

RESEARCH ARTICLE

Rice Plant Counting, Locating, and Sizing Method Based on High-Throughput UAV RGB Images

Xiaodong Bai¹, Pichao Liu^{2*}, Zhiguo Cao³, Hao Lu³, Haipeng Xiong⁴, Aiping Yang⁵, Zhe Cai⁵, Jianjun Wang⁵, and Jianguo Yao²

¹School of Computer Science and Technology, Hainan University, Haikou 570228, China. ²School of Telecommunication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China. ³School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China. ⁴School of Computing, National University of Singapore, Singapore 119077, Singapore. ⁵Agricultural Meteorological Center, Jiangxi Meteorological Bureau, Nanchang 330045, China.

*Address correspondence to: 1220013418@njupt.edu.cn

Rice plant counting is crucial for many applications in rice production, such as yield estimation, growth diagnosis, disaster loss assessment, etc. Currently, rice counting still heavily relies on tedious and time-consuming manual operation. To alleviate the workload of rice counting, we employed an UAV (unmanned aerial vehicle) to collect the RGB images of the paddy field. Then, we proposed a new rice plant counting, locating, and sizing method (RiceNet), which consists of one feature extractor frontend and 3 feature decoder modules, namely, density map estimator, plant location detector, and plant size estimator. In RiceNet, rice plant attention mechanism and positive-negative loss are designed to improve the ability to distinguish plants from background and the quality of the estimated density maps. To verify the validity of our method, we propose a new UAV-based rice counting dataset, which contains 355 images and 257,793 manual labeled points. Experiment results show that the mean absolute error and root mean square error of the proposed RiceNet are 8.6 and 11.2, respectively. Moreover, we validated the performance of our method with two other popular crop datasets. On these three datasets, our method significantly outperforms state-of-the-art methods. Results suggest that RiceNet can accurately and efficiently estimate the number of rice plants and replace the traditional manual method.

Introduction

Rice is one of the three major food crops in the world. Approximately 162 million hectares of land are used for rice cultivation worldwide, with more than 755 million tons of paddy rice produced [1]. Rice plant counting in paddy field is a fundamental work and has a wide range of applications in rice production. In cultivation management, rice counting can be applied to choose a more effective planting density to improve the nutrition competition between crops and weeds [2,3]. In rice growth observation, plant counting can be used for the disaster assessment caused by typhoons and floods [4,5]. Moreover, in crop phenotype analysis, the number of survival rice plants is one of the key metrics in rice breeding. With plant counting, breeders are able selectively to choose parents to hybridize to cultivate the next generation of excellent rice varieties [6]. At present, rice plant counting in China still heavily relies on manual sampling and statistics. The manual counting method has many disadvantages. Initially, plant counting in the field is labor-intensive and cumbersome. Furthermore, because of time and cost constraints,

plant counts can only be performed in a small area in paddy field, making the results unrepresentative. In addition, human error often occurs during tedious manual counting. Finally, the manual observation in field tends to cause irreversible damage to rice. In this paper, we mainly focus on the accurate count of rice plants in paddy field with high-throughput unmanned aerial vehicle (UAV) images. Moreover, we also try to provide higher-level semantic information (plant location and size) that can contribute to downstream researches.

Many counting related techniques have been presented in agricultural research. Decision tree classifier and geometric descriptors were adopted to quantify early-season stand counts in corn [7]. Root nodule counting is utilized to indicate soybean health [8]. The pest population was counted to reduce the risk of crop exposure to pests and diseases [9]. Wheat ear counting approaches were proposed to estimate ear density under field conditions with zenithal color digital images [10] or thermal images [11]. The panicle numbers of pot-grown rice were determined by multiangle imaging and image segmentation [12]. Rice heading stage was automatic observation under field

Citation: Bai X, Liu P, Cao Z, Lu H, Xiong H, Yang A, Cai Z, Wang J, Yao J. Rice Plant Counting, Locating, and Sizing Method Based on High-Throughput UAV RGB Images. *Plant Phenomics* 2023;5:Article 0020. <https://doi.org/10.34133/plantphenomics.0020>

Submitted 26 September 2022

Accepted 9 December 2022

Published 30 January 2023

Copyright © 2023 Xiaodong Bai et al. Exclusive Licensee Nanjing Agricultural University. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License (CC BY 4.0).

conditions by multiclassifier-cascade-based rice spike detection procedures [13]. An indoor grain image acquisition system and grain counting algorithm were proposed for touching hybrid rice [14]. A Faster RCNN (Faster Region Convolutional Neural Network) detection method is utilized to detect wheat ears in an image [15]. Maize tassel counting method TasselNet was proposed by the means of local patch regression [16]. Liu et al. [5] used the image acquisition camera fixed in the field to obtain the rice image in a fixed area and realized the calculation of the rice planting density. Although many excellent approaches have been presented on corn stand [7], root nodule [8], pest [9], wheat ear [11], rice spike [13], etc. counting tasks, none of them can be used to accurate rice plant counting in paddy field. First, unlike the low-cost RGB camera that we use, previous researches may need to apply expensive infrared camera [11]. Second, some previous researches require specific imaging conditions, such as controlled illumination [9], multiangle imaging [14], and near range imaging at a fixed position [5,8,15]. Third, many traditional technologies are utilized in previous methods, such as the color threshold segmentation [12], watershed operation [14], and manual character classification [7,9]. It will make their methods unable to adapt to the light variation in rice field. Last but not least, the counting objects in previous researches are different from us [10,13,16]. Obviously, different image objects and datasets have their own characteristics. For the counting methods well-designed with other datasets, Their performance cannot be guaranteed in the tasks of rice plant counting. Indeed, it is still an open problem to count rice plants in paddy field accurately and efficiently.

Besides agricultural scenes, many achievements have been acquired in crowd counting [17–19], vehicle counting [20], cell counting [21,22], etc. Since crowd counting is the biggest and most successful branch in counting research, we mainly review the researches on crowd counting in this paper. Toward crowd counting, early methods were mainly based on detection [23], which applied low-level feature descriptors [24] and a sliding-window-based detection to obtain the object number. Essentially, these detection-based methods are usually based on classification, which requires a large number of manual bounding box annotations [25,26]. Of course, it is very time-consuming to prepare a large enough dataset for detection-based counting methods. However, severe occlusion, high congestion, and adhesion can make these methods perform poorly. To conquer the above difficulties, later researchers proposed regression-based methods that regress the object density map with the input image and then integrate the density map to obtain the count [27]. Nowadays, deep learning has brought unprecedented success to the research of crowd counting. Zhang et al. [18] are the first to apply deep learning to simultaneously estimate object density maps and object counts. Since then, deep learning has become the mainstream method [5,28–32]. The great breakthrough in crowd counting has important reference significance for the research of crop plant counting. However, there are still many shortcomings in the current deep learning counting methods. To begin with, the influence of irrelevant image background is still not completely solved, which will result in many false positives and a decline in counting accuracy [30,32,33]. In addition, many deep-learning-based counting networks need to apply complex image augmentation to ensure their performance [31,34–36]. Furthermore, the existing deep learning methods are mainly focused on counting. There are still many other applications that have not been excavated [5,28,29,37].

With the popularity of UAVs, the collection of high-throughput crop images has become more accessible and affordable in crop phenotype research [38,39]. Using the acquired rice images in paddy field, we found that we can not only obtain the number of rice plants but also realize their location and size estimation, which can be beneficial to downstream crop phenotype analysis. However, plant location and size estimation have not received enough attention recently. On the basis of the above discussions, the application of UAVs to complete field rice image acquisition and then design new deep learning networks to achieve automatic rice plant counting, locating, and sizing has far-reaching research significance.

In this paper, we present a new deep learning network RiceNet that can realize rice plant counting, locating, and sizing in paddy field with high-throughput RGB images from UAV. After considering the aforementioned shortcomings in current deep-learning-based counting methods, we made the following improvements in the design of RiceNet network: (1) To more effectively characterize the rice plants in high-throughput field images, multiscale features of different semantic information levels are carefully abstracted and incorporated in RiceNet. (2) To guide network to pay attention to more useful information in forward propagation, plant attention mechanism is presented and adopted in RiceNet. (3) Besides plant counting, RiceNet can also provide the plant location and size information by the presented plant location detector (PLD) and plant size estimator (PSE) modules. Those plant location and size information can contribute to downstream crop phenotype researches. In general, sufficient number of manually labeled bounding boxes is required for deep learning network to realize the locating and sizing of rice [40]. Of course, it is time-consuming and laborious to manually label thousands of rice plants with bounding boxes on a high-throughput rice image. On the basis of the reasonable assumption that rice plants are evenly distributed in the image, we initialize pseudo-rice plant bounding boxes from the point-level supervisory. Then, we leverage those pseudo-bounding boxes to guide the regression of rice sizes via an L1 loss. With the same assumption in the obtained high-quality estimated density map, a local nonmaximum suppression (LNMS) is utilized to the estimated density map to get the plant location. Finally, by fusing the obtained location results with the size information given by PSE, RiceNet brings an efficient estimation of the plant size. (4) A new positive-negative loss is proposed in RiceNet, which can be combined with L_{mse} and L_{bce} to enable network parameters to iterate accurately. The utilization of (1), (2), and (4) allows RiceNet to distinguish the rice plants in image and suppress the false positives in background more effectively. All the source code of RiceNet is available at <https://github.com/xdbai-source/Rice-Plant-Counting>.

In the current counting research, most count datasets are aimed at the crowd [19,30], cells [22], spikes [41], or flowers [42]. There is no UAV-based field rice counting dataset right now. In our research, we adopt low-altitude UAV to acquire RGB images of two paddy field from 2018 to 2020. Then, we propose a new UAV-based rice counting (URC) dataset that contains 355 image and 257,793 manual labeled points of rice plants. According to the current literature survey, we are the first to propose a rice plant counting dataset with UAV images, and the first to propose an efficient approach that simultaneously realizes the counting, locating, and sizing of rice. In experiments, our proposed method achieves state of the art on 3

datasets in terms of counting accuracy. In summary, our main contributions in this paper are as follows:

1. We present a new UAV-based high-throughput rice plant dataset (URC dataset) for the counting research of field rice plants. According to the current literature survey, this is the first rice plant counting dataset with UAV RGB images.

2. We propose a new network RiceNet that can achieve automatic, contactless, and accurate counting of rice plants in a large paddy field with UAV RGB imagery. In RiceNet, multi-scale feature fusion, plant attention mechanism, and positive-negative loss are presented and adopted to suppress false positives from the image background to generate high-quality density maps. Experiments show that the proposed RiceNet outperforms the state of the arts on our URC dataset and 2 other popular datasets.

3. With the designed PLD and PSE modules in RiceNet, our method also brings higher-level semantic information such as plant location and size. This high-level semantic information is of great significance to downstream phenotyping research tasks.

Materials and Methods

UAV-based crop RGB image acquisition

The rice images used in the experiment were collected in Nanchang City, Jiangxi Province, China. Its location is shown in Fig. 1A. The rice variety used for observation is indica-type rice. As shown in Fig. 1B, two rice fields were applied to collect the images by a quadcopter UAV (DJI Phantom 4 Advanced) and DJI GS Pro ground station. DJI GS Pro was utilized to automatically generate efficient flight paths and waypoints to complete the aerial photographing missions of the specified paddy fields. In DJI GS Pro, the flight altitude was fixed at 7 m above the ground. The RGB camera on the drone applied the vertical downward shooting. To prevent image blur, the image acquisition method of hovering shooting was adopted at each waypoint. The front and side overlap ratios were set to 80% and 70%, respectively. In the first image acquisition mission, we have to create a new UAV flight plan and set all flight parameters in GS Pro carefully. In the following missions, we only need to

open GS Pro and perform the above flight plan again. Rice images were collected every three days when the rice was between the tillering and jointing development stages from 2018 to 2020. For rainy and windy days when the UAV cannot take off, image acquisition will be postponed by one day. In this way, we collected an average of 25 UAV RGB image sequences per year. The original images with a resolution of $5,472 \times 3,648$ were captured and saved in the UAV SD (Security Digital) card in JPG format. After each flight mission, rice images were exported and stored in the data server.

UAV-based rice counting dataset

According to our existing literature survey, there is still no public UAV-based rice plant counting dataset. To realize the research of rice plant counting in a large paddy field, we present a new URC dataset in this paper. URC dataset contains 355 original high-throughput rice images (size of $5,472 \times 3,648$), which collected between 2018 and 2019. Following the standard practice [41], the center positions of all plants in each image in the URC dataset are manually annotated with dots. Compared with box labeling, point labeling is a more much reasonable and feasible way to represent so many rice plants in UAV images. Specially, we wrote a MATLAB R2018 script to accurately and conveniently label the center of each plant. Afterward, those images were downsampled into $0.25\times$ of the original resolution to facilitate network training. Last, we got a dataset with 355 images (size of $1,368 \times 912$) and 257,793 manually labeled points. Among them, 246 images were randomly selected and used as training images and the remaining 109 images as test images. Each image contains rice plants ranging from 84 to 1,125, with an average of 726 plants per image. In Fig. 2, we provide the rice plant histograms of the images in URC dataset.

The diversity of the rice images collected by UAV in the URC dataset is well illustrated in Fig. 3. For the convenience of demonstration, the image blocks of ten original UAV images are given in Fig. 3. As can be seen in Fig. 3, our dataset contains rice images taken under many lighting and weather conditions. Our URC dataset is not a single scene, and it is very representative. Moreover, we extracted some individual rice plants from



Fig. 1. Image collection process of rice field based on quadcopter UAV. (A) gives the collection location. (B) shows the image collection process in field 1 and field 2, respectively.

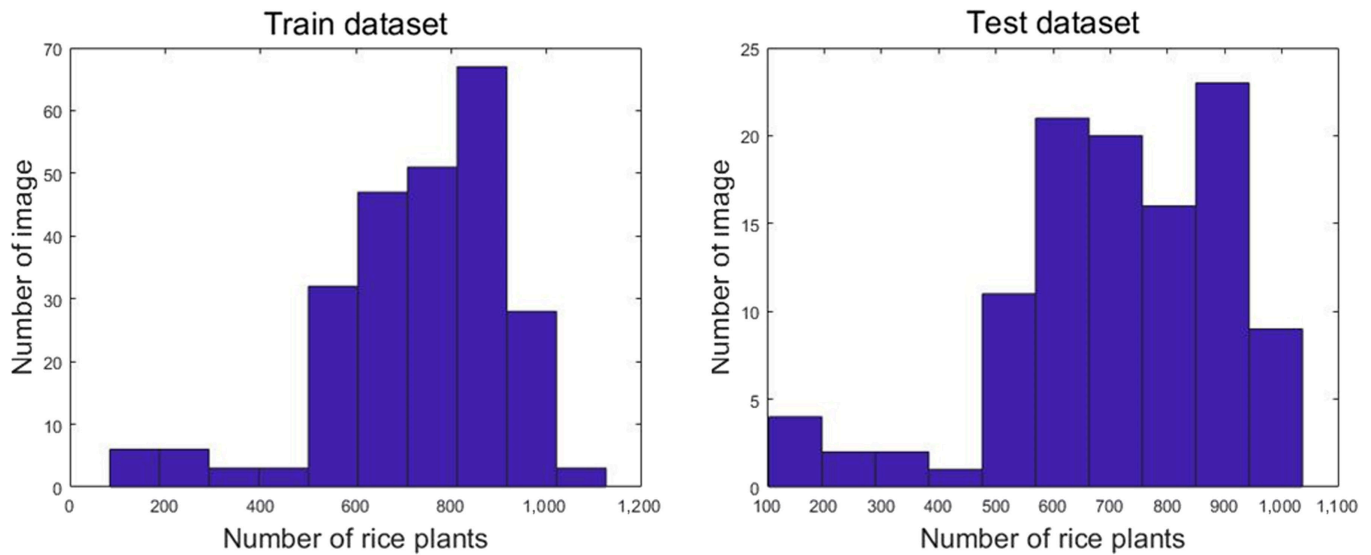


Fig. 2. Statistical distribution of the URC dataset.

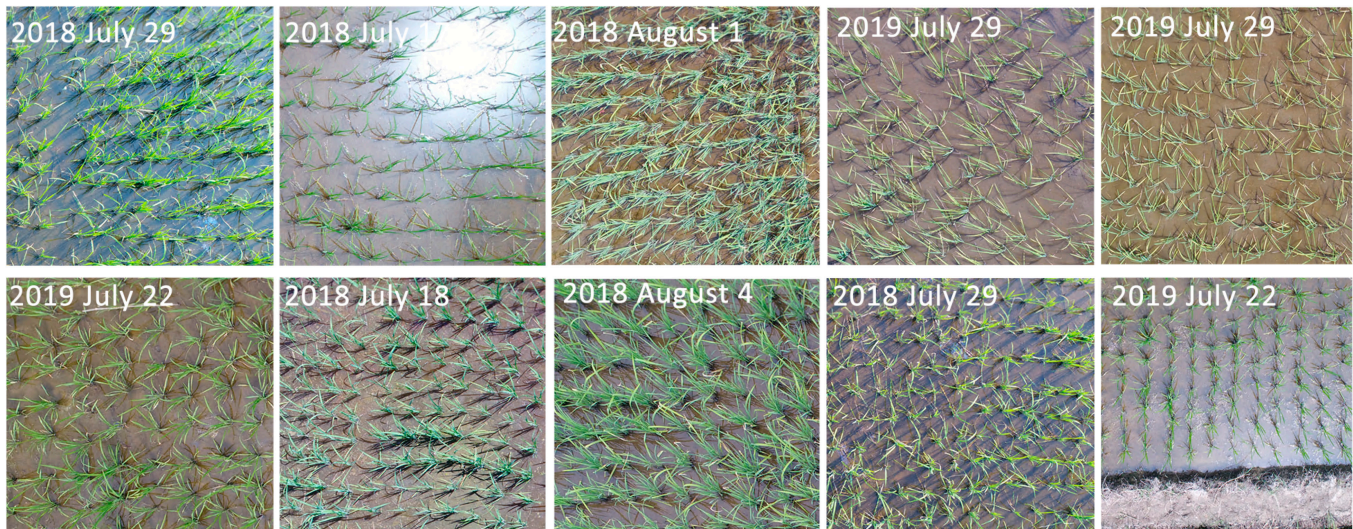


Fig. 3. Examples of image acquisition from different dates.

the images in the URC dataset, as shown in Fig. 4. Figure 4 shows the richness of rice plant traits in aspect ratio, color, orientation, size, etc. As shown in Figs. 3 and 4, URC dataset is a quite challenging dataset for rice plant counting.

Network architecture

In this section, we will introduce our network architecture. As illustrated in Fig. 5, RiceNet consists of a feature extractor and three designed network decoder modules, namely, density map estimator (DME), PSE, and PLD. In RiceNet, the first 13 layers of vgg16_bn [43] are used as the feature extractor. As shown in Fig. 5, an input image is first fed into the feature extractor, and then multilayer feature maps with different semantic levels are extracted. Next, the DME module adopts splicing and upsampling operations to fuse the multilayer features to generate high-quality estimated density map. Rice plant attention mechanism is added in the DME to improve the network's ability to distinguish plants from background. Furthermore,

RiceNet introduces the PLD module with LNMS to extract location information from the estimated density map and the PSE module to estimate the size information of the rice plants. Therefore, RiceNet can realize not only rice counting but also location and size estimation of the rice plants in UAV images. In the following subsections, we will describe each module in detail.

Density map estimator

We use the first 13 convolution layers in VGG16_bn [43] as the feature extractor of RiceNet. By the feature extractor, 4 multi-scale features L1, L2, L3, and L4 were extracted from an input image. The downsampling rates of those feature levels are 1/2, 1/4, 1/8, and 1/16, respectively. Next, we introduce the structure of the designed DME module. As shown in Fig. 5, Parser 1 merges the feature maps of L4 and L3, and Parser 2 merges the output of Parser 1 and the L2 feature map. Then, a density map head combines the output of Parser 2 and L1 to get the initial estimated density map. Specifically, the structures of Parser 1

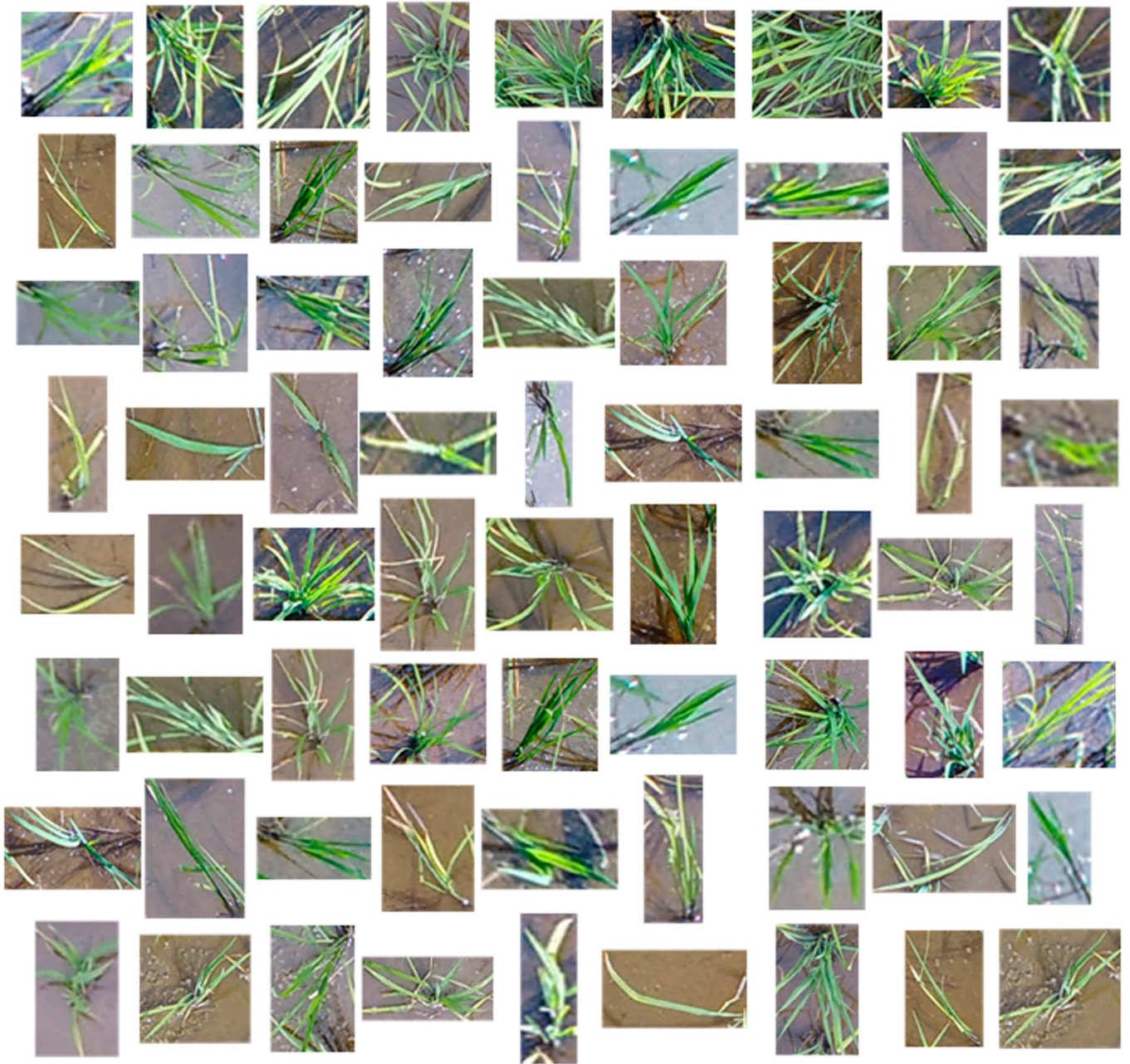


Fig. 4. Diversity of plant morphology in URC dataset.

and Parser 2 are given in Fig. 6. In Parser 1, L4 is upsampled to the same size as L3 and merged with L3. Next, several convolutional layers are employed to get the output. Parser 2 upsamples the output of Parser 1 and merges it with L2. Similar to Parser 1, it is also fed into several convolutional layers before output. Finally, the density map head fuses L1 and the output of Parser 2 to obtain the initial density map (IDM). The resolution of IDM is 1/2 of the original input image.

To further suppress the interference of image background and improve the accuracy of the IDM, plant attention mechanism is presented and adopted in DME module as shown in Fig. 5. The plant attention mechanism is implemented by Parser 3, Parser 4, and attention map head. Their detailed network structures were given in Fig. 6. Similar to Parser 1 and Parser 2, Parser 3 and Parser 4 fuse the L4, L3, and L2 feature maps and then transfer its output into the attention map head.

Attention map head merges the L1 with the output of Parser 4 and then followed by several convolutional layers. Finally, attention map head outputs a plant attention map (PAM), which is 1/2 size of the input image and used to refine IDM into a higher-quality density map. We can see that Parse 1 and Parse 3 share the same network design, and Parse 2 and Parse 4 have the same network design, which can reduce the complexity and difficulty of network model realization.

PAM will assist the IDM to distinguish the rice plants from the background and noise. In detail, element-wise multiplication and a convolution operation are applied on PAM and IDM to generate refined final density maps (FDMs) as Eq. 1:

$$\text{FDM} = \text{Conv}(\text{PAM} \odot \text{IDM}) \quad (1)$$

where \odot means Hadamard product and *Conv* represents convolution operation.

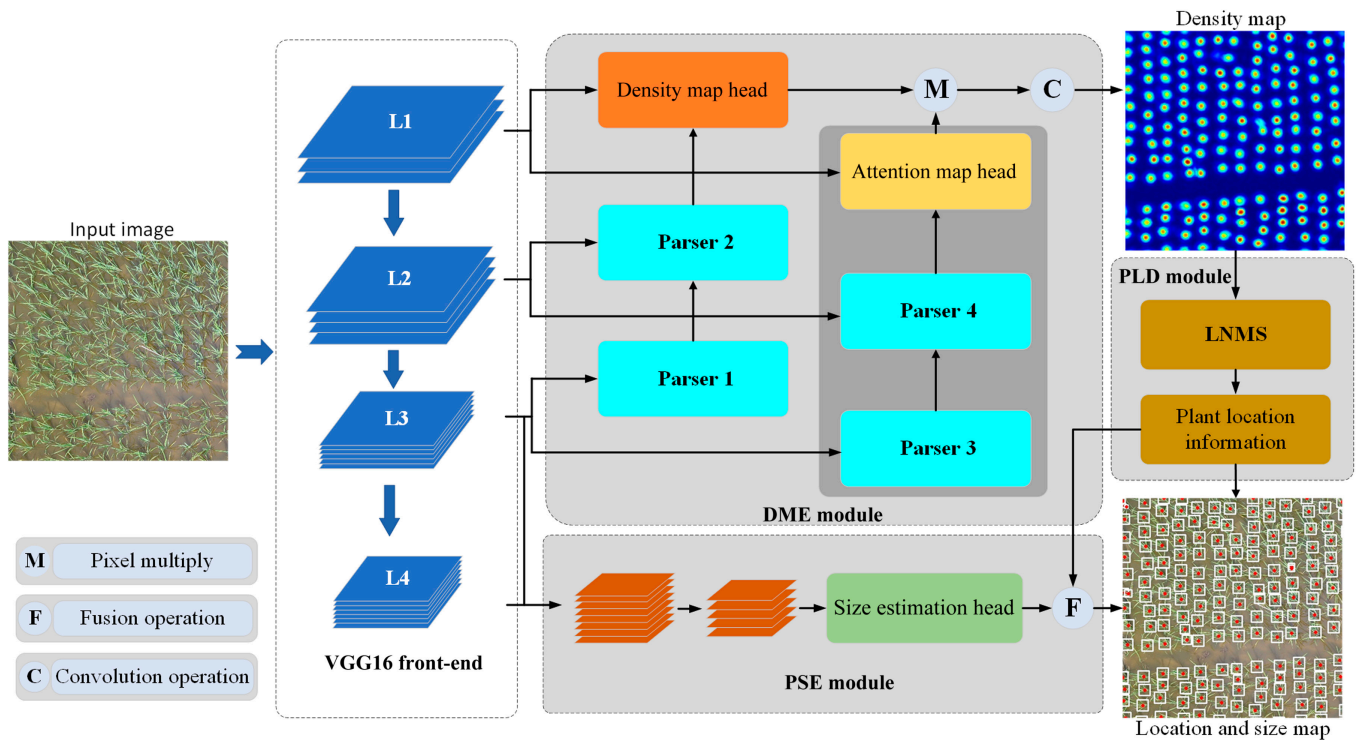


Fig. 5. The overall architecture of the proposed RiceNet. An input image is fed into the feature extractor to generate four feature maps containing different semantic levels. The DME combined with plant attention mechanism to leverage multiple feature maps to generate a high-quality density map. The PLD and PSE are employed to output higher-level semantic information such as size and location.

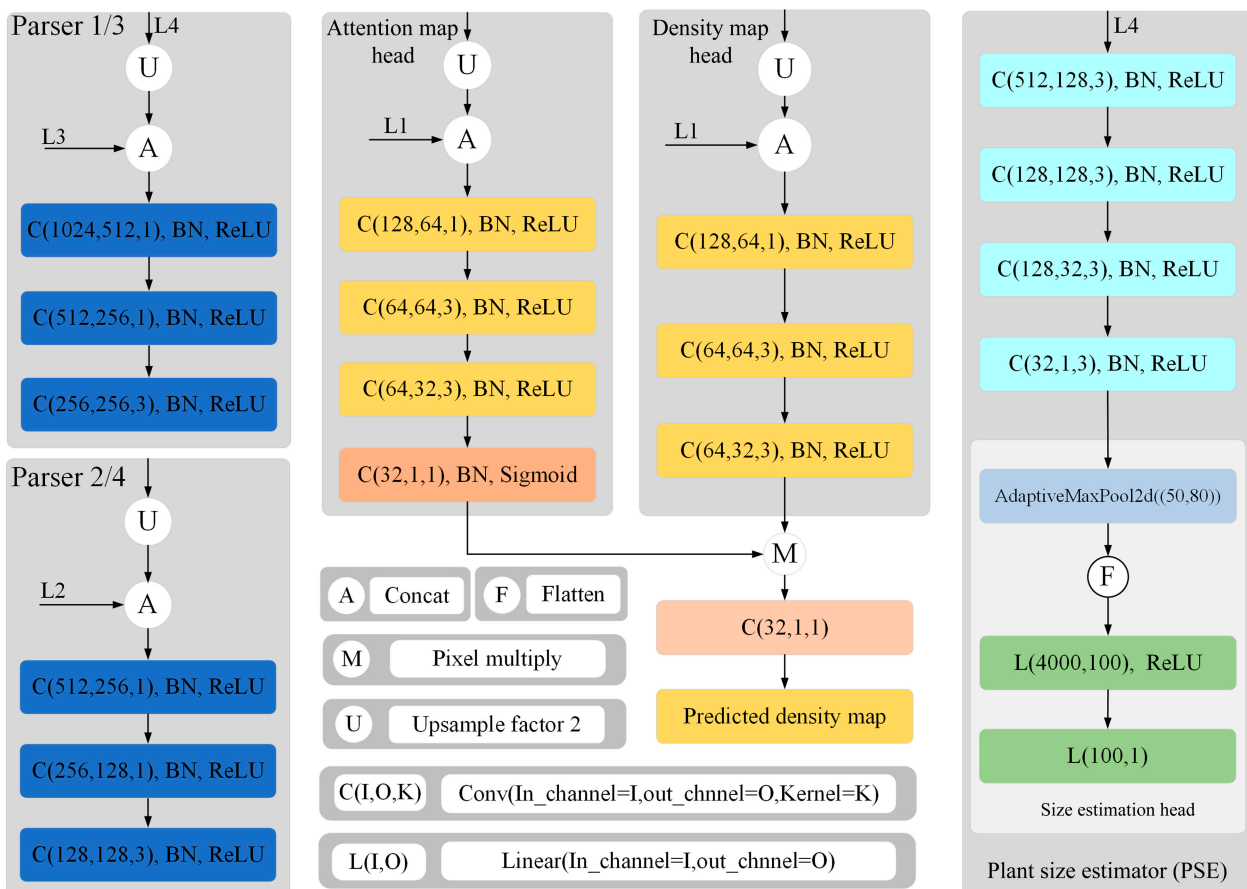


Fig. 6. Module structure diagram of RiceNet. Parser 1/3 and Parser 2/4 respectively integrate feature maps of different levels. Attention map head and density map head are used to generate PAM and IDM, respectively. We can see that PSE is a lightweight module that realizes the size prediction.

Now, we describe how to generate a ground-truth density map $D^{gt}(x)$, which is used for network training. If there is a plant at pixel position x_i , then we can represent it as a delta function $\delta(x - x_i)$. Inspired by Lempitsky and Zisserman [27], we further convolve the above function with a Gaussian kernel G_σ . Thus, suppose that there are N rice plants in an image, and ground-truth density map $D^{gt}(x)$ can be given as:

$$D^{gt}(x) = \sum_{i=1}^N \delta(x - x_i) * G_\sigma(x) \quad (2)$$

Next, we also introduce the generation method of the ground truth of the PAM. For the generated ground-truth density map $D^{gt}(x)$, we believe that the nonzero pixel value region represents the existence probability of the plants. That is to say, the region with zero pixel values in $D^{gt}(x)$ is the image background. Therefore, the ground truth of the PAM can be written as:

$$\forall x_i \in D^{gt}(x), G^{gt}(x_i) = \begin{cases} 1, & \text{if } x_i > 0.001 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Plant location detector

In the third column of Fig. 7, we can see that the density distribution of our estimated the FDM is almost uniformly distributed and independent of each other. Naturally, each peak position on the FDM can be considered as the center position of a plant. Therefore, with a given estimated FDM, the locations of the plants can be obtained through the LNMS. First, we generate a pseudo-size for each image as its ground-truth size with the previous manual point annotations. If there is a plant at pixel position x_j , following Shi et al. [44], then we first calculate the initial object size of point x_j according to the distances to its K -nearest points, and then the pseudo-average size D_{mean} of an image is calculated using Eqs. 4 and 5.

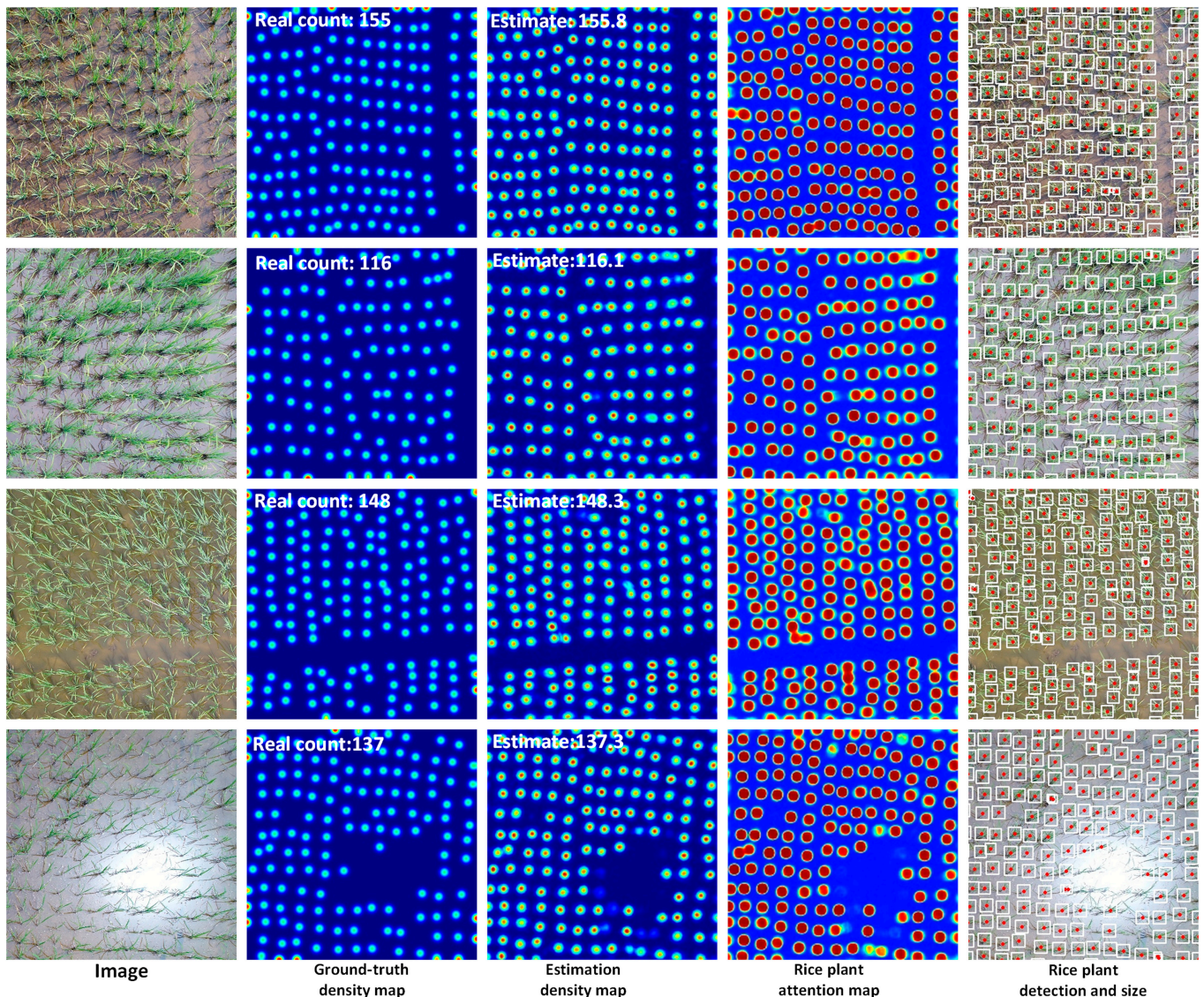


Fig. 7. Visualization of the counting, locating, and sizing results in the URC dataset. The first column is the four test images, the second column is the ground-truth density maps, the third column is the estimated density maps, the fourth column is the PAMs, and the fifth column is the plant location and size prediction results.

$$\bar{x}_j = \frac{1}{K} \sum_{k=1}^K \beta x_{j,k} \quad (4)$$

$$D_{mean} = \frac{1}{M} \sum_{j=1}^M \bar{x}_j \quad (5)$$

where $x_{j,k}$ is the distance between point x_j and its k th nearest neighbor. \bar{x}_j represents the initial object size of x_j . β is a scalar and generally can be set to 0.8. M is the number of the plants. Second, following Zhang et al. [30], we adopt max pooling to obtain local maximum mask map, where the max pooling kernel size is set to $D_{mean} \times 0.3$. Third, the local maximum mask map and the density map are multiplied to obtain all local maxima. However, those local maximum points may even contain some false positives from the background. According to the high-quality FDM generated by the DME module, we observe that the values of those false positives are much smaller than the true positives from the plants. It means that those positions with very small local maxima may originate from the image background. Finally, given an estimated FDM and its maximum pixel value M , an adaptive threshold $M \times 0.3$ can be used to filter out the false positives. In particular, if the maximum value M is less than a small value (set to 0.01), then it means that there are no plants in the input image. The main flow of the PLD module to obtain the location information is shown in Algorithm 1.

Algorithm 1. LNMS algorithm.

Input: FDM - Predicted FDM of a given image I;
 D_{mean} - Pseudo-average size of plant in I.
Output: The coordinates of the plant
1: **Function** Extract coordinates(FDM, D_{mean}):
2: $Max_mask = max_pooling(FDM, kernel_size = (D_{mean} * 0.3, D_{mean} * 0.3))$
3: $Max_mask = (Max_mask = FDM)$
4: $Max_all = Max_mask \times FDM$
5: $adaptive_threshold = max(Max_all) \times 0.3$
6: **if** $max(Max_all) < 0.01$ **then**
7: $coordinates = None$
8: **else**
9: $Max_all[Max_all > adaptive_threshold] = 1$
10: $Max_all[Max_all < 1] = 0$
11: $coordinates = nonzero(Max_all)$
12: **end if**
13: **return** coordinates
14: **end function**

Plant size estimator

Let $\{d_j\}_{j=1}^M$ represent the coordinates generated by Algorithm 1, where $d_j = (x_j, y_j)$ is the two-dimensional coordinates of the center of the j th estimated plant in an image, and M is the total number of coordinates. Then, we can estimate a size value for each plant through the K -nearest neighbors algorithm.

$$\bar{d}_j = \frac{1}{K} \sum_{k=1}^K \beta \sqrt{(d_j - d_k)^2} \quad (6)$$

where \bar{d}_j is the initial object size of d_j , d_k is the k th point closest to d_j , the value of K is 3, and β is a scalar. However, for the plants in sparse region, their sizes obtained by the Eq. 6 will be too large. Therefore, we design a size estimate head structure in our research to regress a size limitation of the plants in the input image. The detailed network structure of the PSE module is given in Fig. 6. In the PSE, L4 feature map is processed by several convolutional layers to further abstract high-level information. Afterward, an adaptive pooling, a flattening process, and 2 linear layers are applied to output the size limitation. We use L1 loss as the loss function of the PSE module and D_{mean} (mentioned in Plant location detector section) as the regression target. Finally, the size estimation comes from the fusion of the PSE and the PLD outputs, as described in Fig. 5. If the output value of PSE module is d_p , then we combine the output of PSE module to bring a more reasonable plant size D_j , which can be written as:

$$D_j = \begin{cases} d_p, & \text{if } d_p < \bar{d}_j \\ \bar{d}_j, & \text{otherwise} \end{cases} \quad (7)$$

Loss functions

In the training phase, we train the plant attention mechanism with a pixel-wise binary cross entropy loss between the network predicted PAM M^p and its corresponding ground truth G^{gt} , which can be given as:

$$L_{bce} = -\frac{1}{N} \sum_{i=1}^N (G_i^{gt} \cdot \log(M_i^p) + (1 - G_i^{gt}) \cdot \log(1 - M_i^p)) \quad (8)$$

where N is the batch size. Moreover, we train the DME module with mean squared error loss between the estimated density map D^p and its corresponding ground-truth map D^{gt} by:

$$L_{mse} = \frac{1}{N} \sum_{i=1}^N (D_i^p - D_i^{gt})^2 \quad (9)$$

To better suppress the false positives in background, we propose a positive-negative loss function. For the generated ground-truth density map $D^{gt}(x)$, we consider that the region with nonzero pixel value in ground-truth map is the positive region. Similarly, the region where the pixel value is zero is the negative region. The probabilities of having objects in the positive region and the negative region are p and 0, respectively. Correspondingly, the probabilities of being the background are $1 - p$ and 1, respectively. Let $\{D^{gt}(x_m)\}_{m=1}^M$ be a density map, where x_m denotes a two-dimensional pixel location and M is the number of pixels in the density map $D^{gt}(x)$. So, we can get the object positive-negative probability map $P^O(x_m)$ and background positive-negative probability map $P^B(x_m)$ by:

$$\forall x_m \in D^{gt}(x_m), P^O(x_m) = \begin{cases} p, & \text{if } x_m > 0 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

$$\forall x_m \in D^{gt}(x_m), P^B(x_m) = \begin{cases} 1-p, & \text{if } x_m > 0 \\ 1, & \text{otherwise} \end{cases} \quad (11)$$

The expected counts for positive region and for the entire negative region are defined as

$$C_O = \sum_{m=1}^M P^O(x_m) \cdot D^p(x_m) \quad (12)$$

$$C_B = \sum_{m=1}^M P^B(x_m) \cdot D^p(x_m) \quad (13)$$

In this case, the summation over the whole density map D^p consists of the object positive–negative counts C^O and the background positive–negative count C^B . Obviously, we would like the background positive–negative count to be zero, and the object positive–negative count is equal to the real count value C^{st} . Thus, we have the following positive–negative loss function,

$$L_{si} = \left| (C^{st} - C_O) + (0 - C_B) \right| \quad (14)$$

where C^{st} is total number of objects, so the total loss function is as follows:

$$L_{all} = L_{mse} + \lambda L_{bce} + \gamma L_{si} \quad (15)$$

Evaluation metric

To assess the difference between the predicted counts and the ground-truth numbers, we applied the commonly used mean absolute error (MAE) and root mean square error (RMSE). The definitions of MAE and RMSE are as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |Z_i - \hat{Z}_i| \quad (16)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Z_i - \hat{Z}_i)^2} \quad (17)$$

where Z_i is real number of rice plant in the i th image, \hat{Z}_i is the estimated total number of plants in the i th image, and N is number of test image. The MAE indicates the accuracy of different approaches, while the RMSE reflects their robustness. A smaller value of MAE and RMSE indicates better network performance.

Implementation details

In the experiment, to reduce computational consumption, the resolution of input images is downsampled into 1/4 of the original resolution of 5,472×3,648. Image blocks in 320 × 320 are randomly cropped from URC dataset images, and then they are randomly horizontal flipped with a probability of 0.5 and processed by a gamma contrast transform with a probability of 0.3 for data augmentation. Adam optimizer is utilized in the training process. The initial learning rate and batch size are set to 1×10^{-4} and 3, respectively. The hyperparameter λ and Υ of the loss function are both set to 0.1. The value of p in the positive–negative loss function is set to 1. We first trained the

DME module to output the estimated high-quality density maps. Afterward, the PSE module is trained with the parameter fine-tuning of the frontend feature extractor. Our experiment is implemented by the PyTorch framework and applied GPU NVIDIA RTX 3090 for acceleration.

Results

Experiment on the URC dataset

In this part, we compared our method with previous state of the arts and analyzed their results on the proposed URC dataset. Table 1 shows the performance of different methods on the URC dataset. The leftmost column is several advanced counting methods and our approach. The third and fourth columns are their MAE and RMSE results. As can be seen from Table 1, our method outperforms other methods by a large margin. The MAE and RMSE of the proposed method reach 8.6 and 11.2, respectively. RiceNet sets the new state of the art with clear advantages over other competitors on the URC dataset. What is striking about the figures in Table 1 is that compared with the TasselNetV2 and FIDTM, the MAE of our method is improved by 70.8% and 67.3%, respectively. We believe that the reason why RiceNet performed better than TasselNetV2 is that RiceNet adds new plant attention mechanism in our DME module and enhances parameter learning capability through new loss function. Compared with the crowd counting approaches including FIDTM, their networks mainly solve the challenges in head scale variation and density distribution imbalance. Obviously, this is not the case for the almost evenly distributed rice plants in paddy field. Thus, our method can focus more on improving the accuracy of plant counting.

In the first and second columns of Fig. 7, we demonstrate four test images in the URC dataset and their ground-truth density map. The third column of Fig. 7 shows the estimated density map by RiceNet. We can see that the estimated high-quality density map in the third column is very close to the ground-truth density map. At the same time, we also give the visualization of the predicting rice PAM in the fourth column of Fig. 7. We can see that the plant attention mechanism in the model promotes the model to focus more on rice plants and plays an important role. To summary up, the adoption of new plant attention mechanism and new loss function (see Ablation experiment section) makes RiceNet more suitable for plant counting, which we believe is the reason why RiceNet performs better than other methods.

Table 1. The performance of different methods on the URC dataset.

Networks	Venue, year	MAE	RMSE
MCNN [30]	CVPR, 2016	25.5	34.1
CSRNet [17]	CVPR, 2018	12.9	17.5
SANet [32]	ECCV, 2018	10.1	13.4
TasselNetV2 [41]	PLME 2019	29.5	39.4
FIDTM [45]	Arxiv, 2021	26.3	31.5
RiceNet	This paper	8.6	11.2

By comparing the second and third columns in the last row of Fig. 7, we can find that the estimated density map given by RiceNet is even slightly better than the ground-truth density map around the image region with strong sun reflection. This reflects that our network has a certain robustness to the sun reflection in paddy field. Actually, how to fully overcome the impact of the strong sun reflection is still an open problem that needs further research.

Experiment on the MTC dataset

In addition, we also did comparison experiment on the MTC dataset. The MTC dataset is a maize tassels count dataset first introduced by Lu et al. [16]. It includes 361 images collected from four experimental fields across China between 2010 and 2015 with six maize cultivars. Maize tassels in the MTC dataset have very large scale and shape changes. The original resolution of those images is $3,648 \times 2,736$, $4,272 \times 2,848$, or $3,456 \times 2,304$. In this dataset, 186 images are randomly selected and used as the training set, and the remaining 175 images are used as the test set. The number of maize tassels in each image varies from 0 to around 100. In the experiment, the short sides of all images in the MTC are proportionally resized to 512 to speed up the calculation. Next, image blocks in the size of 256×256 are randomly cropped and used in the training. Other training details are consistent with the URC dataset. Table 2 shows the performance of different methods on the MTC dataset. As shown in Table 2, RiceNet achieves the best MAE (3.4) and RMSE (5.3), compared with other approaches. It is worth mentioning that our method achieves 81.0% improvement in MAE and 75.8% improvement in RMSE compared to the MCNN. Many images in the MTC dataset have very few maize tassels. This experiment shows that our network can also achieve good performance for the object sparse scenes. We believe that better performance in our method is due to its well object feature extraction and fusion ability between different levels of features. This indicates the superiority of the DME module in our method.

In addition, some counting results in this dataset are demonstrated in Fig. 8. The third column of Fig. 8 shows that the estimated density map is similar to the ground-truth density map and the estimated counting result is close to the real value. We found that several maize tassels were lost in the network estimation density map. This phenomenon is mainly due to the shape and size of these maize tassels that are too special compared to others. As seen from the second row of Fig. 8, our network gives a high-quality estimation density map and

counting result even when there are raindrops in the original RGB maize image. From another aspect, it shows the well performance of our proposed network. Briefly, Table 2 and Fig. 8 indicate that the proposed RiceNet can also achieve good performance on the task of maize tassel counting.

Experiment on the WED

The WED is a widely used wheat ear dataset, which is first introduced by Madec et al. [6]. The wheat images collected in fields have 20 different genotypes. The image resolution is $6,000 \times 4,000$. The number of ears in each image varies from 80 to 170. This dataset includes 236 images, where 165 and 71 images are used for training and testing, respectively. For some genotypes, the color of wheat ears is extremely similar to the color of adjacent leaves. Bounding box annotations are provided in this dataset, while we only use the center point of each box in the experiment so as to unify the comparison of different methods. The resolution of the image is sampled to 1/8 of the original resolution and T is set to 0.01. Image block in 256×256 was randomly cropped from those downsampled images and applied in the training. Other training details are consistent with the URC dataset. Table 3 shows the performance of different methods on the WED. The MAE and RMSE of our method reached 3.7 and 4.6, respectively. In particular, compared to the MCNN, we get 67.8% MAE and 70.5% RMSE improvement. We believe that the plant attention mechanism strengthens the ability of the proposed network to distinguish objects from image backgrounds and improves the counting accuracy. Moreover, we bring some counting results in the WED in Fig. 9. The results in Fig. 9 demonstrate that the network estimated density maps and counting results have a very high accuracy even when the distribution of the crop objects (wheat ears in the WED) is disordered. Table 3 and Fig. 9 indicate the well performance of the proposed RiceNet on the task of wheat ear counting.

Coefficient of determination analysis of counting results

For the count results of the 3 different crop count datasets, we calculated their R^2 values by Eq. 18. As shown in Fig. 10, the R^2 values of 3 datasets reach 0.997, 0.957, and 0.975, respectively. Figure 10 demonstrates that the estimated counting results and the real values obtained by manual counting method have a high correlation. From another side, Fig. 10 indicates

Table 2. Performance of different methods on the MTC dataset.

Networks	Venue, year	MAE	RMSE
MCNN [30]	CVPR, 2016	17.9	21.9
CSRNet [17]	CVPR, 2018	6.9	11.5
BCNet [46]	TCSVT, 2019	5.2	9.2
SFC ² Net [5]	PLPH, 2020	5.0	9.4
TasselNet [16]	PLME, 2017	6.6	9.9
TasselNetV2 [41]	PLME, 2019	5.4	9.2
RiceNet	This paper	3.4	5.3

Table 3. Performance of different methods on the WED.

Networks	Venue, year	MAE	RMSE
MCNN [30]	CVPR, 2016	11.5	15.6
CSRNet [17]	CVPR, 2018	4.2	5.2
BCNet [46]	TCSVT, 2019	4.1	4.9
Faster R-CNN [6]	AGRFOMET, 2019	4.6	5.9
SFC ² Net [5]	PLPH, 2020	4.2	5.1
TasselNet [16]	PLME, 2017	6.8	8.3
TasselNetV2 [41]	PLME, 2019	5.3	6.8
RiceNet	This paper	3.7	4.6

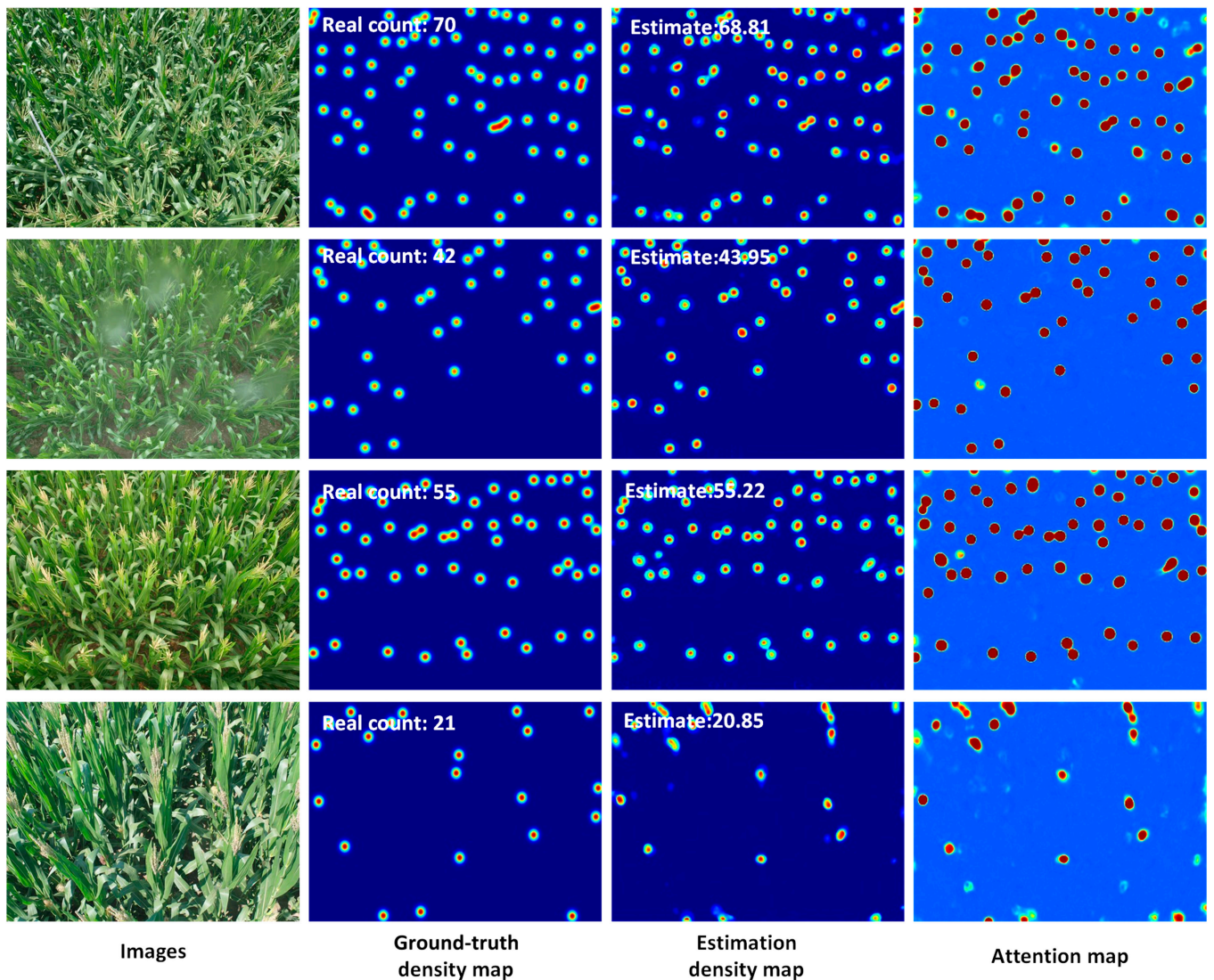


Fig. 8. Visualization of the counting results in the MTC dataset. The first column is the 4 test images, the second column is the ground-truth density maps, the third column is the estimated density maps, and the fourth column is the PAMs.

that our network not only has high counting accuracy but also has good generalization capabilities. According to the plots of coefficients of determination in Fig. 10, there are still some count results of being overestimated or underestimated, especially for the MTC dataset. This phenomenon is mainly due to the more severe shape and color variation of maize tassels in the MTC dataset.

$$R^2 = 1 - \frac{\sum_{i=1}^N (P_i - G_i)^2}{\sum_{i=1}^N (P_i - \bar{G})^2} \quad (18)$$

Ablation experiment

In this section, we verify the effectiveness of the added positive-negative loss with the URC dataset. As shown in the last row of Table 4, the positive-negative loss can improve the counting performance of the proposed RiceNet on the URC dataset. Meanwhile, we incorporate positive-negative loss on other counting methods so as to explore the effectiveness of this loss function on other networks. Table 4 shows the effect of

positive-negative loss in different models on the URC dataset. It can be seen from Table 4 that the performance of MCNN, CSRNet, and TasselNetV2 is improved after adding positive-negative loss. Compared with MCNN, MCNN with positive-negative loss decreases by 54.1% in MAE.

In addition, we analyze the influence of the parameter σ mentioned in Density map estimator section. The σ parameter determines the size of the Gaussian kernel when generating the ground-truth density map. From Table 5, we can see that when the σ value is 6 and the network performs best. We can observe from the Table 5 that the performances with different σ values do not change seriously. Comparable performance was achieved with Gaussian kernel sizes of 2 and 4. The experimental results in Table 5 show the robustness of our method to the Gaussian kernel size parameter.

Estimation of plant location and size

In previous methods, to estimate the plant location and size, it is often necessary to resort to the bounding box labeling method. For rice UAV images with so many rice plants, bounding

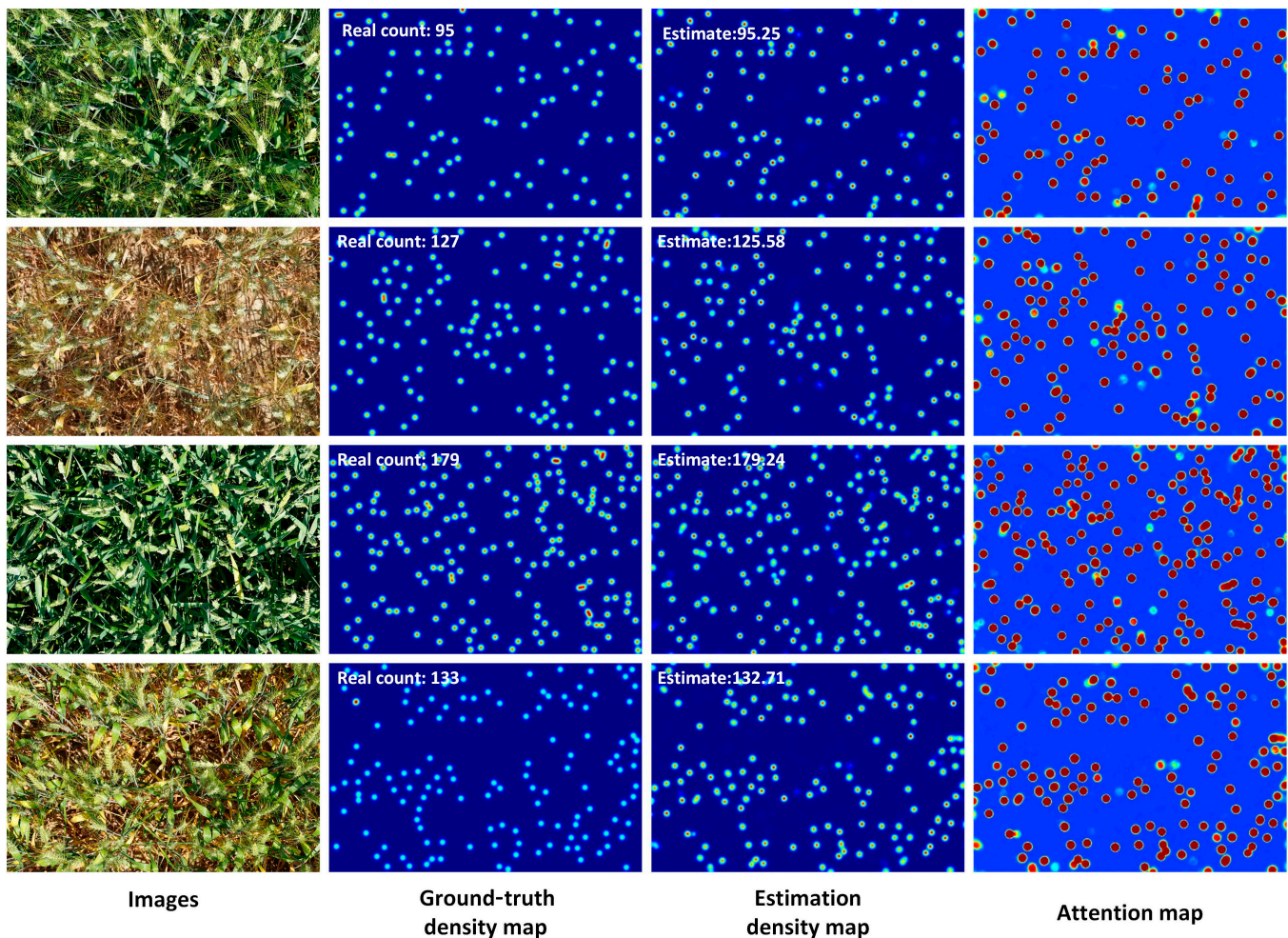


Fig. 9. Visualization of the counting results in the WED dataset. The first column is the 4 test images, the second column is the ground-truth density maps, the third column is the estimated density maps, and the fourth column is the PAMs.

box labeling will be very time-consuming and labor-intensive, and it is not feasible. In contrast, the point labeling method is relatively simple, and it is easier to achieve a large number of labeling of the plants. In our method, on the basis of the reasonable assumption that rice plants are evenly distributed in each URC image, we utilize the distance between manual point annotations to generate pseudo-sizes as a supervision during network training. Thus, we achieve location and size estimation of rice plants with the manual point annotations and point-level supervisory. As seen from the fifth column of Fig. 7, with the designed PLD and PSE modules, RiceNet not only outputs the number of plants but also gives the location (red points) and size (white boxes) information of the rice plants. These location and size information are of great significance for subsequent crop phenotype researches and refined field managements. Since MTC and WED datasets are not satisfied with uniform distribution assumption, their location and size information cannot be calculated by the current version of RiceNet.

Further discussions

In this section, we will further discuss the suggestions in the application of our rice plant counting, locating, and sizing network, the current shortcomings, and the future improvement directions. According to our three years (from 2018 to 2020) of experience in UAV image collection in paddy field and the

after research of deep learning network design, we bring the following suggestions in the application of our RiceNet:

1. It is suggested not to collect UAV images on rainy days. Rainwater will not only cause damage to drone motors and gimbal camera but also cause image quality degradation and make the plant images to blur.
2. It is best to collect UAV images below the fifth-level wind speed. Excessive wind speed in field can enlarge the jitter of the drone resulting in an inaccurate camera autofocus. Moreover, it will change the shape of rice leaves and plants in rice image, giving rise to the decrease in network performance.
3. It is recommended to collect drone images in about 4 hours after sunrise, so as to avoid most fog time in the morning and the rice leaf curls caused by the high temperature at noon.
4. In our research, observers need to come the paddy field and perform drone control in each image collection mission. With the development of drone 5G technology, future 5G + UAVs can achieve remote flight control. In the future, the adoption of this type of UAVs can greatly reduce the difficulty of UAV image collection.
5. Following the usual field management methods, we used an appropriate amount of herbicide in the rice field. The

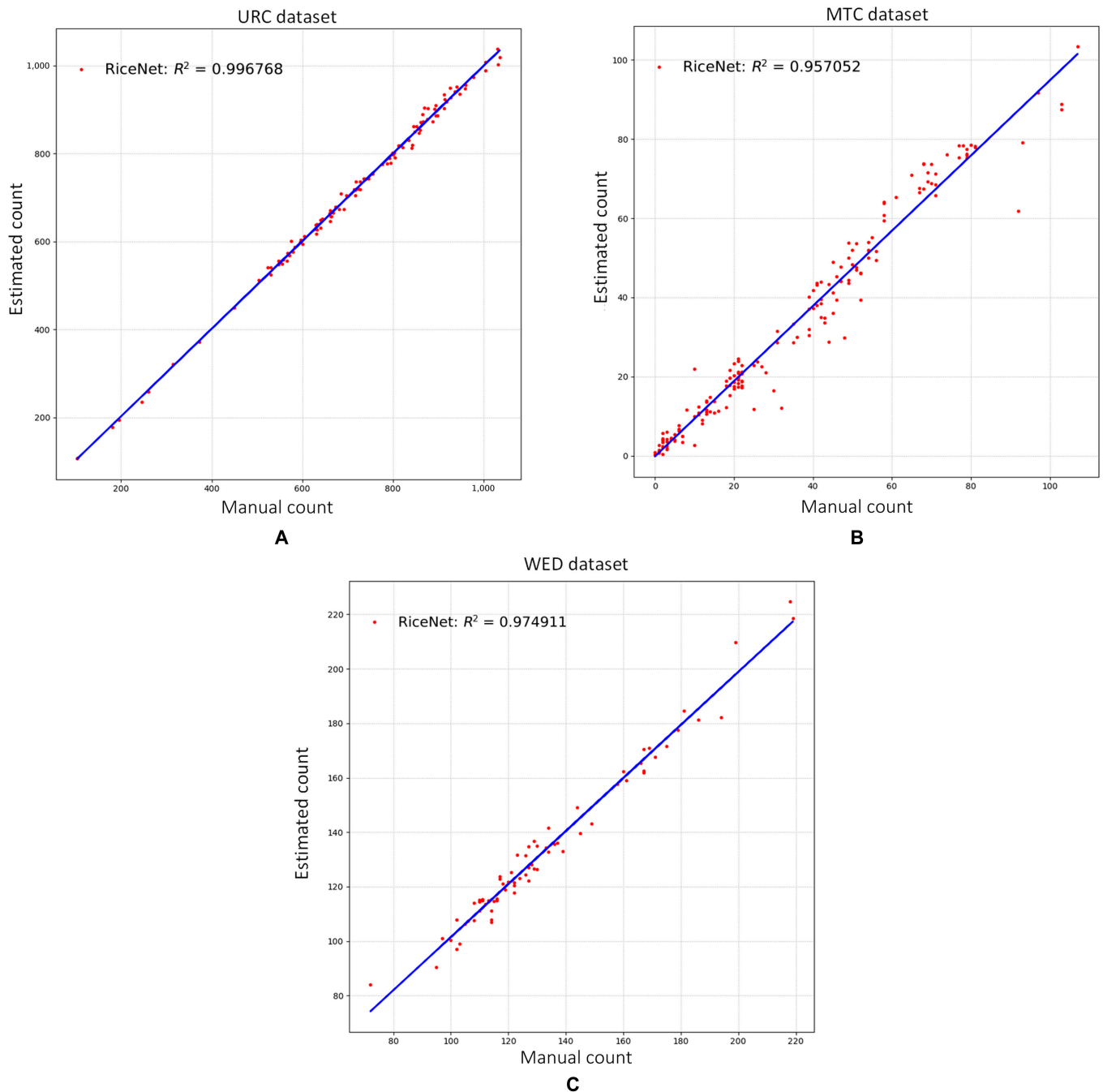


Fig. 10. Coefficients of determination of the RiceNet on different datasets. (A), (B) and (C) are RiceNet counting results on URC, MTC and WED, respectively.

appearance of some weeds (such as Barnyard grass, *Scirpus juncooides*, etc.) is very similar to the rice plants in paddy field. Without using herbicide, our network may degrade the counting performance because of the emergence of weeds.

6. It is recommended to use camera overhead imaging. Oblique imaging will cause occlusion between adjacent rice plants in the obtained image. Moreover, oblique imaging will add challenges to the estimation of plant location and size.

Next, we listed the shortcomings in our current RiceNet counting method and the possible research directions in the future:

1. During the periods of rice irrigation, sun reflection on the water may cause an overexposure region in the obtained image, as shown the last row of Fig. 7. The color information of the pixels in these overexposure region will be severely or completely lost. In future research, the design of new networks that can be more robust to sun reflection in the field will be an important research direction.
2. In RiceNet, multiscale feature fusion, plant attention mechanism, and positive-negative loss are adopted to suppress false positives from the image background. However, false positives still cannot be completely eliminated in the experiment. Therefore, the suppression of false positives from the background is still worth further research.

Table 4. The effect of positive–negative loss in different models on URC dataset.

Networks	Without positive–negative loss		With positive–negative loss	
	MAE	RMSE	MAE	RMSE
MCNN [30]	25.5	34.1	11.7↓	16.5↓
SANet [32]	10.1	13.4	10.8	14.9
CSRNet [17]	12.9	17.5	9.2↓	14.0↓
TasselNetV2 [41]	29.5	39.4	15.1↓	20.3↓
RiceNet	10.6	13.6	8.6↓	11.1↓

Table 5. The effect of sigma in the generation of real density map.

σ	MAE	RMSE
2	8.7	11.5
4	8.8	12.1
6	8.6	11.2
8	9.2	12.5

- In this paper, the network performance of rice plant locating and sizing is qualitatively checked and evaluated with human eyes, since the URC dataset adopted plant point labeling instead of bounding box labeling. In future research, finding how to quantitatively evaluate the network performance of plant locating and sizing will be another important research direction.
- In our project, we only need to focus on improving the accuracy of rice plant counting. All calculations were completed on a GPU-accelerated server after each UAV image collection mission (once every three days). Thus, we do not need to consider about of calculation complexity and running time. In future research, the design of lightweight networks that can run on the field robots or other embedded platforms in real time may also be a good research direction.

Conclusion

This paper proposes a new rice plant counting, locating, and sizing approach with UAV imagery in paddy field. RiceNet includes one multiscale front-end feature extractor and 3 feature decoder modules (DME, PSE, and PLD). Plant attention mechanism and positive–negative loss are presented and utilized in the DME to improve the network’s ability to distinguish backgrounds to generate high-quality density maps. On the basis of the designed PLD and PSE modules, RiceNet not only can output the count number of rice plants but also can realize the

plant location and size estimation with the manual point annotations. To verify our method, we present a new URC dataset, which consists of 355 images and 257,793 manual point annotations. In the experiment, the MAE and RMSE of RiceNet on the URC dataset are 8.6 and 11.2, respectively, which significantly outperforms states-of-the-art methods. Furthermore, we did counting experiments on two other famous crop counting datasets, and our method outperforms other methods by a large margin. In ablation experiments, we demonstrated the effectiveness of the added positive–negative loss and analyzed the influence of the sigma parameter on the counting performance. From the experiments, the proposed RiceNet can realize accurate, contactless, and efficient counting of rice plants in paddy field with UAV imagery. Moreover, RiceNet can provide higher-level semantic information (plant location and size). Access to those information can contribute to other crop phenotyping researches and help advance the development of automated fertilizer applicators, sprayers, etc. Last, we discussed the suggestions in the application of RiceNet, the current shortcomings of RiceNet, and the possible research directions in the future.

Acknowledgment

We would like to thank the field management staff at the Nanchang Key Laboratory of Agrometeorology for continued support of our research. We also would like to thank S. Chen, T. Xie, X. Zhou, D. Chen, and H. Li for contributing to manual image annotation. We thank S. Chen and T. Xie for contributing to the language revision. **Funding:** This work was supported in part by the National Natural Science Foundation of China (grant nos. 61701260, 61876211, and 62271266), the Postgraduate Research and Practice Innovation Program of Jiangsu Province (grant no. SJCX21_0255), and the Postdoctoral Research Program of Jiangsu Province (grant no. 2019K287). **Author contributions:** Conceptualization: X.B., P.L., and Z. Cao; Methodology: P.L., H.L., H.X., and X.B. Validation: P.L. and X.B. Formal analysis and investigation: X.B., A.Y., and Z. Cai. Resources: A.Y., Z. Cai, and J.W. Writing and writing–review: X.B. and P.L. Editing and visualization: all authors. Project administration and funding acquisition: X.B., Z. Cao, and J.Y. All authors have read and agreed to the published version of the manuscript. **Competing interests:** The authors declare that they have no competing interests.

Data Availability

The data presented in this study are available on request from the corresponding author.

References

- Shaheen SM, Antoniadis V, Shahid M, Yang Y, Abdelrahman H, Zhang T, Hassan NEE, Bibi I, Niazi NK, Younis SA, et al. Sustainable applications of rice feedstock in agro-environmental and construction sectors: A global perspective. *Renew Sust Energ Rev.* 2022;153:111791.
- Chauhan BS, Abugho SB. Effects of water regime, nitrogen fertilization, and rice plant density on growth and reproduction of lowland weed *Echinochloa crus-galli*. *Crop Prot.* 2013;54:142–147.
- Zheng H, Chen Y, Chen Q, Li B, Zhang Y, Jia W, Mo W, Tang Q. High-density planting with lower nitrogen application

- increased early rice production in a double-season rice system. *Agron J.* 2020;112:205–214.
4. Blanc E, Strobl E. Assessing the impact of typhoons on rice production in the Philippines. *J Appl Meteorol Climatol.* 2016;55:993–1007.
 5. Liu L, Lu H, Li Y, Cao Z. High-throughput Rice density estimation from transplantation to Tillering stages using deep networks. *Plant Phenomics.* 2020;2020:1375957.
 6. Madec S, Jin X, Lu H, De Solan B, Liu S, Duyme F, Heritier H, Baret F. Ear density estimation from high resolution RGB imagery using deep learning technique. *Agric For Meteorol.* 2019;264:225–234.
 7. Varela S, Dhodda PR, Hsu WH, Prasad P, Assefa Y, Peralta NR, Griffin T, Sharda A, Ferguson A, Ciampitti I. Early-season stand count determination in corn via integration of imagery from unmanned aerial systems (UAS) and supervised learning techniques. *Remote Sens.* 2018;10:343.
 8. J. G. A. Barbedo, Method for automatic counting root nodules using digital images, Paper presented at: Proceedings of the 2012 12th International Conference on Computational Science and Its Applications; 2012 June 18–21; Salvador, Brazil; pp. 159–161.
 9. Rustia DJA, Lin CE, Chung J-Y, Zhuang Y-J, Hsu J-C, Lin T-T. Application of an image and environmental sensor network for automated greenhouse insect pest monitoring. *J Asia Pac Entomol.* 2020;23:17–28.
 10. Jin X, Madec S, Dutartre D, de Solan B, Comar A, Baret F. High-throughput measurements of stem characteristics to estimate ear density and above-ground biomass. *Plant Phenomics.* 2019;2019:4820305.
 11. Fernandez-Gallego JA, Buchailot M, Aparicio Gutiérrez N, Nieto-Taladriz MT, Araus JL, Kefauver SC. Automatic wheat ear counting using thermal imagery. *Remote Sens.* 2019;11:751.
 12. 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;3:211–219.
 13. Bai X, Cao Z, Zhao L, Zhang J, Lv C, Li C, Xie J. Rice heading stage automatic observation by multi-classifier cascade based rice spike detection method. *Agric For Meteorol.* 2018;259:260–270.
 14. Tan S, Ma X, Mai Z, Qi L, Wang Y. Segmentation and counting algorithm for touching hybrid rice grains. *Comput Electron Agric.* 2019;162:493–504.
 15. Hasan MM, Chopin JP, Laga H, Miklavcic SJ. Detection and analysis of wheat spikes using convolutional neural networks. *Plant Methods.* 2018;14:100.
 16. Lu H, Cao Z, Xiao Y, Zhuang B, Shen C. TasselNet: Counting maize tassels in the wild via local counts regression network. *Plant Methods.* 2017;13:79.
 17. Li Y, Zhang X, Chen D. CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes. Paper presented at: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018; pp. 1091–1100.
 18. C. Zhang, H. Li, X. Wang, X. Yang, Cross-scene crowd counting via deep convolutional neural networks. Paper presented at: Proceedings of the 2015 IEEE conference on computer vision and pattern recognition (CVPR); 2015 June 7–12; Boston, MA; pp. 833–841.
 19. Wang Q, Gao J, Lin W, Li X. NWPU-crowd: A large-scale benchmark for crowd counting and localization. *IEEE Trans Pattern Anal Mach Intell.* 2020;43:2141–2149.
 20. Zhao Q, Xiao J, Wang Z, Ma X, Wang M, Satoh S. Vehicle counting in very low-resolution aerial images via cross-resolution spatial consistency and Intra-resolution time continuity. *IEEE Trans Geosci Remote Sens.* 2022;60:4706813.
 21. Erturk IF, Alper Demir M, Akar Ve G, Kulah H. Automatic Cell Counting From Microchannel Images. Paper presented at: Proceedings of the 30th Signal Processing and Communications Applications Conference, SIU 2022; 2022 May 15–18; Safranbolu, Turkey.
 22. Selinummi J, Seppälä J, Yli-Harja O, Puhakka JA. Software for quantification of labeled bacteria from digital microscope images by automated image analysis. *BioTechniques.* 2005;39:859–863.
 23. Dollár P, Wojek C, Schiele B, Perona P. Pedestrian detection: An evaluation of the state of the art. *IEEE Trans Pattern Anal Mach Intell.* 2011;34:743–761.
 24. Viola P, Jones MJ. Robust real-time face detection. *Int J Comput Vis.* 2004;57:137–154.
 25. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, Microsoft coco: Common objects in context. In: Dollár P, Zitnick CL. In, Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. *European conference on computer vision*; Springer, Cham; 2014; pp. 740–755.
 26. Wang X, Chen J, Wang Z, Liu W, Satoh S i, Liang C, Lin C-W. When pedestrian detection meets nighttime surveillance: A new benchmark. *Image.* 2020;20000:40000.
 27. Lempitsky V, Zisserman A. Learning to count objects in images. *Adv Neural Inf Proces Syst.* 2010;23:1324–1332.
 28. Liu L, Qiu Z, Li G, Liu S, Ouyang W, Lin L. Crowd Counting With Deep Structured Scale Integration Network. Paper presented at: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 2020.
 29. Liu L, Wang H, Li G, Ouyang W, Lin L. Crowd counting using deep recurrent spatial-aware network. Paper presented at: Proceedings of the 27th International Joint Conference on Artificial Intelligence; 2018.
 30. Zhang Y, Zhou D, Chen S, Gao S, Ma Y. Single-image crowd counting via multi-column convolutional neural network. Paper presented at: Proceedings of the 2016 IEEE conference on computer vision and pattern recognition (CVPR); 2016 June 27–30; Las Vegas, NV, USA; pp. 589–597.
 31. Wang R, Alotaibi R, Alzahrani B, Mahmood A, Wu G, Xia H, Alshehri A, Aldaheri S. AAC: Automatic augmentation for crowd counting. *Neurocomputing.* 2022;500:90–98.
 32. Cao X, Wang Z, Zhao Y, Su F. Scale aggregation network for accurate and efficient crowd counting. Paper presented at: Proceedings of the 2019 European Conference on Computer Vision (ECCV); 2019; pp. 734–750.
 33. Ma Z, Wei X, Hong X, Gong Y. Bayesian loss for crowd count estimation with point supervision. Paper presented at: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2019; pp. 6142–6151.
 34. Liu Y, Wen Q, Chen H, Liu W, Qin J, Han G, He S. Crowd counting via cross-stage refinement networks. *IEEE Trans Image Process.* 2020;29:6800–6812.
 35. Toha TR, Al-Nabhan NA, Salim SI, Rahaman M, Kamal U, Islam ABMAA. *LC-Net: Localized Counting Network for extremely dense crowds.* *Appl Soft Comput.* 2022;123:108930.
 36. Liu Y, Wang Z, Shi M. Discovering regression-detection bi-knowledge transfer for unsupervised cross-domain crowd counting. *Neurocomputing.* 2022;494:418–431.

37. Thanasutives P, Fukui K-I, Numao M, Kijirikul B. Encoder-decoder based convolutional neural networks with multi-scale-aware modules for crowd counting. Paper presented at: Proceedings of the 25th International Conference on Pattern Recognition, ICPR 2020; 2021 January 10–15; Milan, Italy: Virtual; pp. 2382–2389.
38. Roth L, Barendregt C, Béatrix C-A, Hund A, Walter A. High-throughput field phenotyping of soybean: Spotting an ideotype. *Remote Sens Environ*. 2021;112797.
39. Chen B, Yan Z, Li K, Li P, Wang B, Zuo W, Zhang L. Variational attention: Propagating domain-specific knowledge for multi-domain learning in crowd counting. arXiv. 2021. <https://doi.org/10.48550/arXiv.2108.08023>.
40. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *Adv Neural Inf Process Syst*. 2015;28:91–99.
41. Xiong H, Cao Z, Lu H, Madec S, Liu L, Shen C. TasselNetv2: In-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods*. 2019;15:150.
42. Cai E, Baireddy S, Yang C, Delp EJ, Crawford M. Panicle counting in UAV images for estimating flowering time in sorghum. Paper presented at: Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS; 2021 July 11–16; Brussels, Belgium; pp. 6280–6283.
43. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv. 2014. <https://doi.org/10.48550/arXiv.1409.1556>.
44. Shi Z, Mettes P, Snoek CG. Counting with focus for free. Paper presented at: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2019; pp. 4200–4209.
45. Liang D, Xu W, Zhu Y, Zhou Y. Focal inverse distance transform maps for crowd localization and counting in dense crowd. arXiv. 2021. <https://doi.org/10.48550/arXiv.2102.07925>.
46. Liu L, Lu H, Xiong H, Xian K, Shen C. Counting objects by blockwise classification. *IEEE Trans Circuits Syst Video Technol*. 2020;30:3513–3527.