Development of phenotyping system using low altitude UAV imagery and deep learning

Suxing Lyu[1,2], Noboru Noguchi[3*], Ricardo Ospina[3], Yuji Kishima[3]

(1. *Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa 277-8561, Japan*;
2. *Graduate School of Agriculture, Hokkaido University, Sapporo 060-8589, Japan*;
3. *Research Faculty of Agriculture, Hokkaido University, Sapporo 060-8589, Japan*)

Abstract: In this study, a lightweight phenotyping system that combined the advantages of both deep learning-based panicle detection and the photogrammetry based on light consumer-level UAVs was proposed. A two-year experiment was conducted to perform data collection and accuracy validation. A deep learning model, named Mask Region-based Convolutional Neural Network (Mask R-CNN), was trained to detect panicles in complex scenes of paddy fields. A total of 13 857 images were fed into Mask R-CNN, with 80% used for training and 20% used for validation. Scores, precision, recall, Average Precision (AP), and F1-score of the Mask R-CNN, were 82.46%, 80.60%, 79.46%, and 79.66%, respectively. A complete workflow was proposed to preprocess flight trajectories and remove repeated detection and noises. Eventually, the evident changed in rice growth during the heading stage was visualized with geographic distributions, and the total number of panicles was predicted before harvest. The average error of the predicted amounts of panicles was 33.98%. Experimental results showed the feasibility of using the developed system as the high-throughput phenotyping approach.

Keywords: panicle detection, vision-based phenotyping, deep learning, unmanned aerial vehicle (UAV)
DOI: 10.25165/j.ijabe.20211401.6025

Citation: Lyu S X, Noguchi N, Ospina R, Kishima Y. Development of phenotyping system using low altitude UAV imagery and deep learning. Int J Agric & Biol Eng, 2021; 14(1): 207–215.

1 Introduction

Rice (*Oryza sativa* L.) is the most important grain crop in Asia, and it provides up to 50% of the dietary caloric supply[1]. Its huge market and heavy demand appeal to agricultural researchers and ways to improve the production, yield, and quality of rice are widely studied[2,3]. In agriculture, rice is often researched through visual inspection because during growth stages rice can reveal significant changes in genetic traits, which are visibly observable. Conventional research heavily relied on manual labor-intensive and time-consuming checks. Recently, as computer vision technology has been rapidly developed, applications such as image classification, image segmentation, and object detection, have been highly developed and implemented in the real world[4]. Such computer-vision-based developments have also received widespread attention from agricultural researchers due to their superiority for noninvasive data collection and quantitative studies[5,6]. Meanwhile, computer-vision-based phenotyping has been recognized as a high-throughput approach to identify genetic traits and increase breeding efficiency[7]. For bringing these methods into rice phenotyping analysis, the first step is to choose an appropriate research object, which can not only visibly reflect key changes of rice but also should be meaningful for phenotyping. In this study, the panicle of rice is treated as a major object, as panicle has been identified as a critical trait in which it is regarded as a remarkable genetic parameter[8], and it can also denote a significant transformation between the vegetative stage and the reproductive stage[9]. Given the panicle as the major object, previous computer-vision-based rice phenotyping mainly focuses on segmentation[10,11], detection[12–14] and modeling[15]. Specifically, Ikeda et al.[15] utilized the image-based software, Panicle Structure Analyzer for Rice (PASTAR) and PASTAR Viewer to construct a model of panicle structure automatically. Guo et al.[12] monitored the diurnal peak of flowering through a series of images collected from the natural paddy field. In 2019, the same team updated this research by replacing convolutional neural networks as the backbone for panicle detection[13]. Duan et al.[10] designed a color feature-based artificial neural network (ANN) for multi-angle image segmentation. Lately, Xiong et al.[11] showed that their convolutional-neural-network (CNN)-based deep learning model was the most effective model for identifying panicle segmentation and for reliability across different cases. In addition, because of the similarity between rice and wheat, it is also worth reviewing studies on the wheat panicle. Zhu et al.[16] proposed a coarse-to-fine method for wheat ear detection. Sadeghi-Tehran et al.[17] proposed a similar hierarchical approach to extract and abstract the features of low-to-high levels features step-by-step and to enhance the color contrast by decorrelation stretching. In conclusion, machine-learning and deep-learning-based algorithms of computer vision have entered the domain in recent years due to their remarkable performance[7,18]. Nevertheless, given the notoriously heterogeneous field conditions, it is hard to extrapolate from the results of the phenotypic analysis performed in controlled environments in laboratory and indoor fields to natural conditions[19].

Unmanned aerial vehicles (UAVs) have been recognized as one

Received date: 2020-07-15 **Accepted date:** 2020-12-23

Biographies: Suxing Lyu, PhD candidate, research interest: UAV remote sensing and computer vision, Email: johnlyu94@outlook.com; **Ricardo Ospina**, PhD, JSAM Member, Postdoctoral fellow, research interest: agricultural robots, vehicle robotics. Email: rospina@bpe.agr.hokudai.ac.jp; **Yuji Kishima,** PhD, Professor, research interest: genetic diversity, plant breeding, Email: kishima@abs.agr.hokudai.ac.jp.

***Corresponding author: Noboru Noguchi**, PhD, JSAM Member, Professor, research interest: agricultural robots, vehicle robotics. Research Faculty of Agriculture, Hokkaido University, Kita-9, Nishi-9, Kita-Ku, Sapporo, Hokkaido, 060-8589, Japan. Tel: +81-11-706-3847, Email: noguchi@ bpe.agr.hokudai.ac.jp.

of the effective field-based high-throughput phenotyping platforms due to the boost in UAV manufacturing and to their flexibility and the high quality of their data collection[20]. For utilizing computer-vision-based methods more flexibly, the UAV-based platform works with visual detection and segmentation is under rising trend. Ampatzidis and Partel.[21] designed a low-cost and automated system of phenotyping by UAV to detect, count, and locate trees for further detailed individual tree evaluation. Gnädinger and Schmidhalter[22] proposed a method to assess the emergence of maize by counting the number of maize plants using UAV aerial images. Jin et al.[23] estimated the density of wheat crops at emergence through segmenting corps from very low altitude UAV images. Zhou et al.[14] implemented a UAV-platform for rice panicle detection and compared the performance using different CNNs. In conclusion, aerial images from UAVs at low altitudes have been shown to have extensive applications and to be useful for high-throughput phenotyping, especially on segmentation and the detection of individual crops. Given the state-of-the-art computer vision technologies, the combination of deep-learning-based detection and aerial photogrammetry at low altitudes is promising.

In this study, a vision-based panicle phenotyping system was developed by combining a cutting-edge deep learning model, Mask region-based CNN (Mask-R-CNN; He et al.[24]), with consumer-grade UVAs. The aim of this system is to construct a complete workflow, including panicle detection, mapping, and prediction of the total number of panicles during the heading stage. This system is hoped to be light and easy-to-go for providing a practicable implementation for farmers. Therefore, only two consumer-grade UVAs, DJI Mavic Pro and DJI Mavic 2 Zoom (DJI Corporation, Shenzhen, China), and their embedded cameras were utilized, excluding high-load six-rotor UAVs with heavy and expensive cameras. Two-year experiments were conducted to evaluate the performance. The main difficulties in achieving the goal are summarized as follows:

◆ A strict standard of the training dataset is required. The dataset should consider not only the size but also the diversities of panicles. The diversities include different cultivation densities, varieties, natural conditions, and changes in crop growth. The model trained on this dataset should be generalized enough.

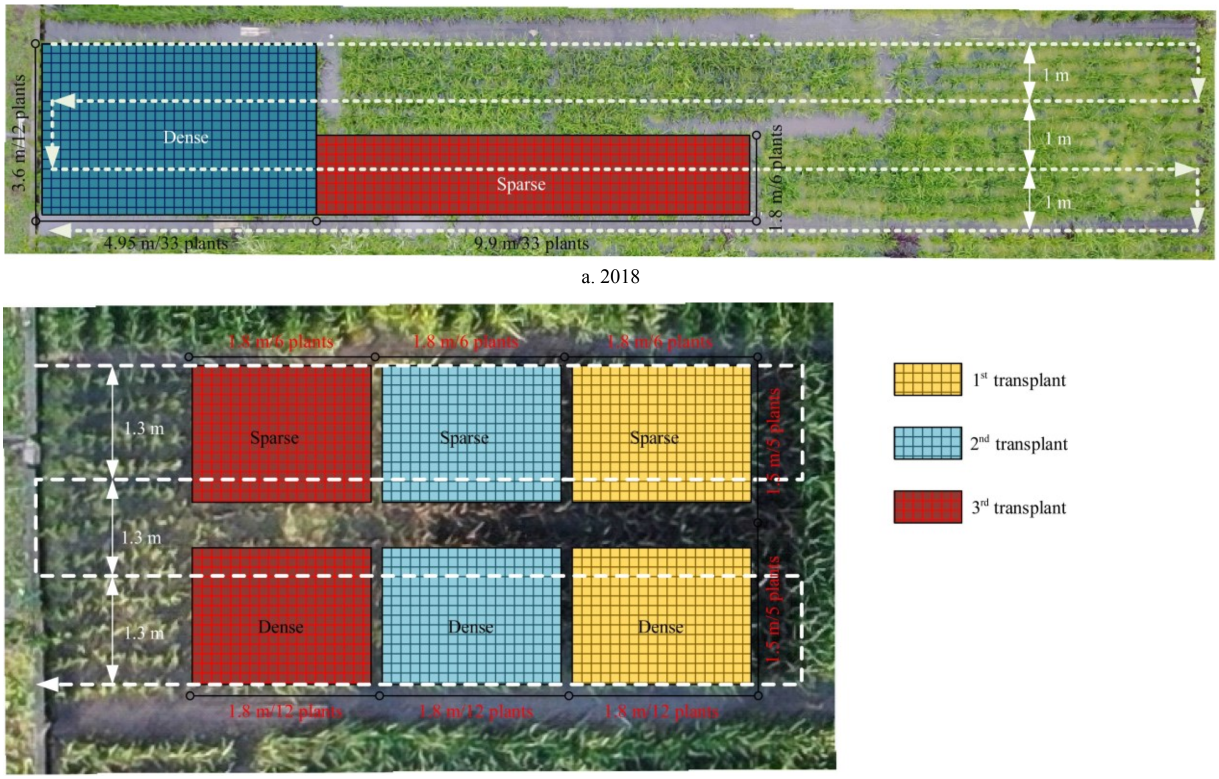
◆ Due to the limited ability of the utilized UAVs and their embedded cameras, flights must be very close to the ground. The images acquired at proximal altitudes contain numerous homogeneous scenes and noises caused by changing lights, the reflectance of water surface, and waggling crops moved by the wind from nature or rotors. In this situation, stitching images as an orthographic image is hard, because feature points among images cannot be matched correctly. From numerous images with overlays, eliminating noises and extracting effective information are challenges.

There are mainly three contributions to this study. (1) A large rice panicle dataset with pixel-level annotations was built from scratch; (2) The trained model was implemented and tested on different varieties, cultivation densities and UAVs to confirm the generalizability of the model; (3) A complete workflow, including extracting panicle locations and eliminating noises, was proposed for predicting the total number of panicles before harvest without stitching orthographic images.

2 Materials and methods

2.1 Study area

The study was held at the paddy fields of the farm at Hokkaido University, Sapporo, Japan. The experiments in 2018 and 2019 were conducted at different paddy fields on the farm, as shown in Figure 1.



a. 2018

b. 2019

Figure 1 Experimental paddy fields in 2018 (a) and in 2019 (b)

In 2018, rice was sown on 6 May and was transplanted to the field on 7 June. There were two types of cultivation. (1) The dense cultivation (covered by blue grids) involved the planting of rice at 15-cm vertical intervals and 30-cm horizontal intervals. A total of 396 Kitaake plants (an improved rice variety from Hokkaido) were arranged in 33 rows and 12 columns. (2) The sparse cultivation (covered by red lines) involved the planting of individuals at 30-cm vertical intervals and 30-cm horizontal intervals. A total of 198 plants were arranged in 33 rows and 6 columns. Specifically, the plants under sparse cultivation were crossed by Kitaake and Kokusyokuto-2 (a Hokkaido landrace rice generally denoted as A58).

In 2019, besides two types of cultivation densities, three growth stages of plants were planned. The variety was Nanatsuboshi (an improved rice variety from Hokkaido). Rice was sown on 9 May, 24 May, and 5 June, and transplanted to the field on 4 June, 18 June, and 8 July, respectively. Half of all areas were planted with the dense cultivation, with the plants arranged in 12 rows and 5 columns with 15-cm vertical intervals and 30-cm horizontal intervals. The other half of the areas were planted with sparse cultivation, with the plants arranged in 6 rows and 5 columns with 30-cm vertical intervals and 30-cm horizontal intervals. A total of six areas were defined on the basis of their transplanting dates and cultivation densities.

2.2 Hardware settings

The downward wind from the rotors can strongly shake stems and affect the image quality, as growing panicles have weak stems. Hence, the small-scale UAVs are suitable options, as they can fly at a very low altitude with a relatively weak wind generated by the rotors. In 2018, the DJI Mavic Pro was utilized. After acquiring data in 2018, DJI released two brand new products, the Mavic 2 Pro and Mavic 2 Zoom. Specifically, the Mavic 2 Zoom is equipped with an optical zoom function that can zoom in twice the original focal length. Therefore, in 2019, the DJI Mavic 2 Zoom replaced the DJI Mavic Pro. The specifications of the UAVs are shown in Table 1.

Table 1 Specifications of DJI Mavic Pro and DJI Mavic 2 Zoom

	DJI Mavic Pro	DJI Mavic 2 Zoom
Sensor	1/2.3” CMOS	1/2.3" CMOS
FOV (Field of View)	78.8° (26 mm)	83° (24 mm)-48° (48 mm)
Aperture	f/2.2	f/2.8 (24 mm)–f/3.8 (48 mm)
Resolution	4096×2160	3840×2160
FPS	24	30

In 2018, the flight speed was 0.3 m/s, and the altitude was 1.2 m. The gimbal angle was kept vertical to the ground. In 2019, the flight speed was 0.3 m/s. The flight speed and gimbal angle were set at 2.5 m and −90°, respectively. Camera settings were set to auto with a 4K video mode whenever the UAVs flew in the fields. One thing that should be noted that the zoom function of DJI Mavic 2 Zoom was maintained at an enlargement of 2× (FOV 48°, f/3.8) except for the flights on 20 August, 21 August, and 22 August.

2.3 Field experiments and image acquisition

In 2018, videos acquired on 4, 6, 8, 12, and 22 August were used to extract frames. A total of 400 frames were equably selected as the training and validation dataset for the deep learning model. These selected frames contained the entire period of the heading stage before maturing and were able to confirm panicles under heterogeneous nature conditions. On the one hand, even though all frames were annotated with pixel-level polygons as much as possible, labeling all objects is impossible when a single frame could contain more than 200 panicles. On the other hand, feeding a full 4K size image is far out of the graphic card memories. To avoid missing annotations and to improve the quality of the dataset, the original frames (4096×2160) were split into small tiles (128×128). Tiles containing annotations of less than 32 pixels were further filtered from the dataset for avoiding extremely small objects in the dataset. Ultimately, 13 857 images with 38 799 annotations were obtained. Eighty percent of the dataset was randomly selected as the training set, and the rest was treated as the validation set. The dataset acquired in 2018 were only used to train and validate the deep learning model.

In 2019, videos were acquired on 6, 7, 9, 10, 12, 13, 20, 21, 22, and 27 August. Frames from each video were extracted by a one-second interval, and then they would match to corresponding geographic information. These frames are shot by the DJI Mavic 2 Zoom instead of the DJI Mavic Pro, and they met new nature conditions at the different paddy fields planted a different rice variety. These differences can confirm that the frames used in 2019 are never seen by the trained deep learning model. Furthermore, the results of evaluating the system performance on these frames should be creditable.

2.4 Mask R-CNN

The Mask R-CNN[24] belongs to the R-CNN series[25] and is modified on the Faster R-CNN[26] by adding the FCN (Fully Convolutional Network)[27] for the pixel-level masking of objects. In conclusion, the Mask R-CNN, mainly consists of (1) a residual learning network (ResNet)[28] and an FPN (Feature Pyramid Network)[29] as the backbone for feature extraction over an entire image; (2) a region proposal network (RPN)[26] for proposing candidate object-bounding boxes; (3) RoIAlign for resolving the misalignment; (4) the head layer for bounding-box recognition (classification and regression); and (5) an FCN for mask prediction. We implemented the Mask R-CNN from an open-source project[30] on Github to fit the requirements in this study.

On the configuration of the training progress, only two classes, the background, and panicles were defined. The detailed configuration of the training progress included (1) the ResNet101 as a backbone; (2) a batch size of 2 on a single graphic card; (3) a learning rate of 0.001 for 30 epochs only training head layers and a learning rate of 0.0001 for 60 epochs training all layers; (4) no resizing of the input image (128×128); (5) beginning from pre-trained COCO weights instead of beginning from random weights; (6) RPN anchor scales of 8, 16, 32, 64, and 128; (7) on-the-fly augmentation. Specifically, data augmentation was conducted with either no change or by randomly selecting 1 to 3 types of (1) horizontal flip; (2) vertical flip; (3) one randomly selected affine translation from rotations of 90°, 180°, and 270°; (4) random gaussian blur with sigma values from 0 to 5; (5) random scaling from 1× to 2×.

During inference progress, a sliding window (1024×1024) with 50% overlap was applied to scan the 4K full image. In every step, the pre-trained model detects the contents within a sliding window with a detection confidence of 85%. Non-maximum suppression (NMS)[31] was implemented to delete repeated detection among the overlapping sliding windows.

2.5 Workflow of predicting the total numberof panicles

2.5.1 Flight trajectory preprocessing

The centroid status of the UAV was recorded as flight logs by a frequency of every 100 ms. Each log consists of (1) a geographic location (Latitude, Longitude, Altitude); (2) a timestamp; (3) three rotation angles (Pitch, Roll, Yaw); (4) a gimbal angle. Original geographic locations were recorded in the WGS84 coordinate (EPSG: 4326), and then they were converted to Tokyo/UTM zone 54N (EPSG:3095) coordinate, which used the Tokyo geographic 2D coordinate reference system (CRS) as its base CRS and the UTM zone 54N as its projection. Therefore, the format of recorded locations was converted to (*X*, *Y*, *Z*).

Given the sensor noises and missing logs, directly matching frames to their corresponding logs can lead to errors. Thus, the Kalman smoother[32] was utilized for preprocessing flight trajectories. The Kalman smoother is an extension based on the Kalman filter, and it is a backward algorithm. The Kalman smoother can provide imputed values for missing values in time series. To apply the Kalman smoother needs to set a dynamic model. In this study, the dynamic model was extended from a 2D model[33] and then was applied to explain the motion of flight with parameters using Equations (1)-(6):

$$x_k = \begin{pmatrix} X_k \\ Y_k \\ Z_k \\ s_k^{(X)} \\ s_k^{(Y)} \\ s_k^{(Z)} \end{pmatrix} \tag{1}$$

$$H_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{2}$$

$$\phi_{k-1} = \begin{bmatrix} 1 & 0 & 0 & \Delta_t & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta_t & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta_t \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

$$Q_k = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_s & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_s & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_s \end{bmatrix} \tag{4}$$

$$P_0 = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_s^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_s^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_s^2 \end{bmatrix} \tag{5}$$

$$R = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} \tag{6}$$

where, x_k-state at *k*, Equation (1). The elements X_k, Y_k, and Z_k are the coordinates in the *X*, *Y*, and *Z* directions at *k*, respectively, m; $s_k^{(X)}$, $s_k^{(Y)}$ and $s_k^{(Z)}$ are the velocities of *X*, *Y*, and *Z* at *k*, respectively, m/s. H_k-measurement matrix; ϕ_{k-1} -system matrix giving the state x_k from x_{k+1}; Q_k-system noise covariance matrix; P_0-the initial estimate of state error covariance; *R*-measurement noise covariance matrix; The value of σ is an estimate of GPS noise. In this study, σ was 5 m. σ_s was the average velocity of each flight trajectory. Describing motion modeling and the Kalman smoother in more detail is beyond the scope of this study. Thus, for more details, we recommend checking the mentioned research, as only the critical parameters are described here.

2.5.2 Mapping panicle distribution

Frames were extracted from videos with an interval of one second. Each frame would be matched to a corresponding log from the preprocessed flight trajectory by aligning the same timestamps. Continuously, panicles in the frames were detected by the trained model, and their locations were recorded in the pixel coordinate (*u*, *v*). Eventually, these locations of detected panicles were converted from the image coordinate (*x*, *y*, *z*) into the UTM coordinate (*X*, *Y*, *Z*). The transformation equations[34] are shown as Equations (7) and (8).

$$\begin{cases} x = u - \dfrac{\bar{u}}{2} \\ y = v - \dfrac{\bar{v}}{2} \\ z = \dfrac{-\sqrt{\dfrac{\bar{u}^2}{2} + \dfrac{\bar{v}^2}{2}}}{\tan\dfrac{FoV}{2}} \end{cases} \tag{7}$$

where, *u* and *v* are the coordinates in the pixel coordinate; $\bar{u}$ and $\bar{v}$ are the width and height of the frame, respectively; FoV is the diagonal angle of the frame; *x*, *y*, and *z* are the coordinates of the image coordinate.

$$\begin{pmatrix} X_p \\ Y_p \\ Z_p \end{pmatrix} = -R_3(\psi)R_2(\theta)R_1(\phi)Q_1(\rho_1)Q_2(\rho_2)\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} X_d \\ Y_d \\ Z_d \end{pmatrix} \tag{8}$$

where, X_p, Y_p, and Z_p are the panicles' locations in the real-world coordinate converted from *x*, *y*, and *z* in the image coordinate. X_d, Y_d, and Z_d are the recorded centroid locations of the UAV from logs. $R_1(\phi)$, $R_2(\theta)$, and $R_3(\psi)$ are the rotation matrices for the roll ϕ, pitch θ, and yaw ψ angles, respectively. ρ_1 and ρ_2 are the rotation angle and tilt angle of the gimbal, respectively. In this study, ρ_1 was 0. The results of mapping the locations of panicles and the flight trajectory into the real-world background are shown in Figure 2, where green points denote panicles and red points denote the discrete records of the flight trajectory. Furthermore, only the experimental area in 2019 is specified, and the panicles within this area are visualized as shown in Figure 3, where blue points denote the panicle locations.

2.5.3 Prediction of the amounts of panicles

As frames were extracted with overlap, panicles contain unwanted repetition and noises, such as false detections. Counting the number of panicles directly from the experimental area can only acquire incorrect values. The clustering algorithm, called Density-based Spatial Clustering of Applications with Noise (DBSCAN)[35], was implemented for eliminating the repetition and noises. The DBSCAN does not rely on a pre-defined number of clusters and is friendly for the arbitrary shape of samples. The required parameters of DBSCAN are the maximum distance between two samples and the minimum number of samples in a neighborhood. The DBSCAN outputs clusters and noises based on the input data. In this study, the panicles belonging to the same cluster were seen as the same panicle and would be further computed for the centroid of the cluster. The locations of the centroids were treated at the new locations of panicles of eliminating the repetition. The noises would be directly deleted. The parameters of the DBSCAN are set the maximum distance as 2 mm and the minimum number as 2.

Even though the DBSCAN can remove most repetition and noises, the errors and uncertainties still exist. To predict a relatively accurate number of panicles, fitting the number of panicles by time-series into a growth model is an option. Nonlinear regression models are widely used in agricultural research. Generally,

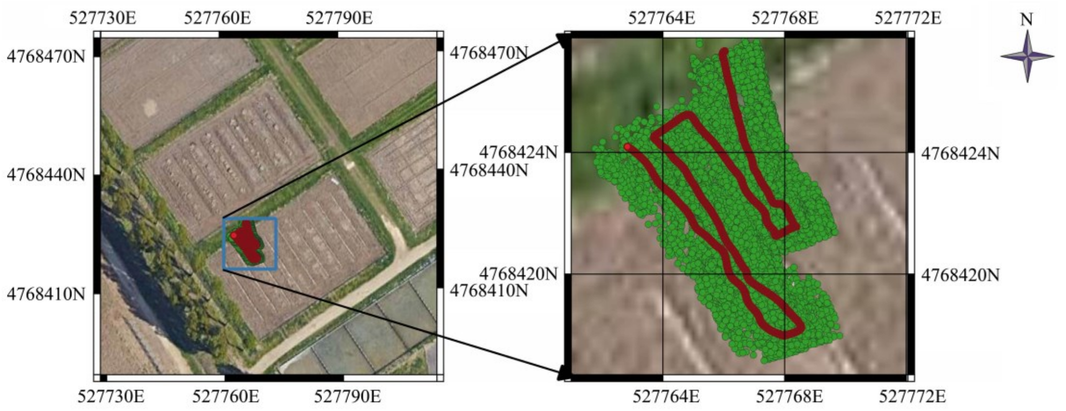


Figure 2 Mapping flight trajectory and panicle locations in the UTM coordinate

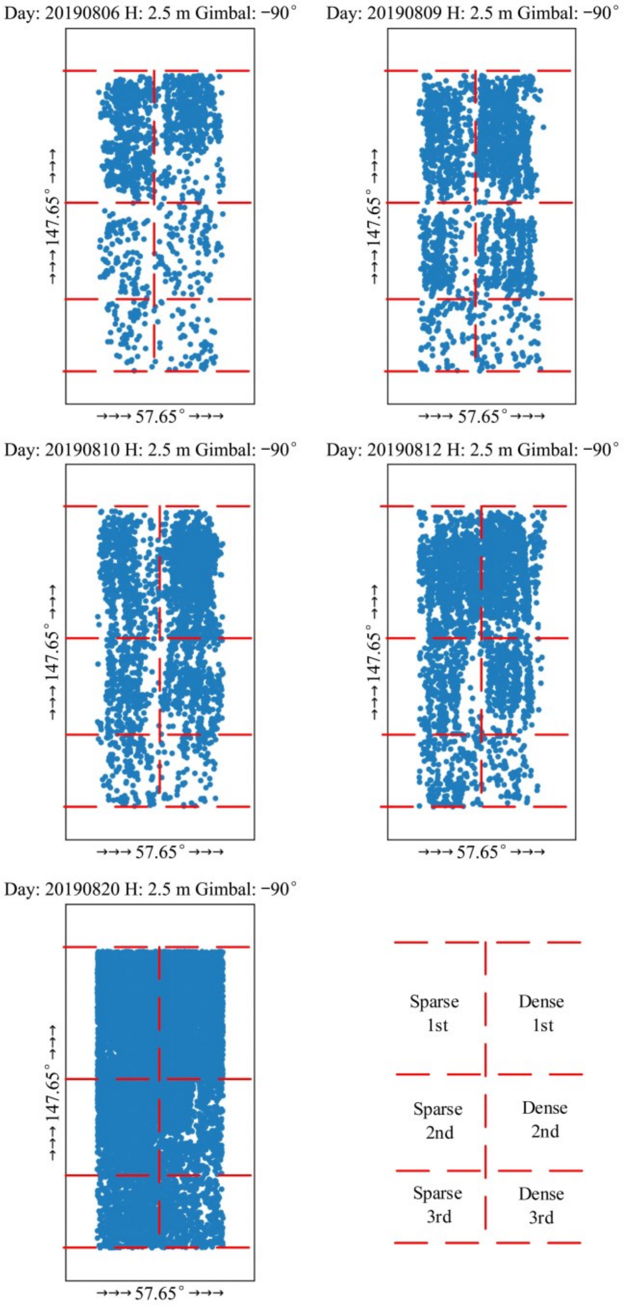


Figure 3 Panicles remapped within the experimental area in 2019

the sigmoid model also called the S-shaped model, has advantages in interpretability and prediction[36]. Meanwhile, rice growth was explored as a biphasic growth pattern[37], which was a mixture representation of two logistic curves that one was from the vegetative stage and the other was from the reproductive stage. Specifically, the total weight of panicles was acted as the dependent variable in the model during the reproductive stage. As the total weight is also tightly related to the total number of panicles, in this study, the number of panicles was set as the response variable, and the day after transplanting was set as the explanatory variable. The formulation of the growth model is shown as Equation (9).

$$Y = \frac{Y_{asym}}{1+e^{-k(t-t_m)}} + E \tag{9}$$

where, Y_{asym} is the asymptotic value; t_m is the point that the growth rate is maximum; k is the steepness of the fitted curve; E is the error. The initial guesses were given the maximum value of Y to Y_{asym}, the median date of t to t_m, 1 to k and the minimum value of Y to E. The growth model was fitted by the least-squares method.

2.6 Evaluation indices

2.6.1 Evaluation indices of Mask R-CNN

The performance of the Mask R-CNN was evaluated through three metrics under intersection over union (IoU) over 50%: Average Precision (AP), Precision, Recall, and $F1$-score. The value of IoU shows whether the overlap ratio that detected the bounding rectangle covers the corresponding annotation. Precision reflects the percentage of correct predictions. Recall measures the proportion of correct predictions among all annotations. On the precision-recall curve, AP represents the area under the curve. $F1$-score indicates the balanced accuracy of predictions. Equations (10)-(13) are shown as follows:

$$Recall = \frac{TP}{TP+FN} \tag{10}$$

$$Precision = \frac{TP}{TP+FP} \tag{11}$$

$$AP = \int_0^1 P(r)dr \tag{12}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision \times Recall} \tag{13}$$

2.6.2 Evaluation indices of growth model

R^2, Equation (14), and Root Mean Square Error (RMSE), Equation (15), are computed for evaluating how well the models are fitted. $\bar{y}$ is the mean of observed data y_i. f_i is the predicted data from the fitted model.

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})} \tag{14}$$

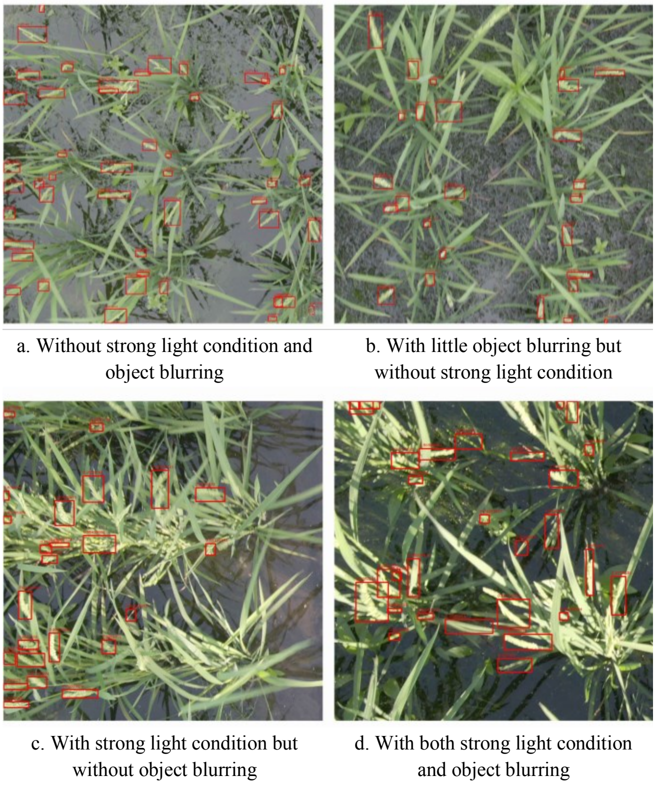
$$RMSE = \sqrt{\frac{\sum_i (y_i - f_i)^2}{n}} \tag{15}$$

3 Results and discussion

3.1 Accuracy of Mask R-CNN

The dataset acquired in 2018 was split into two parts, 80% as the training set and 20% as the validation set. The remaining 20% of the dataset was selected as the validation set with a detection coefficient of 85%. The best-trained time of the Mask R-CNN was selected, and the metrics, AP, precision, recall, and F1-score on the validation set 79.46%, 82.46%, 80.06%, and 79.66%, respectively.

Given the complex environments in natural fields, although the trained model performed well on the validation set, it is hard to say that the model is good enough and is well generalized under natural conditions. Therefore, visual inspections of several common scenes were performed by applying the best-trained model to real detection of the frames acquired in 2019. In general, strong light conditions causing overexposure and the blurring of objects are common in the paddy rice field. Four general scenes were randomly selected to check the reliability of the pre-trained Mask R-CNN, as shown in Figure 4. Figure 4a is the scene shot without strong light condition and object blurring; Figure 4b is the scene shot with little object blurring but without strong light condition; Figure 4c is the scene shot with the strong light condition but without object blurring; Figure 4d is the scene shot with both strong light condition and object blurring. In conclusion, the Mask R-CNN can handle the complicated nature conditions with acceptable overall performance.



a. Without strong light condition and object blurring

b. With little object blurring but without strong light condition

c. With strong light condition but without object blurring

d. With both strong light condition and object blurring

Figure 4 Visual inspection of the results of Mask R-CNN under four common scenes shot:

3.2 Interpretation of rice growth

The manual check would be used to compare the growth trend with mapping results, as shown in Figure 3, to interpret the growth changes represented by the changes in mapping panicle distributions. The manual check recorded the ratio of crops where the first panicle had emerged from the rice stem, as shown in Figure 5. From the perspective of the record, all crops in the paddy field came into the heading stage first in the dense cultivation areas. The sparse cultivation areas in the second and third transplant regions grew better than the dense cultivation areas. The third transplanted area could not enter the full heading status until harvest.

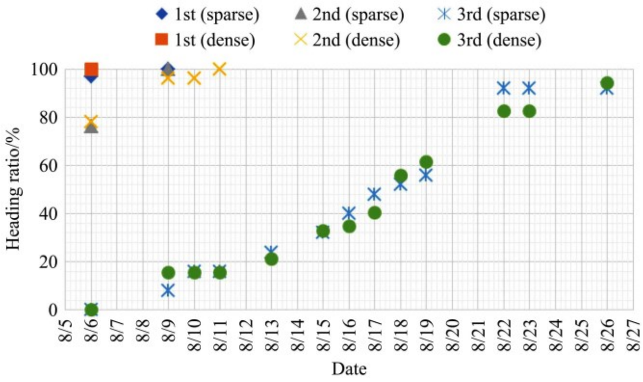


Figure 5 Records of the ratio of panicles in the heading stage

From the perspective of the panicle distributions, the sparse and dense regions of the earliest transplant showed the fastest growing speed compared with the other two transplants. Meanwhile, crops in the second transplant areas were still growing separately with little crossing. The third transplant region grew slowly because the appropriate growth window was missed. With reference to the cultivation density of the various regions, the dense cultivation of the first transplant showed the most rapid growth overall, and the sparse cultivation of the second transplant had greater panicle density than the dense cultivation within the same transplant on 20 August. In conclusion, the changes in panicle distributions generally obey the same growth trend as what is recorded by a manual check. The proposed system can reflect the differences during rice growth and should be reliable enough.

3.3 Accuracy of the prediction

In 2019, the harvest was conducted on 27 August. The final number of panicles was counted at harvest. To compare the real number of panicles with a predicted total, the growth model would predict the amounts on 27 August by the growth model. As the third transplant areas, 3S and 3D, could not enter the full heading status, these two areas would not fit the assumption of the growth model. Therefore, they were removed from the prediction. Eventually, the results are shown in Table 2 and the fitted models are shown in Figure 6, where the numbers of titles denote the sequence of transplant date (1, 2, or 3) and the capitals denote the cultivation type, D (dense) or S (sparse). The errors in Table 2 are normalized as $\frac{|Predicted\ Amount - Truth\ Amount|}{Truth\ Amount}\times 100\%$. As shown in Table 2, the model of 1D area has the best fitness and acquires the minimum error. Relatively, the 1S and 2D areas show unexpected results. The predicted amount of the 1S area is much larger than the true amount. On the contrary, the predicted amount of the 2S area is far less than the truth. The model belonging to 2D cannot even show the straightforward turning change due to its slowest growth status. The potential factors causing such unreliable results will be discussed through combing with Figure 6.

Table 2 Results of panicle amount prediction

No.	R^2	RMSE	Truth amount	Predicted amount	Error
1S	0.70	194.62	452	699	54.64%
1D	0.91	88.79	795	714	10.18%
2S	0.73	84.85	418	317	24.16%
2D	0.78	60.81	618	328	46.92%

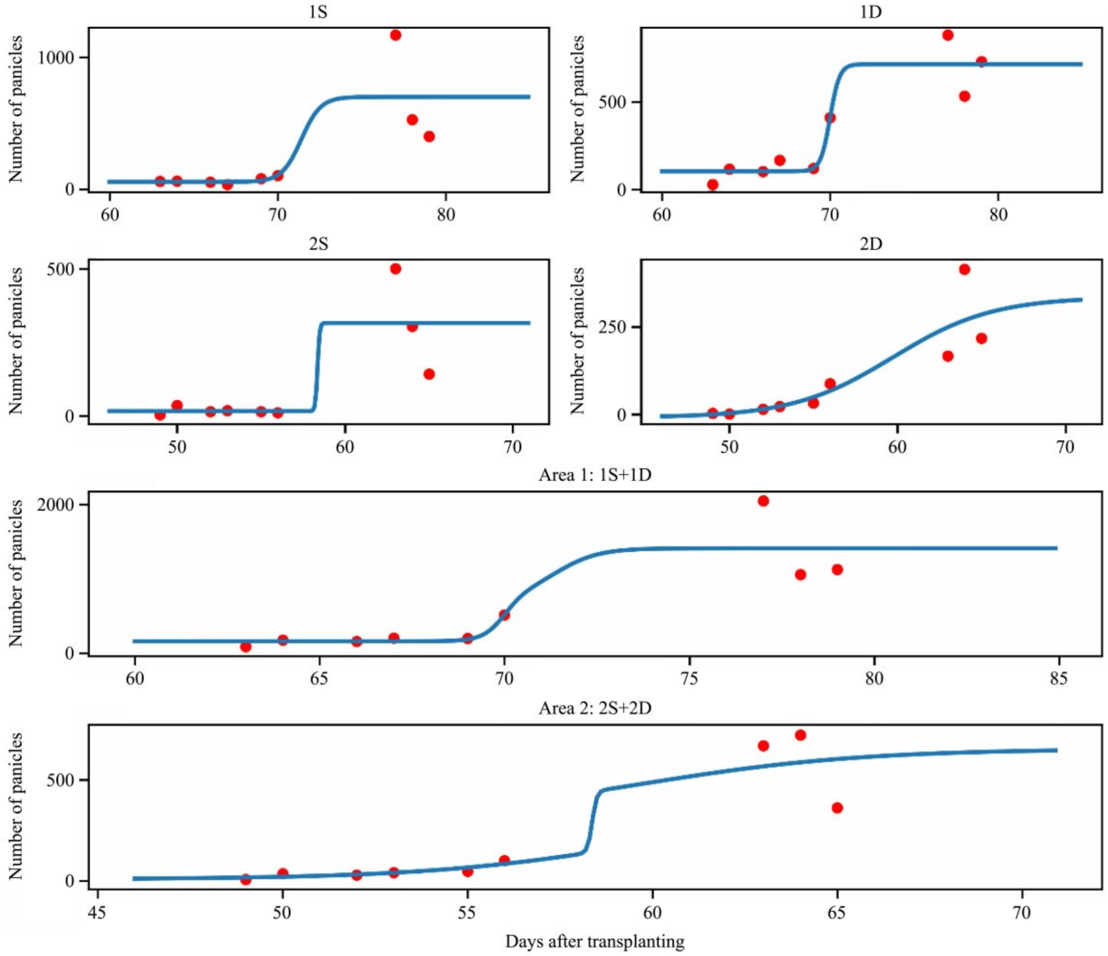


Figure 6 Panicle growth model

In conclusion, the overall R^2 is higher than 70%, the average RMSE is 107.27 and the average error is 33.98%. The reasons potentially causing errors of the prediction are concluded as follows:

◆ Due to the continuous rainfall for about one week in the middle of August, it should be mentioned that no flights were conducted in this period. This is the critical factor that there were no records during the important turning period except the 1D area. Because the 1D area was the first area that all crops in the 1D area could have at least one panicle emerging from the stem, the 1D area should be the first one to enter the fast growth stage. Thus, the result of the 1D area performed well.

◆ On the other hand, due to the loss of records during the turning period, the final predictions of other areas were highly affected by the last three records. However, within the last three records, most rice had been in the mature stage, and panicles had changed to yellow with a larger size of fruits. The changes of color and texture made huge differences of panicles between the heading stage and the mature stage. The larger size of panicles made the view of frames overcrowded. These two reasons caused the disappointing accuracy of the trained Mask R-CNN. The DBSCAN algorithm could also be affected and lead to errors in removing repetition and noises.

3.4 Errors and uncertainties

The combination of light consumer-level UAVs and state-of-the-art computer vision technology for high-throughput analysis provides a new method of field-based phenotyping that eliminates labor-intensive and time-consuming work. However, there are errors and uncertainties that can drop the overall performance of the proposed system. The key points are discussed to improve the proposed system in the future and avoid potential faults.

(1) Blurring: The accuracy of the Mask R-CNN when using an over blurred image can be significantly low. The strong wind from the rotors or natural wind can cause such heavily blurring. Improper flight speed can also cause the same result because of the relatively slow shutter speed. The flier should consider the balance between height and resolution and should try to avoid taking off when the wind speed is too great.

(2) Cover: Panicles at the lower stem are hidden under the top canopy in the image and covered by leaves during the early heading stage. Even though panicles had grown, they were covered by other panicles that densely surrounded them. Choosing an appropriate observation window need to be carefully decided.

(3) Size: On the one hand, panicles can be too small to detect during the early heading stage. Small object detection is a difficult challenge for computer vision tasks. Extremely small objects only contain little information, and scaling images can cause information loss. In this study, each full image was divided into many patches with overlap, and the patches were fed to the trained Mask R-CNN without scaling. On the other hand, most panicles are excessively large and intertwine with surrounding panicles by harvest time. Specifically, such chaotic scenes can highly affect the accuracy of the detector. Therefore, it is recommended to implement the proposed system before rice enters the mature stage.

(4) Maturity: The dataset acquired in 2018 did not include

mature panicles. When rice enters the mature stage, the features of the panicle change greatly in terms of color, texture, shape, and size. It was difficult for the trained Mask R-CNN to recognize mature panicles. Meanwhile, most crops in the paddy field had entered the mature stage after 20 August, and the trained Mask R-CNN predicted objects with relatively poor accuracy.

In addition to the factors mentioned, many uncertainties in the natural environment reflect the complex and heterogeneous condition of real fields. These uncertainties are difficult to control and should be carefully addressed in these specific scenes.

3.5 Discussion

Even though the proposed platform has several advantages, some deficiencies must be overcome. (1) Uncertainties caused by natural conditions, such as changing illumination, rainfall, and strong winds, can greatly affect the performance of the proposed platform, and so the platform is not useful in extreme weather. (2) The Mask R-CNN heavily relies on computational resources, and it cannot provide real-time processing. For solving the mentioned problems, we have contemplated mounting a neutral density (ND) filter to prevent overexposure, and accelerating computation using a cloud server in the future. Moreover, in this study, the overlaps and repeated detections were not solved perfectly. Sampling analysis, instead of continuous detection, would be an option to resolve this issue. Tracking the detected objects through continuous video frames could be another solution to resolving the filtering overlap.

4 Conclusions

Using the proposed system makes it possible to intuitively visualize the change in rice growth status during the heading stage. In particular, the growth speed and panicle distribution among different parts were clearly visualized during the middle and preliminary stages of the rice heading. Under a proper flight configuration, the Mask R-CNN performed very well in capturing complicated scenes and achieved precision, recall, AP, and F1-score of 82.46%, 80.60%, 79.46%, and 79.66% respectively. This performance was verified using another rice variety in 2019 and was sufficiently generalized. The light consumer-level UAVs, as a high-throughput phenotyping platform, provide an opportunity to implement this system not only in the limited area of a whole field but also over widespread agricultural production areas due to their excellent flexibility. UAVs can easily enter any location that cannot be observed using tractors and fixed platforms. Furthermore, the light consumer-level UAVs are relatively cheaper than high-load UAVs, and their operations are novice-friendly. The original intention of this study is to develop an easy-to-go system for those who will use it in real agricultural production.

Up to the present, agricultural production has met the revolution of which labor-intensive and time-consuming processes have been transformed by the predominance of high technology. The increase in the use of high technology promotes agricultural information data-driven methods as new methods covering all aspects of agricultural production. For the proposed system, it can see the turning point that multidisciplinary fusion leads to new development. The proposed system can be applied not only to phenotyping analysis but can also be expected to become a source of agricultural data. Given the high resolution of the generated maps, data can be treated as an accurate reflection of ground conditions or as a sampling of data that has value in other applications. The idea of using such a phenotyping system for future research is offered.

[References]

[1] Mosleh M K, Hassan Q K, Chowdhury E H. Application of remote sensors in mapping rice area and forecasting its production: A review. Sensors, 2015; 15(1): 769–791.

[2] Shamshiri R R, Ibrahim B, Balasundram S K, Taheri S, Weltzien C. Evaluating system of rice intensification using a modified transplanter: A smart farming solution toward sustainability of paddy fields in Malaysia. Int J Agric & Biol Eng, 2019; 12(2): 54–67.

[3] Shamshiri R R, Ibrahim B, Ahmad D, Che Man H, Wayayok A. An overview of the System of Rice Intensification for Paddy Fields of Malaysia. Indian J Sci Technol, 2018; 11(18): 1–16.

[4] Feng X, Jiang Y, Yang X, Du M, Li X. Computer vision algorithms and hardware implementations: A survey. Integration the VLSI Journal, 2019; 69: 309–320.

[5] Li L, Zhang Q, Huang D. A Review of Imaging Techniques for Plant Phenotyping. Sensors, 2014; 14(11): 20078–20111.

[6] Haw C L, Ismail W I W, Kairunniza-Bejo S, Putih A, Shamshiri R. Colour vision to determine paddy maturity. Int J Agric & Biol Eng, 2014; 7(5): 55–63.

[7] Mochida K, Koda S, Inoue K, Hirayama T, Tanaka S, Nishii R, et al. Computer vision-based phenotyping for improvement of plant productivity: a machine learning perspective. GigaScience, 2019; 8(1): 1–12.

[8] Tang L, Zhu Y, Hannaway D, Meng Y, Liu L, Chen L, et al. RiceGrow: A rice growth and productivity model. NJAS - Wageningen J Life Sci, 2009; 57(1): 83–92.

[9] Yoshida S. Fundamentals of Rice Crop Science. Los Banos: The International Rice research Institute, 1981; 279p.

[10] Duan L, Huang C, Chen G, Xiong L, Liu Q, Yang W. Determination of rice panicle numbers during heading by multi-angle imaging. Crop J, 2015; 3(3): 211–219.

[11] Xiong X, Duan L, Liu L, Tu H, Yang P, Wu D, et al. Panicle-SEG: A robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. Plant Methods, 2017; 13(1): 1–15.

[12] Guo W, Fukatsu T, Ninomiya S. Automated characterization of flowering dynamics in rice using field-acquired time-series RGB images. Plant Methods, 2015; 11(1): 7. doi: 10.1186/s13007-015-0047-9.

[13] Desai S V, Balasubramanian V N, Fukatsu T, Ninomiya S, Guo W. Automatic estimation of heading date of paddy rice using deep learning. Plant Methods, 2019; 15(1): 76. doi: 10.1186/s13007-019-0457-1.

[14] Zhou C, Ye H, Hu J, Shi X, Hua S, Yue J, et al. Automated counting of rice panicle by applying deep learning model to images from unmanned aerial vehicle platform. Sensors, 2019; 19(14): 3106. doi: 10.3390/s19143106.

[15] Ikeda M, Hirose Y, Takashi T, Shibata Y, Yamamura T, Komura T, et al. Analysis of rice panicle traits and detection of QTLs using an image analyzing method. Breed Sci, 2010; 60(1): 55–64.

[16] Zhu Y, Cao Z, Lu H, Li Y, Xiao Y. In-field automatic observation of wheat heading stage using computer vision. Biosyst Eng, 2016; 143: 28–41.

[17] Sadeghi-Tehran P, Sabermanesh K, Virlet N, Hawkesford M J. Automated method to determine two critical growth stages of wheat: Heading and flowering. Front Plant Sci, 2017; 8: 252. doi: 10.3389/fpls.2017.00252.

[18] Kamilaris A, Prenafeta-Boldú F X. Deep learning in agriculture: A survey. Comput Electron Agric, 2018; 147: 70–90.

[19] Araus J L, Cairns J E. Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci, 2014; 19(1): 52–61.

[20] Yang G, Liu J, Zhao C, Li Z, Huang Y, Yu H, et al. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. Front Plant Sci, 2017; 8: 1111. doi: 10.3389/fpls.2017.01111.

[21] Ampatzidis Y, Partel V. UAV-based high throughput phenotyping in citrus utilizing multispectral imaging and artificial intelligence. Remote Sens, 2019; 11(4): 410. doi: 10.3390/rs11040410.

[22] Gnädinger F, Schmidhalter U. Digital counts of maize plants by unmanned aerial vehicles (UAVs). Remote Sens, 2017; 9(6): 544. doi: 10.3390/rs9060544.

[23] Jin X L, Liu S Y, Baret F, Hemerlé M, Comar A. Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. Remote Sens Environ, 2017; 198: 105–114.

[24] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. Proc IEEE Int Conf Comput Vis, 2017; 42(2): 386–397.

[25] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014; pp.580–587.

[26] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell, 2017; 39(6): 1137–1149.

[27] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell, 2017; 39(4): 640–651.

[28] He K, Zhang X, Ren S, Sun J. Deep Residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016; pp.770–778.

[29] Lin T-Y Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. 2017; pp.936–944.

[30] Waleed A. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. Github 2017; Available: https://github.com/matterport/Mask_RCNN. Acceseed on [2018-12-01].

[31] Neubeck A, Van Gool L. Efficient Non-Maximum Suppression. In: 18th International Conference on Pattern Recognition (ICPR'06). Hong Kong: IEEE, 2006; pp.850–885.

[32] Fletcher S J. Kalman filter and smoother. In: Data Assimilation for the Geosciences. Elsevier; 2017. pp.765–782.

[33] Lee W C, Krumm J. Trajectory preprocessing. In: Zheng Y, Zhou X, editors. Computing with Spatial Trajectories. New York: Springer New York, 2011; pp.3–33.

[34] Sugiura R, Noguchi N, Ishii K. Remote-sensing technology for vegetation monitoring using an unmanned helicopter. Biosyst Eng, 2005; 90(4): 369–379.

[35] Ester M, Kriegel H P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, 1996; pp.226–231.

[36] Archontoulis S V, Miguez F E. Nonlinear regression models and applications in agricultural research. Agron J, 2015; 107(2): 786. doi: 10.2134/agronj2012.0506.

[37] Sheehy J E. Bi-phasic growth patterns in rice. Ann Bot, 2004; 94(6): 811–817.