, H.; Ren, W. Universal detection of curved rice panicles in complex environments using aerial images and improved YOLOv4 model. *Front. Plant Sci.* **2022**, *13*, 1021398. [[CrossRef](#)]
19. Zhang, S.; Qu, C.; Ru, C.; Wang, X.; Li, Z. Multi-Objects Recognition and Self-Explosion Defect Detection Method for Insulators Based on Lightweight GhostNet-YOLOV4 Model Deployed Onboard UAV. *IEEE Access* **2023**, *11*, 39713–39725. [[CrossRef](#)]
20. Du, S.; Zhang, P.; Zhang, B.; Xu, H. Weak and Occluded Vehicle Detection in Complex Infrared Environment Based on Improved YOLOv4. *IEEE Access* **2021**, *9*, 25671–25680. [[CrossRef](#)]
21. Li, K.; Zhuang, Y.; Lai, J.; Zeng, Y. PFYOLOv4: An Improved Small Object Pedestrian Detection Algorithm. *IEEE Access* **2023**, *11*, 17197–17206. [[CrossRef](#)]
22. Zhang, Y.; Lu, X.; Li, W.; Yan, K.; Mo, Z.; Lan, Y.; Wang, L. Detection of Power Poles in Orchards Based on Improved Yolov5s Model. *Agronomy* **2023**, *13*, 1705. [[CrossRef](#)]
23. Wang, J.; Zhang, F.; Zhang, Y.; Liu, Y.; Cheng, T. Lightweight Object Detection Algorithm for UAV Aerial Imagery. *Sensors* **2023**, *23*, 5786. [[CrossRef](#)]
24. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)]
25. Mishra, D. Mish: A Self Regularized Non-Monotonic Activation Function. *arXiv* **2019**, arXiv:1908.08681.
26. Wu, J.; Yang, G.; Yang, X.; Xu, B.; Han, L.; Zhu, Y. Automatic Counting of in situ Rice Seedlings from UAV Images Based on a Deep Fully Convolutional Neural Network. *Remote Sens.* **2019**, *11*, 691. [[CrossRef](#)]
27. Dong, L.; Zhang, H.; Yang, K.; Zhou, D.; Shi, J.; Ma, J. Crowd Counting by Using Top-k Relations: A Mixed Ground-Truth CNN Framework. *IEEE Trans. Consum. Electron.* **2022**, *68*, 307–316. [[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.