

A UAV Open Dataset of Rice Paddies for Deep Learning Practice

Ming-Der Yang ^{1,2}, Hsin-Hung Tseng ^{1,2,*}, Yu-Chun Hsu ^{1,2}, Chin-Ying Yang ³, Ming-Hsin Lai ⁴ and Dong-Hong Wu ⁴

¹ Department of Civil Engineering, and Innovation and Development Center of Sustainable Agriculture, National Chung Hsing University, Taichung 402, Taiwan; mdyang@nchu.edu.tw (M.-D.Y.), d107062002@mail.nchu.edu.tw (Y.-C.H.)

² Pervasive AI Research (PAIR) Labs, Hsinchu 300, Taiwan

³ Department of Agronomy, National Chung Hsing University, Taichung 40227, Taiwan; emiyang@nchu.edu.tw

⁴ Crop Science Division, Taiwan Agricultural Research Institute, Taichung 413, Taiwan; mhlai@tari.gov.tw (M.-H.L.); dhwu@tari.gov.tw (D.-H.W.)

* Correspondence: d108062001@mail.nchu.edu.tw; Tel.: +886-4-22840440 (ext. 300)

Abstract: Recently, unmanned aerial vehicles (UAVs) have been broadly applied to the remote sensing field. For a great number of UAV images, deep learning has been reinvigorated and performed many results in agricultural applications. The popular image datasets for deep learning model training are generated for general purpose use, in which the objects, views, and applications are for ordinary scenarios. However, UAV images possess different patterns of images mostly from a look-down perspective. This paper provides a verified annotated dataset of UAV images that are described in data acquisition, data preprocessing, and a showcase of a CNN classification. The dataset collection consists of one multi-rotor UAV platform by flying a planned scouting routine over rice paddies. This paper introduces a semi-auto annotation method with an ExGR index to generate the training data of rice seedlings. For demonstration, this study modified a classical CNN architecture, VGG-16, to run a patch-based rice seedling detection. The k-fold cross-validation was employed to obtain an 80/20 dividing ratio of training/test data. The accuracy of the network increases with the increase of epoch, and all the divisions of the cross-validation dataset achieve a 0.99 accuracy. The rice seedling dataset provides the training-validation dataset, patch-based detection samples, and the ortho-mosaic image of the field.

Keywords: open dataset; deep learning; CNN; training data; UAV images; rice seedling

Citation: Yang, M.-D.; Tseng, H.-H.; Hsu, Y.-C.; Yang, C.-Y.; Lai, M.-H.; Wu, D.-H. A UAV Open Dataset of Rice Paddies for Deep Learning Practice. *Remote Sens.* **2021**, *13*, 1358. <https://doi.org/10.3390/rs13071358>

Academic Editor: Francesco Nex

Received: 10 February 2021

Accepted: 30 March 2021

Published: 1 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Underlying the global climate change and a two billion increase of world population in next projected 30 years [1,2], sufficient yielding of grain crops has been considered in many countries as one of the most important issues to maintain food security. Remote sensing for land use [3–6] and agricultural monitoring [7–9] from satellites have been greatly adopted since the space era [10]. Satellites carry multispectral sensors, hyperspectral sensors, panchromatic sensors, and synthetic aperture radar that have been widely used for land use classification, agricultural monitoring and management, and disaster assessment [11–14]. The often-used satellites, such as Landsat, SPOT, Sentinel, and RADARSAT, provide a monthly- to weekly-level revisiting cycle and up to meter-level spatial resolution [15–18]. However, limited by the temporal and spatial resolution, satellite images usually cannot provide real-time and highly detailed data for precision agriculture [19]. Thanks to the development of mechanical and electronic techniques, unmanned aerial vehicles (UAVs) have been broadly applied to the remote sensing field. Compared to satellite remote sensing, UAVs process many advantages, such as ultra-high spatial reso-

lution, flexible monitoring ability, and reasonable cost. Thus, UAVs have performed various notable applications on combining multispectral data, thermal data, and field information to classify crop species, assess disasters, and monitor plant growth [20–23].

With the development of computing power and a great number of UAV images, deep learning techniques have been reinvigorated and performed many results in agricultural applications. Egli and Höpke [24] developed a lightweight convolutional neural network (CNN) for automated tree species classifying with high-resolution UAV images. Chen et al. [25] applied an object detection network on counting strawberries with ultra-high-resolution UAV images for yield prediction. Yang et al. [26] applied deep-learning to UAV images to estimate rice lodging over a vast area. Li et al. [27] proposed an improved object detection model for high-precision detection of hydroponic lettuce seedlings. Pearse et al. [28] applied a CNN model for tree seedlings detection to map and monitor the regeneration of forests in UAV images. Oh et al. [29] applied the object detection technique to cotton seedling counting in UAV images to analyze the plant density for precision field management.

Although deep-learning applications on UAVs are numerous, the UAV datasets vary with applications and are limited for free access. The commonly used image datasets are CIFAR-10, ImageNet-1000, and COCO [30–32], which were released by Canadian Institute For Advanced Research (CIFAR), ImageNet Large Scale Visual Recognition Challenge (ILSVRC), and Microsoft, respectively. Images in the above-mentioned datasets are for general purpose use, in which the objects, views, and applications are for ordinary scenarios. The images acquired from the UAVs are mostly from a look-down perspective. The significant difference of viewing angles on the same objects results in a different context that degrades the applicability of the general-use dataset on UAVs deep-learning applications.

Rice is one of the major grain crops worldwide, over half of the world's population consumes rice as the staple food, and over 85% of consumption accounts for Asia [33,34]. To precisely estimate the grain yield and quality of rice, exploring the hill number of rice seedlings is a key component for cultivation density and uniform maturity of precision agriculture. This paper collected UAV images of rice seedlings in-field at the early stage of growth with a UAVs' look-down perspective. For demonstration, the rice seedling dataset was adopted to identify the number and position of rice seedlings using a lightweight CNN classification architecture. The proposed CNN model is trained with a 5-fold cross-validation dataset, which reduces the effect of bias data on the model. In addition, the performance is evaluated by classification accuracy.

The aim of this paper is to provide a platform of UAV image dataset of rice paddy for data sharing by making labeled and unlabeled data findable and accessible through domain-specific repositories. For this scope, this paper focuses on the description of the dataset, including what methods used for collecting and producing the data, where the dataset may be found, and how to use the data with useful information and a showcase.

2. Dataset Description

This section introduces the data descriptors of the rice seedling dataset available at <https://github.com/aipal-nchu/RiceSeedlingDataset>.

2.1. Data Introduction

The datasets published at GitHub consists of orthomosaic image, training-validation dataset, and the demo dataset. The orthomosaic image (see Figure 1) is the image stitched from a series of nadir-like view UAV images. The dataset provides 13 images for consecutive growth stages, which were imaged in 2018, 2019, and 2020 as listed in Table 1. All the images are georeferenced in TWD97 / TM2 zone 121 (EPSG: 3826) projected coordinate system. The training-validation dataset (green bounding area) was generated by the method discussed in Section 2.4, and images were saved under a specific subfolder by each class. The demonstration dataset (red bounding area) is used for the test of object

detection. This study clipped eight square images with an $8\text{ m} \times 8\text{ m}$ area, which contains approximately one thousand hills of rice seedlings in each image. The detail of the demo dataset is discussed in Section 3.1.

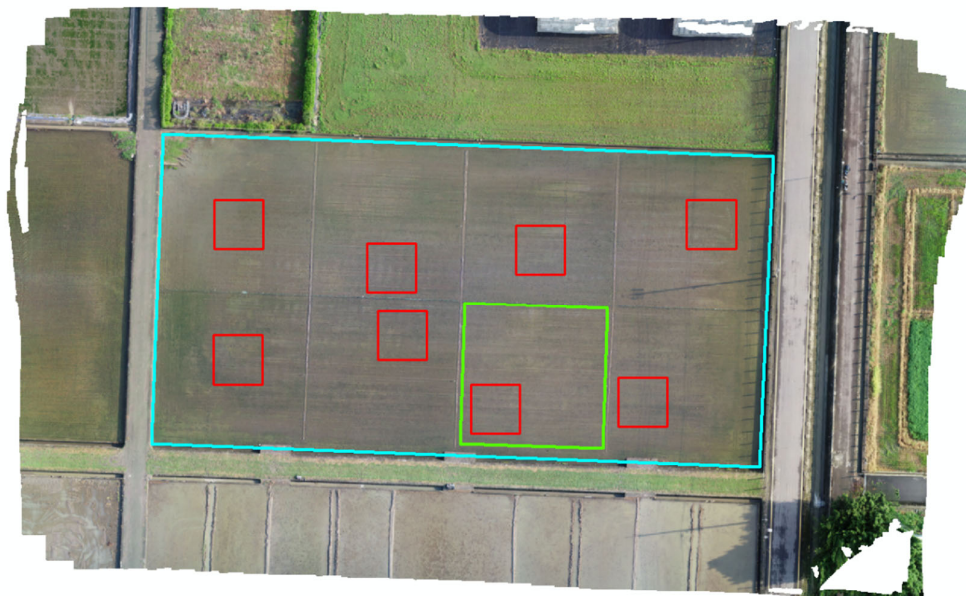


Figure 1. An overview of field No. 80 (cyan bounding area). Image acquired on 7th August 2018. The green bounding area represents the area for training-validation dataset, and the red bounding area represents the area for the object detection demonstration dataset.

Table 1. A description of image files published on GitHub.

Filename	Description	Disk Space
2018-08-07_ARI80_20m_Orthomosaic.tif	orthomosaic image	465 MB
2018-08-14_ARI80_20m_Orthomosaic.tif	orthomosaic image	610 MB
2018-08-23_ARI80_20m_Orthomosaic.tif	orthomosaic image	556 MB
2019-03-26_ARI78_20m_Orthomosaic.tif	orthomosaic image	485 MB
2019-04-02_ARI78_20m_Orthomosaic.tif	orthomosaic image	418 MB
2019-08-12_ARI78_20m_Orthomosaic.tif	orthomosaic image	503 MB
2019-08-20_ARI78_20m_Orthomosaic.tif	orthomosaic image	605 MB
2020-03-12_ARI78_40m_Orthomosaic.tif	orthomosaic image	278 MB
2020-03-17_ARI78_40m_Orthomosaic.tif	orthomosaic image	317 MB
2020-03-26_ARI78_40m_Orthomosaic.tif	orthomosaic image	385 MB
2020-08-12_ARI78_40m_Orthomosaic.tif	orthomosaic image	330 MB
2020-08-18_ARI78_40m_Orthomosaic.tif	orthomosaic image	382 MB
2020-08-25_ARI78_40m_Orthomosaic.tif	orthomosaic image	402 MB
RiceSeedlingClassification.tgz	training-validation dataset	426 MB
RiceSeedlingDetection.tgz	detection training dataset	10.9MB
RiceSeedlingDemo.tgz	detection demonstration dataset	48.5MB

2.2. Training-Validation Dataset

The training-validation dataset was collected by a multi-rotor UAV flying a planned scouting routine over a paddy operated by Taiwan Agricultural Research Institute (TARI) in Wufeng District, Taichung, Taiwan. The data were collected on 7, 14, and 23 August 2018, between 07:03 am and 08:00 am local time. The UAV flew at a constant altitude and carried an RGB sensor with an approximate nadir view for the duration of data collection

using a 4-rotor commercial range UAV, DJI Phantom 4 Pro (Da-Jiang Innovations, Shenzhen, PRC) [35]. The equipped sensor is an RGB sensor with a 1-inch diagonal size and a focal length of 8.8 mm. The sensor parameters are listed in Table 2.

The UAV flew nominally at a 20 m altitude above ground to generate a spatial resolution of 5.3 mm/pixel. The ground speed was between 1.8 and 2.2 m/s and was relatively constant during data collection. Figure 2a depicts the data collecting area over a satellite image, and Figure 2b depicts the flight routes (white dots) and orthomosaic image overlapped on the satellite image. The designed route overlap was 80% and side-overlap was 75%, resulting in a total number of 349, 299, and 443 images, respectively. The detail of the training-validation data collection missions is listed in Table 3.

2.3. Expansion Dataset

To test the impact of environmental disturbances, additional UAV datasets acquired in 2019 and 2020 were also provided. The data were acquired in field No. 78, which is located next to field No. 80. In 2020, image data were acquired from an RGB sensor, DJI Zenmuse X7, which is an interchangeable lens camera and is equipped with a lens with 24 mm focal length [36]. The detail of this sensor is listed in Table 2. The designed flight height was 40 m subject to a narrow FOV and a high sensor resolution to acquire the approximately same spatial resolution as the 2018 and 2019 UAV datasets.

The expansion data provide more UAV paddy images for challenging test. Amongst, several images appear the influence of environmental disturbances, such as the variety of illuminations, weather, soil moisture, and seedling sizes, and the presence of algae, that can be treated as expansion image datasets. Some examples were shown in Figure 3. To adapt to these disturbances, users could augment the data through photometric and geometric transformations or add noise to the original training set to learn more robust features [37]. The detail of the expansion data acquisition missions is listed in Table 4.

Table 2. Unmanned aerial vehicle (UAV) sensor parameters.

Sensor	DJI Phantom 4 Pro [35]	DJI Zenmuse X7 [36]
Resolution (H × V)	5472 × 3648	6016 × 4008
FOV (H° × V°)	73.7° × 53.1°	52.2° × 36.2°
Focal Length (mm)	8.8	24
Sensor Size (H × V mm)	13.2 × 8.8	23.5 × 15.7
Pixel Size (μm)	2.41 × 2.41	3.99 × 3.99
Image Format	JPG	JPG
Dynamic Range	8 bit	8 bit

Table 3. Information of UAV training-validation data collection missions.

Study Area	No. 80 field		
Sensor	DJI Phantom 4 Pro		
Acquisition Date	7th August 2018	14th August 2018	23rd August 2018
Time	07:19–07:32	07:03–07:13	07:41–08:00
Weather	Mostly clear	Mostly clear	Partly Cloudy
Avg. Temperature (°C)	28.7	27.8	28.6
Avg. Press (hPa)	997.7	992.2	987.9
Flight Height (m)	21.4	20.8	22.9
Spatial Resolution (mm/pixel)	5.24	5.09	5.57
Forward Overlap (%)	80	80	80
Side Overlap (%)	75	75	80
Collected Images	349	299	443
Coverage Area (ha)	1.38	1.18	1.33

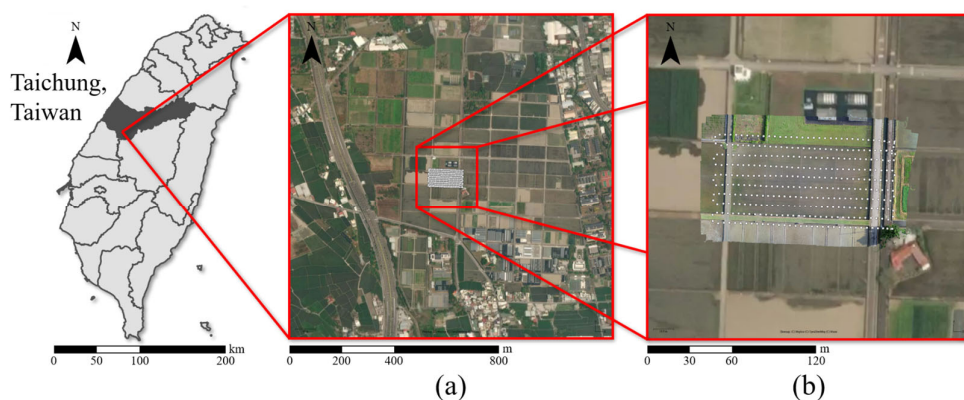


Figure 2. (a) The data collecting area over a satellite image; (b) the flight routes (white dots) and orthomosaic image.

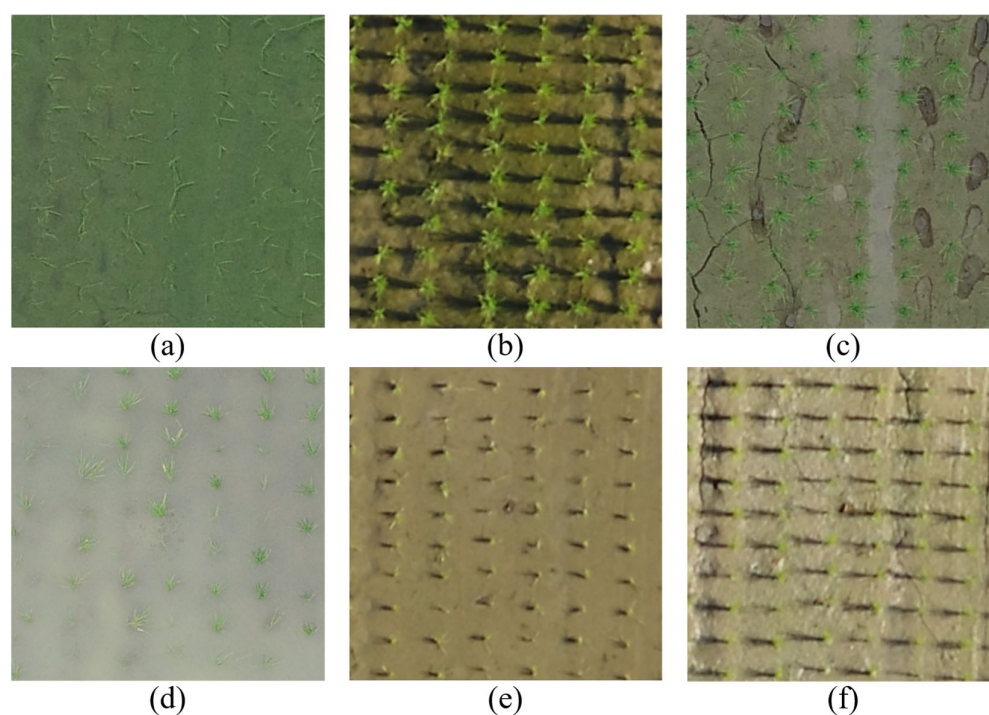


Figure 3. Examples of various environmental disturbances. (a) presence of algae and low contrast; (b) presence of algae and high contrast; (c) low contrast and low illumination. (d) low contrast, high illumination, and flooded paddy; (e) small seedling size and wet soil; (f) small seedling size and dry soil.

Table 4. Information of UAV expansion data collection missions.

Study Area		No. 78 field								
Sensor	DJI Phantom 4 Pro						DJI Zenmuse X7			
Acquisition Date	26th March 2019	2nd April 2019	12th August 2019	20th August 2019	12th March 2020	17th March 2020	26th March 2020	12th August 2020	18th August 2020	25th August 2020
Time	09:40–10:05	09:19–09:48	14:23–14:44	08:16–08:36	09:54–10:07	09:27–09:42	08:58–09:12	09:00–09:12	08:34–08:46	08:16–08:29
Weather	Clear	Cloudy	Cloudy/ occasional rain	Partly cloudy	Partly cloudy	Clear	Clear	Clear	Clear	Clear
Avg. Temperature (°C)	22.6	21.2	29.1	28.5	22.0	23.6	27.8	32.4	29.8	30.7
Avg. Press (hPa)	1011.7	1011.3	994.2	997.8	1009.8	1011.5	1006.9	1005.2	999.2	996.3

Flight Height (m)	20.2	21.3	18.6	19.1	42.2	41.9	42.0	41.8	40.2	40.2
Spatial Resolution (mm/pixel)	5.04	5.33	4.62	4.78	6.38	6.38	6.38	6.37	6.36	6.36
Forward Overlap (%)	80	80	80	80	80	80	80	80	80	80
Side Overlap (%)	80	80	80	80	80	80	80	80	80	80
Collected Images	583	631	615	596	250	250	250	250	250	251
Coverage Area (ha)	1.17	1.25	1.17	1.18	1.59	1.60	1.58	1.59	1.59	1.60

2.4. Data Preprocessing

UAV images were orthorectified and stitched through a commercial software, Agisoft Metashape (St. Petersburg, Russia) [38], to form a single orthomosaic image. To extract the rice seedlings rapidly, this paper introduced a semi-auto annotation method through an excess-green-minus-excess-red index (ExGR) to enhance the greenness of the images [39]. Yen's thresholding method was applied to obtain a binary map [40]. Then, a morphological process was employed to enhance the object features, and then the centric point of every object was calculated using the contour extraction from the OpenCV library [41]. Finally, rice seedling objects can be subset and saved as single images one by one, or generate the annotations for object detection training set. The workflow of preprocessing is shown in Figure 4.

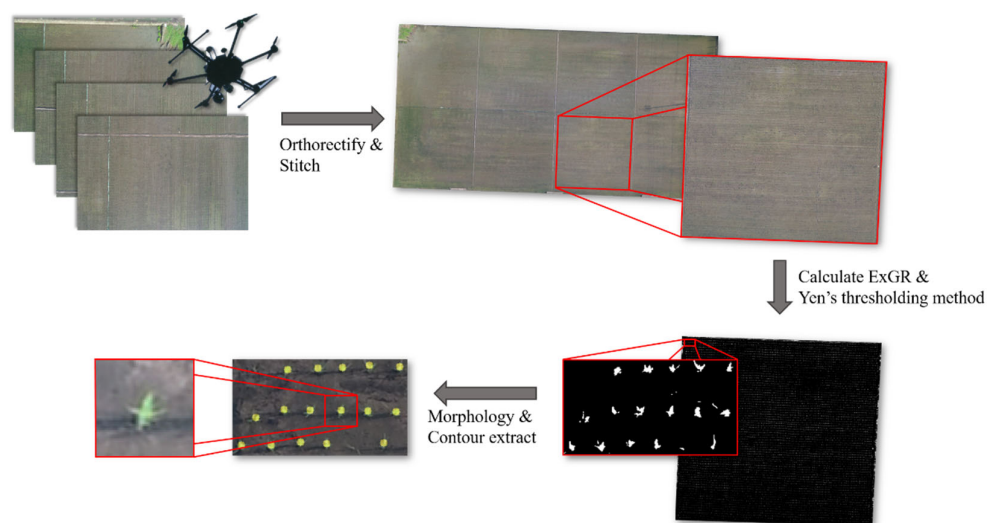


Figure 4. The workflow of semi-auto annotation.

2.5. UAV Dataset of Rice Seedling Classification

One paddy image selected from the UAV dataset acquired on 7th August 2018 is adopted as training data for rice seedling classification. Training samples of UAV images extracted by binarization and morphological processing (discussed in Section 2.4) were manually verified by agricultural experts. The classes of the UAV images in this dataset are categorized into rice seedling and arable land, in which each class contains 28K and 26.5K samples, respectively. The dataset comprises two annotated classes (Figure 5), 54.6K samples in total, and 48×48 pixels in size of each image.

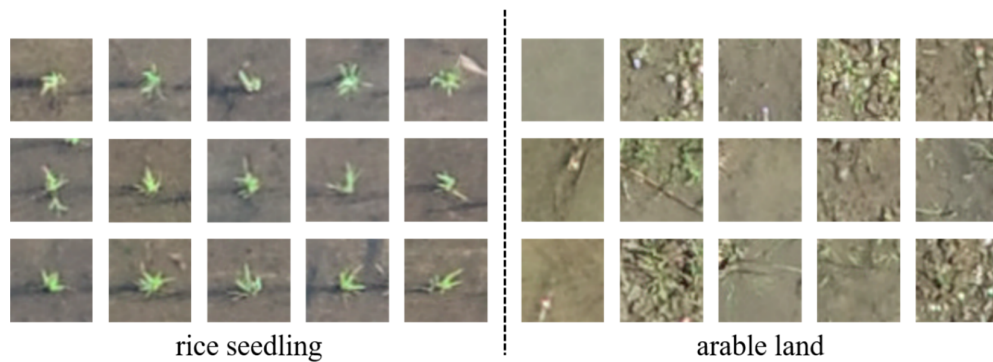


Figure 5. Examples of classes of the rice seedling classification dataset.

Table 5 shows the number of samples for each class for training, validation, and testing of classification. The dataset was split in an 80/20 ratio of training/test data, which is the most commonly adopted in deep learning applications [42]. Besides, a 10% subset of the test samples was used to validate the training result. A total of 43.7K samples were used for training. 1.1K samples and 9.8K samples were used for validation and testing, respectively.

Table 5. The number of images used for training, validation, and testing in the rice seedling dataset.

Class	Training Samples	Validation Samples	Testing Samples	Total Samples
Rice Seedling	22,438	561	5048	28,047
Arable land	21,265	532	4784	26,581
Total	43,703	1093	9832	54,628

2.6. UAV Dataset of Rice Seedling Detection

In this paper, annotations of object detection were provided with three serial missions, 7th, 14th, and 23rd August 2018. The training and validation images were cropped from eight subsets into 600 training samples, in which each subset generates 25 training samples with a size of 320×320 pixels and each sample contains approximately 50 seedlings. The annotations were generated in PASCAL VOC [43] format by a graphical image annotation tool, Labelling [44]. An example of these XML files is given in Appendix A to show the information about image size, classes, and coordinates of bounding boxes. Examples of three growth stages of the rice seedling detection dataset were shown in Figure 6.

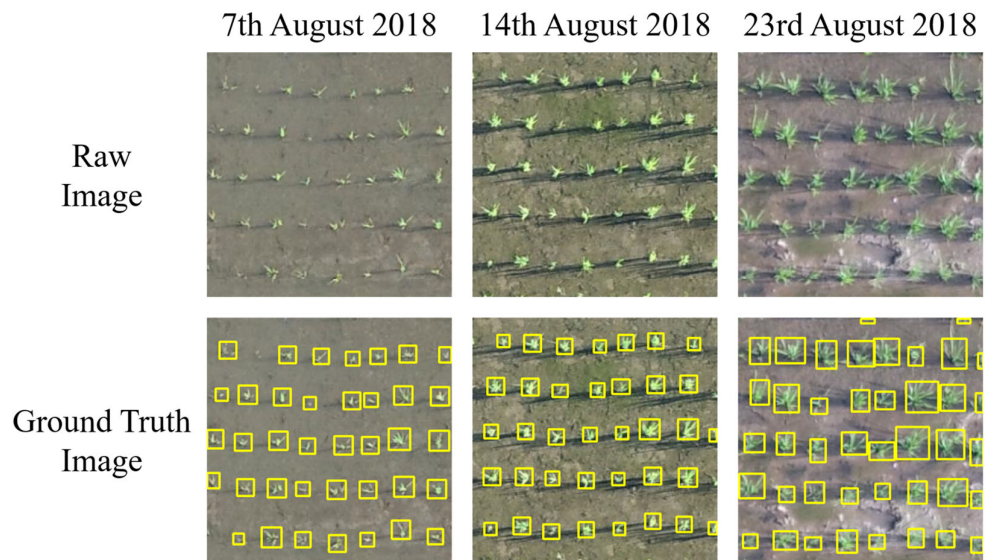


Figure 6. Examples of three growth stages of the rice seedling detection dataset.

3. Data Application

The rice seedling dataset was fed to a deep learning classifier for training. The training phase involves hyperparameter tuning, including learning rates, decay ratio of learning rates, batch sizes, and the number of epochs. This study modified a classical CNN architecture, VGG-16 [45], to demonstrate a simple classification.

3.1. Demonstration of Rice Seedling Detection

To demonstrate data application to patch-based object detection scenarios, this paper clipped 8 images from the orthomosaic image (Figure 7) with a region of 8×8 meters and a size of 1527×1527 pixels for each image. The object detection annotation of ground truth is also provided for the eight demo images in PASCAL VOC.

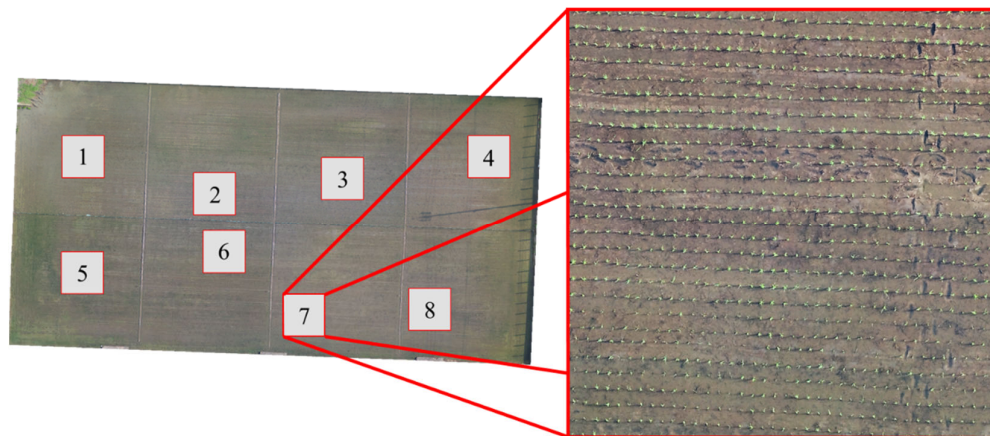


Figure 7. Subset 7 was clipped from the orthomosaic image. (The spotlighted image is enhanced in contrast for a clearer view, not in true color.).

3.2. Classification Model

This paper performed the image classification with the dataset using a convolutional neural network (CNN) algorithm, which was modified from the classical algorithm, VGG-16, due to its promising classification architecture. The model was redesigned with a relatively simple network structure by keeping the iconic stack-convolution structure but

reducing the number of convolution layers, filters, and fully-connected layers that decrease the number of parameters in the training phase to mitigate an overfitting problem. The visualized architecture of the network is shown in Figure 8. The input images are in 48×48 pixels and contain three visible bands (R, G, B). Table 6 shows the layer parameters of the model.

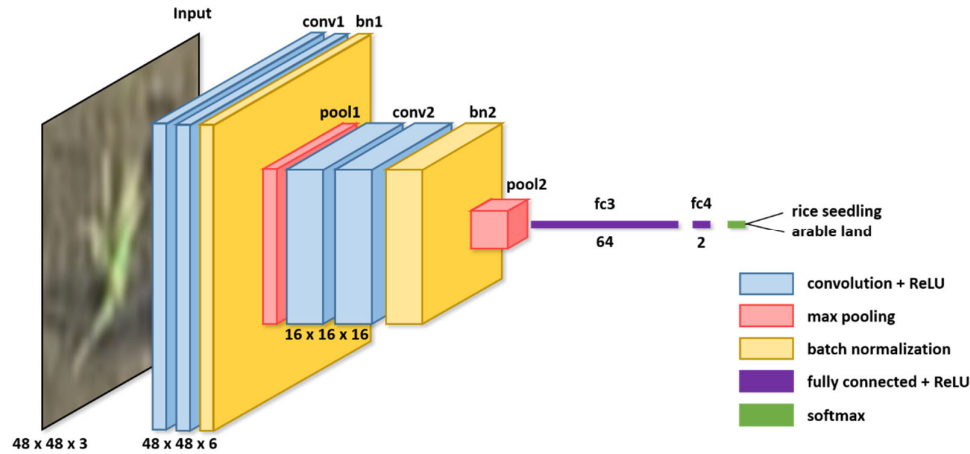


Figure 8. The architecture of the proposed network for rice seedling classification.

Table 6. Layer parameters for the proposed network.

Layer	Parameter	Activation Function
Input	$48 \times 48 \times 3$	—
Convolution 1_1 (conv1_1)	6 filters (3×3), stride 1, padding same	ReLU
Convolution 1_2 (conv1_2)	6 filters (3×3), stride 1, padding same	ReLU
Batch Normalization 1 (bn1)	—	—
Pooling 1 (pool1)	Max pooling (3×3), stride 3	—
Convolution 2_1 (conv2_1)	16 filters (3×3), stride 1, padding same	ReLU
Convolution 2_2 (conv2_2)	16 filters (3×3), stride 1, padding same	ReLU
Batch Normalization 2 (bn2)	—	—
Pooling 2 (pool2)	Max pooling (4×4), stride 4	—
Flatten	—	—
Full Connect 3 (fc3)	64 nodes	ReLU
Dropout	Dropout rate 0.1	—
Full Connect 4 (fc4)	2 nodes	ReLU
Output	—	Softmax

The layers in CNN are defined as follows:

1. The first two convolution layers both comprise 6 filters and a kernel size of 3×3 pixels. Each convolution layer is followed by a rectified linear unit (ReLU) operation. This conception is adopted from a VGG-16 network architecture, so-called stack convolution, which can achieve barely the same result with fewer parameters and computations than a larger convolution kernel. Besides, the convolution operation uses the same-padding option, which expands the boundary pixels before the convolution operation to remain the same size as the input tensor.

2. The stacked convolution layers are followed by a batch-normalization operation, which speedups the convergence and prevents the problem of gradient vanishing, and a max-pooling layer with a kernel size of 3×3 pixels and a stride of 3.

3. The second stacked convolution layer and batch-normalization layer use the same manner as the first one, except that the convolution kernel is set to 16 filters. The batch-

normalization layer is followed by a max-pooling layer with a kernel size of 4×4 pixels and a stride of 4.

4. The first full connection layer comprises 64 neurons, followed by a ReLU and a dropout operation. The dropout operation is proposed to eliminate overfitting as it trains only some randomly active neurons. The rate of the dropout was set to 0.1.

5. The second full connection layer has three neurons, which represent two classes of images in the rice seedling dataset, followed by ReLU operation. The output layer is a softmax activation function by forcing the sum of the output values equal to 1.0. This activation function also limits each output value between 0–1, which means the probability of each class.

3.3. Performance Evaluation

The evaluation metrics were adopted in this study to evaluate the classification model as the following description in detail [46].

3.3.1. Precision

Precision is the ratio of the correct classification to the total number of classifications in the specific class. A low precision indicates a large number of false positives. Precision can be represented as:

$$Precision_c = \frac{TP_c}{TP_c + FP_c}, \quad (1)$$

where TP_c depicts the positive class correctly classified by the model, and FP_c depicts the model misclassifies the samples as the positive class.

3.3.2. Recall

The recall is the ratio of the number of the correct classifications to the total number of samples. A high recall indicates a small number of misclassified samples. Recall can be represented as:

$$Recall_c = \frac{TP_c}{TP_c + FN_c}, \quad (2)$$

where FN_c depicts the model misclassified the samples as the negative class.

3.3.3. Accuracy

Accuracy is the fraction of the correctness of the model, and is calculated as the sum of correct classification divided by all the classifications as:

$$Accuracy = \frac{TP + TN}{TP + TB + FP + FN}, \quad (3)$$

where TN_c depicts that the model correctly classifies the samples as the negative class.

3.3.4. F1-score

F1-score quantifies the harmonic mean between precision and accuracy. This metric usually presents the robustness of the classifying task, and can be calculated as:

$$F1 - score_c = 2 \times \frac{Precision_c \times Recall_c}{Precision_c + Recall_c}. \quad (4)$$

3.4. Model Training

This study adopted python programming language to implement the preprocessing workflow and the classification. The deep learning architecture is TensorFlow version 2.2 [47], and the used libraries are skimage, matplotlib, and numpy.

In the beginning of training, a Gaussian distribution was applied to initialize the weights of the layers randomly. To train the network, an adaptive moment estimation (Adam) optimizer [48] was adopted with an initial learning rate of $5E-5$, batch size of 128, and the number of epochs of 20. To avoid the possible bias in the particular division of the training dataset, k-fold cross-validation was introduced. k was set to 5 to obtain an 80/20 dividing ratio of training/test data. The accuracy of the network increases with the number of epoch, and all the divisions of the cross-validation dataset achieve a 0.99 accuracy and close to 1.0 (Figure 9). All the divisions of cross-validation datasets show a steady increase in validation accuracy and a steady descend in the loss. In Figure 9, the model has good performance without overfitting. To choose the best model from the five models, this paper compared the validation-accuracy of each model, which are all above 99.9%, and the one with the lowest validation-loss (the fifth model) was chosen for the evaluation and demonstration of patch-based rice seedling detection.

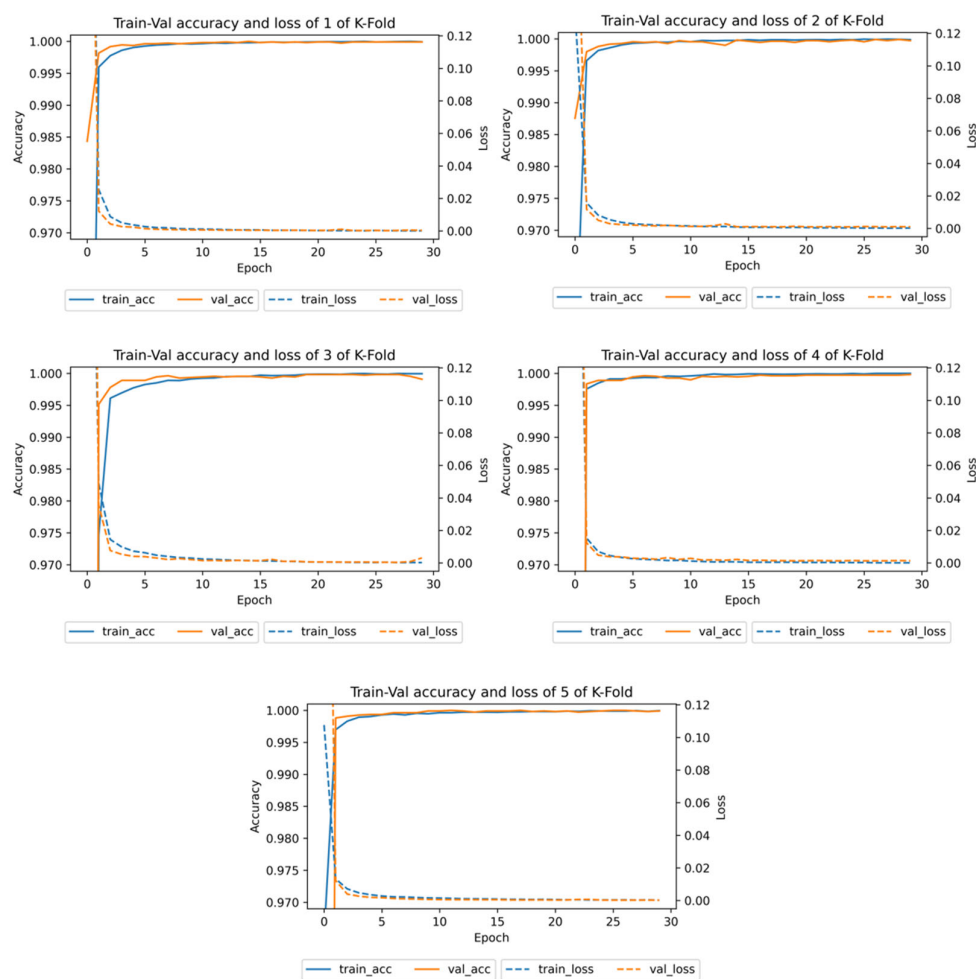


Figure 9. Accuracy and loss of five divisions of cross-validation dataset of the proposed network during training and validation.

3.5. Model Evaluation and Detection Demonstration Results

Five divisions of the test dataset were tested with the evaluation metrics using the second model, which was discussed in Sections 3.3 and 3.4, as shown in Table 7. The results indicate the model possessing a superior classification ability on all five test datasets with all F1-scores of every class reaching 99.9%.

Table 7. Model evaluation on five divisions of cross-validation datasets.

Fold	Rice Seedling			Arable land			Accuracy
	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	
1	99.98	100.00	99.99	100.00	99.98	99.99	99.99
2	99.98	99.98	99.98	99.98	99.98	99.98	99.98
3	99.98	99.98	99.98	99.98	99.98	99.98	99.98
4	99.98	99.95	99.96	99.94	99.98	99.94	99.96
5	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Figure 10 shows the post-process of the patch-based rice seedling detection in the Subset 7 demo image shown in Figure 7. The rice seedling detection consists of an overlapped patch-based image detection and a post-process of heatmaps. Images for detection are subset into many patches with an overlap (also called sliding window) to form a long sequence of image sets with a size of 48×48 pixels that are applied to the proposed classification model to output the probability of each pixel in each class.

The classification results were reordered to form a heatmap (Figure 10a) in which the size is identical to the original image. Then a threshold of 0.99 as the confidence of the classification (Figure 10b). An erosion operation with a diamond-shaped filter was applied to simply disconnect the slightly adjacent objects (Figure 10c). Finally, the findContours() function from OpenCV was applied to extract objects, and called the boundingRect() function to get the top-left position of objects and the width and height of the bounding boxes. To visualize the bounding boxes, the boxes were drawn with yellow and the width of 2 pixels on the raw image (Figure 10d).

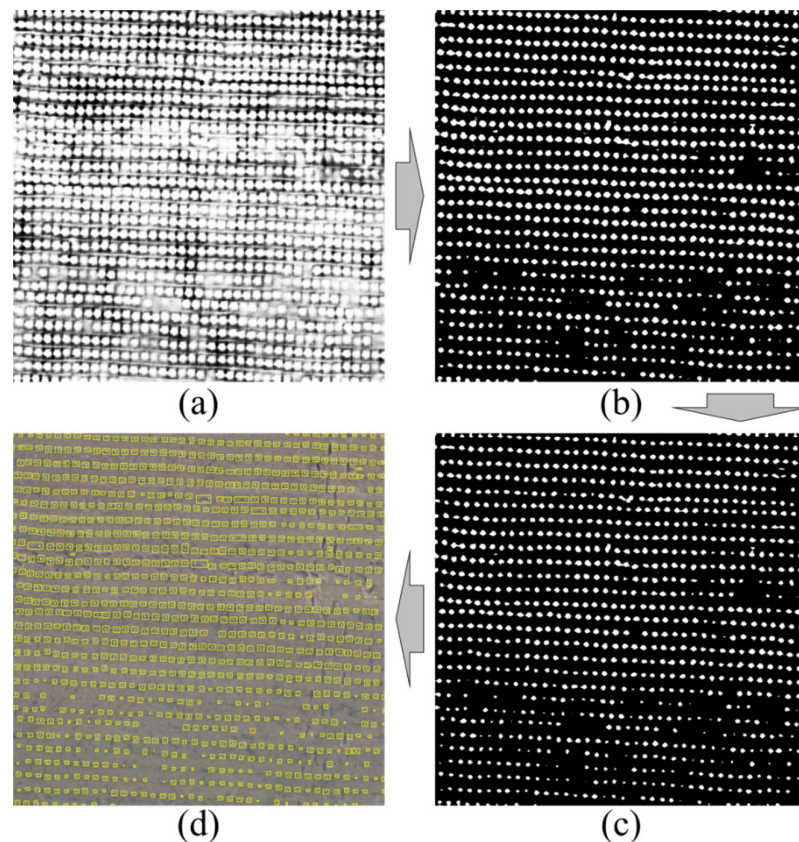


Figure 10. The post-process of patch-based rice seedling detection in Subset 7. (a) heatmap of classification results; (b) result image obtained from a binarization process with a 0.99 threshold;

(c) result image obtained from a diamond-shaped erosion process; (d) RGB image with drawn bounding boxes of detected objects for visualization.

The comparison of the prediction image and ground truth image of Subset 1 is presented in Figure 11, in which the detected seedlings were drawn with the yellow bounding boxes. Due to the limited layout, the remaining images can be accessed from the web. Table 8. Comparison of the hill number of rice seedlings from patch-based detection and the ground truth. Amongst, Subset 1 and Subset 4 got an above 10% error rate in the number of the detected rice seedlings. To explore this issue, this paper focused on the highly undetected areas in these two images. The comparison between prediction images and ground truth images is shown in Figure 12. The undetected rice seedlings are visually smaller than the detected rice seedlings. This paper also provides images for two consecutive growing stages after 7th August. The comparison between success and failure shows that the undetected rice seedlings are generally smaller than the detected rice seedlings.

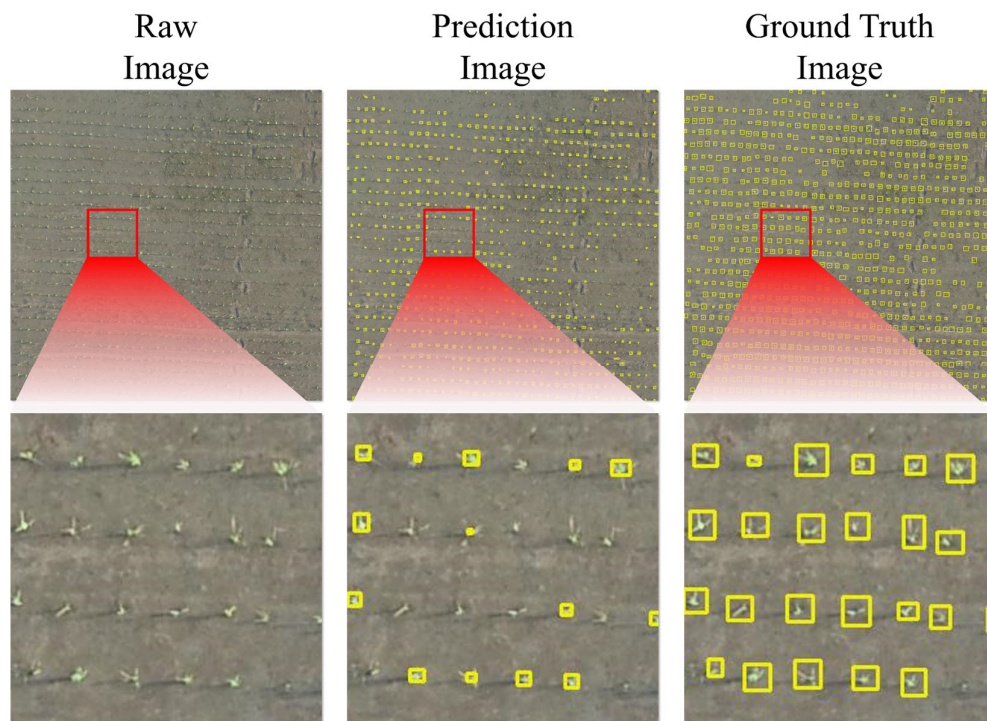


Figure 11. Comparison of the prediction images and ground truth images of Subset 1 in the detection demonstration.

Table 8. Comparison of the hill number of rice seedlings from patch-based detection and the ground truth.

Subset No.	1	2	3	4	5	6	7	8
Prediction	735	1006	1037	809	1004	1050	1017	1032
Ground truth	898	1000	1019	964	971	1002	1033	1005
Error (%)	18.15	0.60	1.77	16.08	3.40	4.79	1.55	2.69

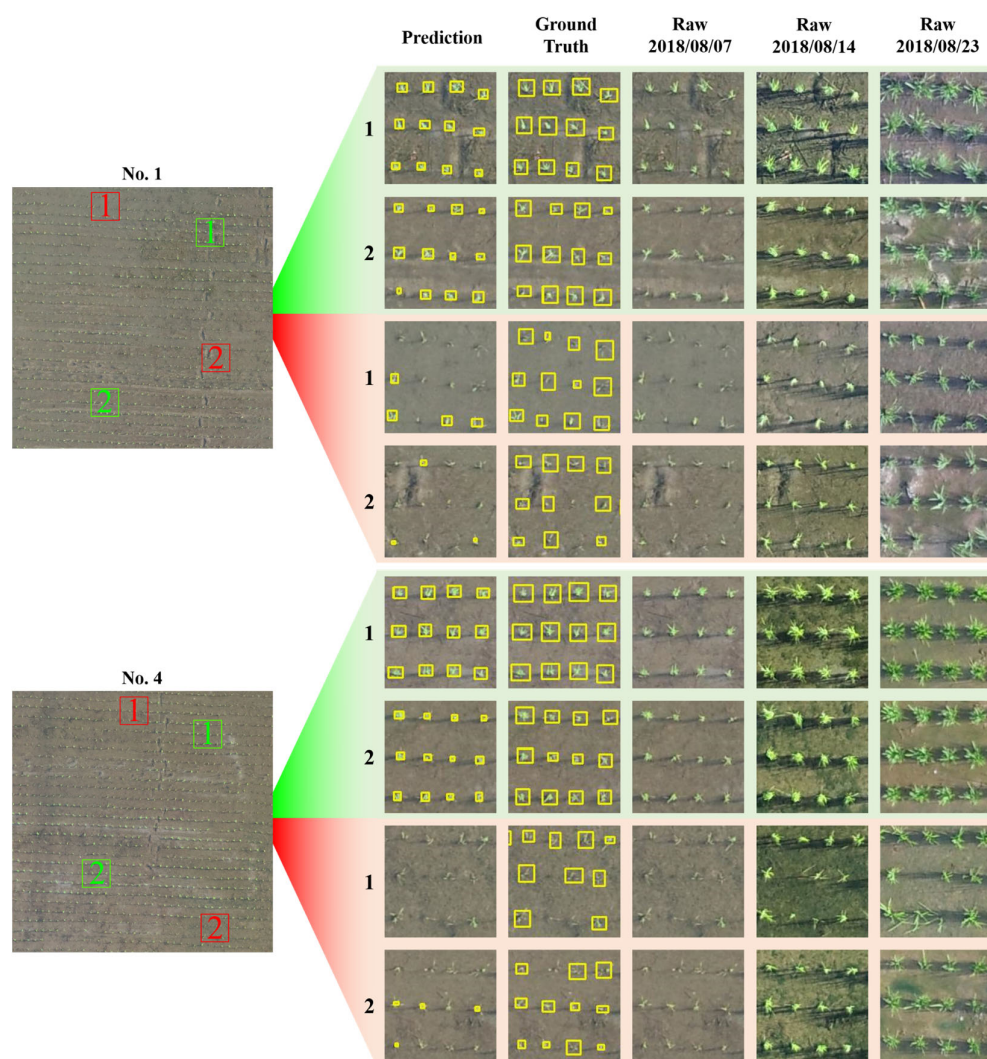


Figure 12. A focused comparison of the detection results between prediction and ground truth images for four subareas in Subset 1 and Subset 4. The images of three successive growth stages explain the difference of growing condition. The green areas depict the successful detection, and the red areas depict the worse detection.

4. Conclusions

This paper provides a verified semi-auto annotated UAV dataset for rice seedlings identification. The dataset is described in the information of acquisition, preprocessing, and usage in a CNN classification model. A classification model as an example was provided for demonstration of model training and prediction using the dataset. Also, this paper demonstrated the patch-based rice seedlings detection to show the ability for object detection and plant counting. The results and performance evaluation confirm the applicability of this dataset on rice seedling counting. For data sharing, all datasets including the training-validation dataset, patch-based detection samples, and the orthomosaic image of the field are available on the open link. <https://github.com/aipal-nchu/RiceSeedlingDataset>

Author Contributions: conceptualization, M.-D.Y. and H.-H.T.; methodology, M.-D.Y. and H.-H.T.; software, H.-H.T., and Y.-C.H.; validation, H.-H.T., and Y.-C.H.; investigation, M.-H.L., and D.-H.W.; formal analysis, M.-D.Y., H.-H.T., Y.-C.H., M.-H.L., and D.-H.W.; writing—original draft preparation, M.-D.Y. and H.-H.T.; writing—review and editing, M.-D.Y. and H.-H.T.; visualization, Y.-C.H. and H.-H.T.; supervision and project administration, M.-D.Y. and C.-Y.Y.;

funding acquisition, M.-D.Y. and C.-Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Ministry of Science and Technology, Taiwan, under Grant Number 108-2634-F-005-003.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://github.com/aipal-nchu/RiceSeedlingDataset>.

Acknowledgments: This research is supported through Pervasive AI Research (PAIR) Labs, Taiwan, and “Innovation and Development Center of Sustainable Agriculture” from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

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

Appendix A

Table A1. An example of XML file of detection annotation.

```

<annotation>
  <folder>raw</folder>
    <filename>1.tif</filename>
    <path>data/demo/raw/1.tif</path>
    <source>
      <database>RiceSeedlingDetection</database>
    </source>
    <size>
      <width>320</width>
      <height>320</height>
      <depth>3</depth>
    </size>
    <segmented>0</segmented>
    <object>
      <name>RiceSeedling</name>
      <pose>Unspecified</pose>
      <truncated>0</truncated>
      <difficult>0</difficult>
      <bndbox>
        <xmin>159</xmin>
        <ymin>283</ymin>
        <xmax>181</xmax>
        <ymax>305</ymax>
      </bndbox>
    </object>
    <object>
      ...
  </annotation>

```

References

1. Brown, M.E.; Funk, C.C. Food security under climate change. *Science* **2008**, *319*, 580–581.
2. Pison, G. The population of the world. *Popul. Soc.* **2019**, *569*, 1–8.
3. Yang, M.D.; Yang, Y.F.; Hsu, S.C. Application of remotely sensed data to the assessment of terrain factors affecting the Tsao-Ling landslide. *Can. J. Remote Sens.* **2004**, *30*, 593–603.
4. Yang, M.D.; Lin, J.Y.; Yao, C.Y.; Chen, J.Y.; Su, T.C.; Jan, C.D. Landslide-induced levee failure by high concentrated sediment flow—A case of Shan-An levee at Chenyulan River, Taiwan. *Eng. Geol.* **2011**, *123*, 91–99.
5. Yang, M.D.; Su, T.C.; Hsu, C.H.; Chang, K.C.; Wu, A.M. Mapping of the 26 December 2004 tsunami disaster by using FORMOSAT-2 images. *Int. J. Remote Sens.* **2007**, *28*, 3071–3091.
6. Lin, J.Y.; Yang, M.D.; Lin, B.R.; Lin, P.S. Risk assessment of debris flows in Songhe Stream, Taiwan. *Eng. Geol.* **2011**, *123*, 100–112.

7. Xiao, X.; Boles, S.; Frolking, S.; Li, C.; Babu, J.Y.; Salas, W.; Moore III, B. Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. *Remote Sens Environ.* **2006**, *100*, 95–113.
8. Gebbers, R.; Adamchuk, V.I. Precision agriculture and food security. *Science* **2010**, *327*, 828–831.
9. Ozdogan, M.; Yang, Y.; Allez, G.; Cervantes, C. Remote sensing of irrigated agriculture: Opportunities and challenges. *Remote Sens.* **2010**, *2*, 2274–2304.
10. Downs, S.W. Remote Sensing in Agriculture. 1974, Huntsville, AL: NASA. Available online: <https://ntrs.nasa.gov/api/citations/19740009927/downloads/19740009927.pdf> (accessed on 4 January 2021).
11. Peña, J.M.; Gutiérrez, P.A.; Hervás-Martínez, C.; Six, J.; Plant, R.E.; López-Granados, F. Object-Based Image Classification of Summer Crops with Machine Learning Methods. *Remote Sens.* **2014**, *6*, 5019–5041.
12. Becker-Reshef, I.; Justice, C.; Sullivan, M.; Vermote, E.; Tucker, C.; Anyamba, A.; Small, J.; Pak, E.; Masuoka, E.; Schmaltz, J.; et al. Monitoring global croplands with coarse resolution earth observations: The Global Agriculture Monitoring (GLAM) project. *Remote Sens.* **2010**, *2*, 1589–1609.
13. Atzberger, C. Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sens.* **2013**, *5*, 949–981.
14. Sanders, K.T.; Masri, S.F. The energy-water agriculture nexus: The past, present and future of holistic resource management via remote sensing technologies. *J. Clean. Prod.* **2016**, *117*, 73–88.
15. Landsat Missions—Landsat 8. Available online: <https://www.usgs.gov/core-science-systems/nli/landsat/landsat-8> (accessed on 6 January 2021).
16. SPOT7—Earth Online. Available online: <https://earth.esa.int/eogateway/missions/spot-7> (accessed on 6 January 2021).
17. Sentinel-2—Missions—Resolution and Swath—Sentinel Handbook. Available online: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/instrument-payload/resolution-and-swath> (accessed on 6 January 2021).
18. RADARSAT Constellation. Available online: <https://earth.esa.int/web/eoportal/satellite-missions/r/rcm> (accessed on 6 January 2021).
19. Zhang, C.; Kovacs, J.M. The application of small unmanned aerial systems for precision agriculture: A review. *Precis. Agric.* **2012**, *13*, 693–712.
20. Kwak, G.; Park, N. Impact of Texture Information on Crop Classification with Machine Learning and UAV Images. *Appl. Sci.* **2019**, *9*, 643.
21. Yang, M.D.; Huang, K.S.; Kuo, Y.H.; Tsai, H.P.; Lin, L.M. Spatial and Spectral Hybrid Image Classification for Rice Lodging Assessment through UAV Imagery. *Remote Sens.* **2017**, *9*, 583.
22. Yang, M.D.; Boubin, J.G.; Tsai, H.P.; Tseng, H.H.; Hsu, Y.C.; Stewart, C.C. Adaptive autonomous UAV scouting for rice lodging assessment using edge computing with deep learning EDANet. *Comput. Electron. Agric.* **2020**, *179*, 105817.
23. Yang, C.Y.; Yang, M.D.; Tseng, W.C.; Hsu, Y.C.; Li, G.S.; Lai, M.H.; Wu, D.H.; Lu, H.Y. Assessment of Rice Developmental Stage Using Time Series UAV Imagery for Variable Irrigation Management. *Sensors* **2020**, *20*, 5354.
24. Egli, S.; Höpke, M. CNN-Based Tree Species Classification Using High Resolution RGB Image Data from Automated UAV Observations. *Remote Sens.* **2020**, *12*, 3892.
25. Chen, Y.; Lee, W.S.; Gan, H.; Peres, N.; Fraisse, C.; Zhang, Y.; He, Y. Strawberry Yield Prediction Based on a Deep Neural Network Using High-Resolution Aerial Orthoimages. *Remote Sens.* **2019**, *11*, 1584.
26. Yang, M.D.; Tseng, H.H.; Hsu, Y.C.; Tsai, H.P. Semantic Segmentation Using Deep Learning with Vegetation Indices for Rice Lodging Identification in Multi-date UAV Visible Images. *Remote Sens.* **2020**, *12*, 633.
27. Li, Z.; Li, Y.; Yang, Y.; Guo, R.; Yang, J.; Yue, J.; Wang, Y. A high-precision detection method of hydroponic lettuce seedlings status based on improved Faster RCNN. *Comput. Electron. Agric.* **2021**, *182*, 106054.
28. Pearse, G.D.; Tan, A.Y.S.; Watt, M.S.; Franz, M.O.; Dash, J.P. Detecting and mapping tree seedlings in UAV imagery using convolutional neural networks and field-verified data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 156–169.
29. Oh, S.; Chang, A.; Ashapure, A.; Jung, J.; Dube, N.; Maeda, M.; Gonzalez, D.; Landivar, J. Plant Counting of Cotton from UAS Imagery Using Deep Learning-Based Object Detection Framework. *Remote Sens.* **2020**, *12*, 2981.
30. CIFAR-10 and CIFAR-100 Datasets. Available online: <https://www.cs.toronto.edu/~kriz/cifar.html> (accessed on 6 January 2021).
31. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20–25 June 2009; pp. 248–255.
32. Lin, Y.T.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. In Proceedings of the 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
33. FAOSTAT—New Food Balances. Available online: <http://www.fao.org/faostat/en/#data/FBS> (accessed on 4 January 2021).
34. Muthayya, S.; Sugimoto, J.D.; Montgomery, S.; Maberly, G.F. An overview of global rice production, supply, trade, and consumption. *Ann. N. Y. Acad. Sci.* **2014**, *1324*, 7–14.
35. Phantom 4 Pro—DJI. Available online: <https://www.dji.com/phantom-4-pro?site=brandsite&from=nav> (accessed on 6 January 2021).
36. Zenmuse X7 Specs—DJI. Available online: <https://www.dji.com/zenmuse-x7/info#specs> (accessed on 12 March 2021).
37. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60.

-
38. AgiSoft Metashape Professional 1.6.5. (Software); AgiSoft LCC: St. Petersburg, Russia, 2020.
 39. Meyer, G.E.; Neto, J.C. Verification of Color Vegetation Indices for Automated Crop Imaging Applications. *Comput. Electron. Agric.* **2008**, *63*, 282–293.
 40. Yen, J.C.; Chang, F.J.; Chang, S. A new criterion for automatic multilevel thresholding. *IEEE Trans. Image Process.* **1995**, *4*, 370–378.
 41. Contour Features—Open Source Computer Vision (OpenCV). Available online: https://docs.opencv.org/4.5.1/dd/d49/tutorial_py_contour_features.html (accessed on 6 January 2021).
 42. Lever, J.; Krzywinski, M.; Altman, N. Model selection and overfitting. *Nat. Methods* **2016**, *13*, 703–704.
 43. The PASCAL Visual Object Classes Homepage. Available online: <http://host.robots.ox.ac.uk/pascal/VOC/> (accessed on 6 January 2021).
 44. Labellmg. Available online: <https://github.com/tzutalin/labellmg> (accessed on 7 January 2021).
 45. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
 46. Sokolova, M.; Japkowicz, N.; Szpakowicz, S. Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In Proceedings of the 19th Australian Joint Conference on Artificial Intelligence (AI), Hobart, Australia, 4–8 December 2006; pp. 1015–1021.
 47. TensorFlow. Available online: <https://www.tensorflow.org/> (accessed on 7 January 2021).
 48. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.