Computers and Electronics in Agriculture, Volume 173, June 2020, Article number 105375
DOI:10.1016/j.compag.2020.105375

Computers and Electronics in Agriculture

Manuscript Details

Manuscript number	COMPAG_2019_1895_R2
Title	MHW-PD: a robust rice panicles counting algorithm based on deep learning and multi-scale hybrid window
Article type	Research Paper

Abstract

In-field assessment of rice panicle yields accurately and automatically has been one of the key ways to realize high- throughput rice breeding in the modern smart farming. However, practical rice fields normally consist of many different, often very small sizes of panicles, particularly when large numbers of panicles are captured in the imagery. In these cases, the integrity of panicle feature is difficult to extract due to the limited panicle original information and substantial clutters caused by heavily compacted leaves and stems, which results in poor counting efficacy. In this paper, we propose a simple, yet effective method termed as Multi-Scale Hybrid Window Panicle Detect (MHW-PD), which allows the identification and counting of rice panicles robustly independent of the panicle number (density) in the scene. On the basis of quantifying and analyzing the relationship among the receptive field, the size of input image and the average dimensions of panicles, the MHW-PD gives dynamic strategies for choosing the appropriate feature learning network and constructing adaptive multi-scale hybrid window (MHW), which maximizes the richness of panicle feature. Besides, a fusion algorithm is involved to remove the repeated counting of the broken panicles to get the final panicle number. With extensive experimental results, the MHW-PD has achieved ~87% of panicle counting accuracy; and the counting accuracy just decreases by ~8% when the number of panicles per image increases from 0 to 80, which shows better in stability than all the competing methods adopted in this work. The MHW-PD is demonstrated qualitatively and quantitatively that is able to deal with high density of panicles.

Keywords Rice; Panicle counting; Deep learning; Multi-Scale Hybrid window; Faster- RCNN;

Corresponding Author Haiyan Jiang

Corresponding Author's Institution Nanjing Agricultural University

Order of Authors Xu Can, Haiyan Jiang, Peter Yuen, Zaki Zaki, Chen Yao**Corresponding Author's Institution**

Highlights

（1） A counting algorithm is developed for in-field rice panicles with high density.

（2） The appropriate CNN is chosen by analyzing receptive field and panicle size.

（3） A MHW is calculated quantitatively to maximize the richness of panicle feature.

（4） A fusion module is involved to remove the repeated counting of broken panicle.

（5） Stability and robustness of MHW-PD is demonstrated by several experiments.

MHW-PD: a robust rice panicles counting algorithm based on deep learning and multi-scale hybrid window

Xu Can[1], Jiang Haiyan[1, 2*], Peter Yuen[3], Zaki Ahmad Khan[1], Chen Yao[1]

1 College of Information science & Technology, Nanjing Agricultural University
Nanjing 210095, Jiangsu, China

2 National Engineering & Technology Center for Information Agricultural,
Nanjing Agricultural University Nanjing 210095, Jiangsu, China

3 Electro-Optics & Remote Sensing, Centre for Electronics Warfare, Information &
Cyber (CEWIC), Cranfield University, Swindon, U.K

Abstract

In-field assessment of rice panicle yields accurately and automatically has been one of the key ways to realize high-throughput rice breeding in the modern smart farming. However, practical rice fields normally consist of many different, often very small sizes of panicles, particularly when large numbers of panicles are captured in the imagery. In these cases, the integrity of panicle feature is difficult to extract due to the limited panicle original information and substantial clutters caused by heavily compacted leaves and stems, which results in poor counting efficacy. In this paper, we propose a simple, yet effective method termed as Multi-Scale Hybrid Window Panicle Detect (MHW-PD), which focuses on enhance the panicle features to detect and count the large number of small-sized rice panicles in the in-field scene. On the basis of

Published by Elsevier. This is the Author Accepted Manuscript issued with: Creative Commons Attribution Non-Commercial No Derivatives License (CC:BY:NC:ND 4.0).
The final published version (version of record) is available online at DOI:10.1016/j.compag.2020.105375. Please refer to any applicable publisher terms of use.

quantifying and analyzing the relationship among the receptive field, the size of input image and the average dimensions of panicles, the MHW-PD gives dynamic strategies for choosing the appropriate feature learning network and constructing adaptive multi-scale hybrid window (MHW), which maximizes the richness of panicle feature. Besides, a fusion algorithm is involved to remove the repeated counting of the broken panicles to get the final panicle number. With extensive experimental results, the MHW-PD has achieved ~87% of panicle counting accuracy; and the counting accuracy just decreases by ~8% when the number of panicles per image increases from 0 to 80, which shows better in stability than all the competing methods adopted in this work. The MHW-PD is demonstrated qualitatively and quantitatively that is able to deal with high density of panicles.

Keywords： Rice; Panicle counting; Deep learning; Multi-Scale Hybrid window; Faster-RCNN;

1 Introduction

The main diet of the population in Asia is predominately rice, thus the monitoring of rice yield accurately is crucially important to the growers for the prediction of harvest and the development of strategic growth plan. The yield of cereal crops, such as rice, is largely determined by three agronomic indicators: the kernel number, the seed setting rate and the 1000-grain weight(Slafer et al., 2014). Previous researches (Ferrante et al., 2017; Jin et al., 2017)have shown that the number of kernels per unit area is the most relevant agronomic traits to grain yield. However, this number of grains per unit area not only relates to the seed setting rate, but also it is strongly dependent on the number of panicle per unit area. Therefore, it is desirable for the breeders to obtain the number of panicles per unit area quickly and accurately. At present, this is often achieved through counting manually in most rice cultivation or breeding research, which costs huge amount of time and labor. Furthermore, due to the great morphological similarity between different plants in the field, and also the subjectivity in individual observers, it is very error-prone for counting rice panicles manually particularly in large-scale production scenarios. Therefore, a fast and relatively accurate automatic counting method is needed: for both production as well as scientific research needs such as phenotyping work.

Automatic counting method based on machine vision technology is considered to be an effective alternative to manual counting, and successful precedents such as the counting of plant leaves(Aich et al., 2017; Barré et al., 2017; Dobrescu et al., 2017; Giuffrida et al., 2016) and fruits(Maldonado Jr et al., 2016; Mussadiq et al., 2015; Stein et al., 2016) have been reported. The effectiveness of this automatic counting method is heavily dependent on the ability of the machine to recognize the targets. In terms of automatic counting of rice panicles, the existing panicle recognition methods can be divided into two main categories: the segmentation technique which bases on colour and/or textural features and the candidate region-based classification methods. Panicle segmentation method (Cointault et al., 2008; Pound et al., 2017) extracts the colour or texture of the panicle, and the rice panicles are segmented from the background before they are counted. Zhou et al. (Zhou et al., 2018) employed principal component analysis to extract representative features of wheat from RGB images such as colour, texture and edge for wheat panicle segmentation, and ~80% of count accuracy by using a trained dual support vector machine has been reported. Fernandez et al.(Fernandez-Gallego et al., 2018) proposed a fast low-cost wheat panicle segmentation algorithm which uses Laplacian, Median and Maxima (LMM) filters to remove clutter backgrounds and had achieved good panicle counting results. The panicle segmentation method is of a low computational complexity algorithm but the result is sensitive to the illumination conditions of the imagery data (Guo et al., 2015).

The candidate region classification is the method that clusters features over the spatial domain. The key of the algorithm is the generation of candidate regions, through features such as color or texture and the candidate regions are subsequently formed by using the hysteresis threshold of the I2 color plane (Duan et al., 2015) and the Laws texture energy over the input image(Qiongyan et al., 2017). This method eliminates more of the clutter background than that of the segmentation approach, hence it achieves better counting accuracy to some extents. Alternative approach that utilizes superpixel technique for improving the quality of the candidate region generation through better preservation of boundary information and to reduce boundary adhesions, has been widely explored(Lu et al., 2016). Some authors employed simple linear iterative clustering for the generation of superpixel and then classified the region candidates using convolutional neural network (Xiong et al., 2017) or classifier trained based on colour feature(Du et al., 2019). Further study using more effective segmentation method that utilize superpixel in different scales and couple with a trained linear regression model for counting different varieties of rice panicles has also been reported(Olsen et al., 2018).

The recent work had made the better use of the powerful feature learning capabilities of the CNN (Convolutional Neural Network, CNN). More sophisticated feature learning that utilizes a full convolution network for counting field wheat spikelet have reported a counting accuracy of about 86%(Alkhudaydi et al., 2019). Other method(Hasan et al., 2018) used the R-CNN(Girshick et al., 2014) for wheat panicle identification counting, for the object detection algorithm focus on solving the composite problem of classification and localization. The latest work(Madec et al., 2019) introduced the Faster-RCNN(Ren et al., 2015) method into wheat panicle counting and got a 91% counting accuracy. For the rice panicles we focus on, they will droop due to their self-weight on the maturity-stage, which means the crowded panicles cram together with leaves and even occluded by leaves locally. Meanwhile, the size of the panicles in the image tends to reduce when high density of panicles, e.g. >50 panicles/image, is captured by the camera. In this case, the very limited information (color/textural/spatial) of the panicle, which is embedded closely in substantial amount of clutter background, greatly reduces the feature learning efficiency of the existing object detection algorithms(He et al., 2015; Liu et al., 2016;

Redmon et al., 2016; Redmon et al., 2017) and inevitably resulting in large counting error. Thus, there is a real need to develop a new auto approach to allow a rapid counting of the scene with large number of small-sized rice panicles per image.

2 Principles and designs of the MHW-PD for panicle counting

2.1 Analysis of application of Faster-RCNN

Faster-RCNN is one of the representative detection algorithms based on regions(Han et al., 2018), which features the strengths of algorithmic structures like that of the RCNN(Girshick et al., 2014), the SPP-Net(He et al., 2015) and the Fast-RCNN (Girshick, 2015). As shown in Figure 1, Faster-RCNN has capabilities such as feature learning, candidate region generation, target classification and positional frame generation. When Faster-RCNN learns feature based on a CNN, one important point is the receptive field, which is defined by the region in the input space that corresponds to any pixel on a particular CNN's feature map. In the circumstances when train a model to make classification and location, the receptive field of every position on the feature map have to span over all the anchors that the target/object represents. Otherwise the feature vectors of the anchors will not have enough information to make predictions, leading some objects missed by detection model. This is particular true when the target in question is relatively small in physical size in comparison to that of the background objects, for example, the small-sized rice panicles here in our scenario.

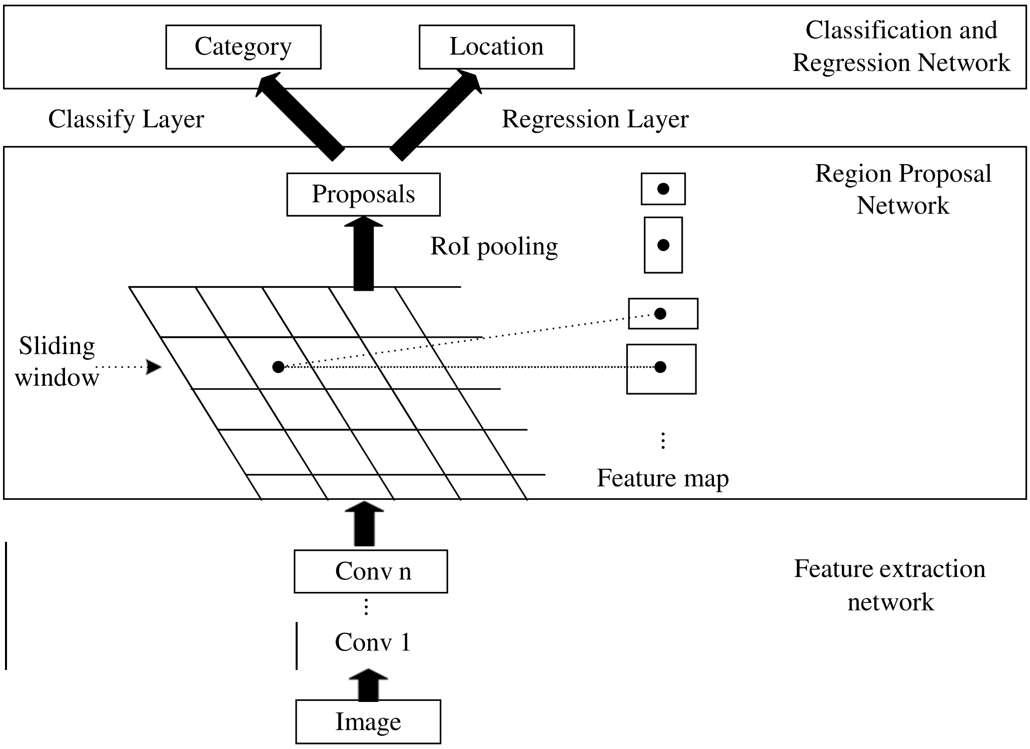


Fig. 1 Outlines the schematic layout of the Faster-RCNN network

2.2 Overall design of the MHW-PD

The objective of the paper is to report an adaptive multi-scale hybrid window (MHW) pre-processing technique to enhance the signal to noise ratio of the panicle features in the input image, and to couple it with Faster-RCNN network to achieve robust counting accuracy for the large number of small-sized panicles in image. For the problem of information loss in the process of learning small-sized panicles feature, we firstly designed a dynamic mechanism for selecting feature learning network, which is based on the relationship between the size of the rice panicle and the dimension of the receptive field. Secondly, we dynamically calculated the hybrid windows in different scales by partitioning the image into subsections by quantifying the relationship between the input image size and the feature learning network parameters. This helps to reduce the background complexity by suppressing the clutter background particularly when the number of rice panicles increases. The framework of MHW-PD (Figure 2) consists of the following work flow: a) select feature learning network dynamically; b) calculate the structure of the hybrid windows; c) train the automatic rice panicle counting model based on the Faster-RCNN; d) fuse the same rice panicle which has been partitioned into several entities to remove the multiple counting; e) output the final number of rice panicles count of the test image.

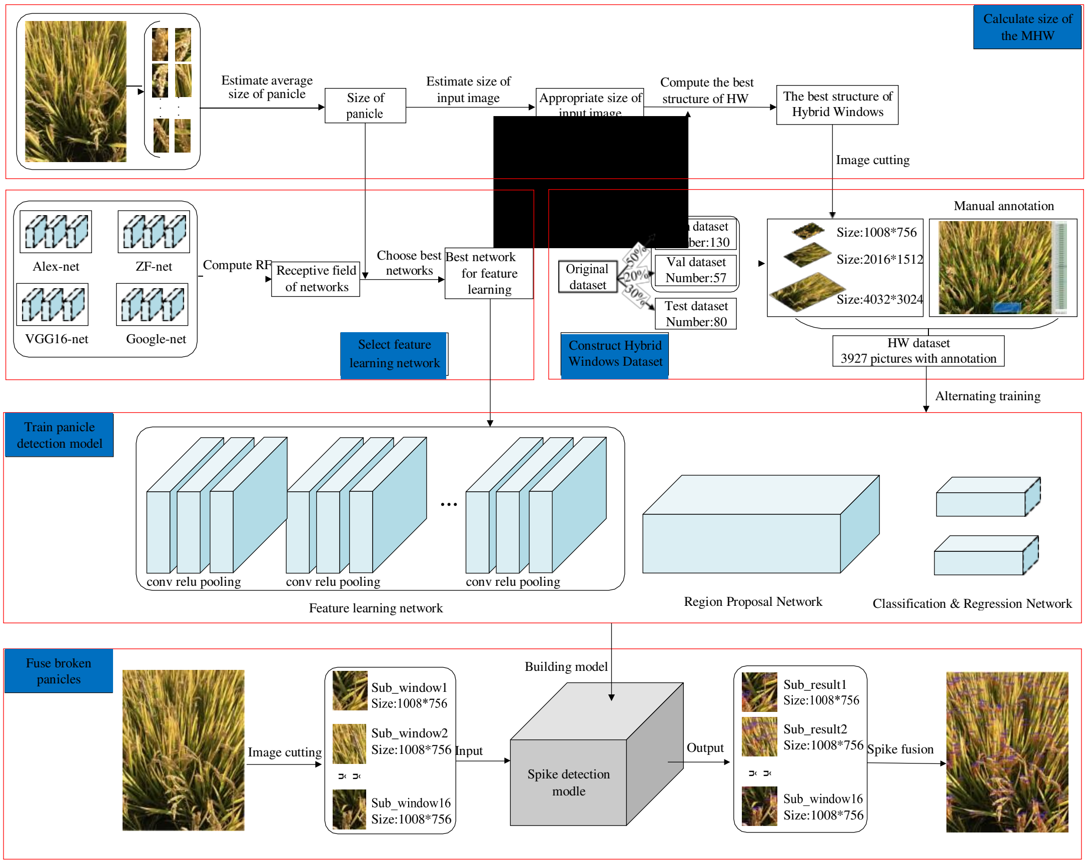

Calculate size of the MHW
Estimate average size of panicle
Size of panicle
Estimate size of input image
Appropriate size of input image
Compute the best structure of HW
The best structure of Hybrid Windows
Image cutting
[redacted]
Alex-net
ZF-net
VGG16-net
Google-net
Compute RF
Receptive field of networks
Choose best networks
Best network for feature learning
Select feature learning network
Original dataset
50%
20%
30%
dataset
ber:130
Val dataset Number:57
Test dataset Number:80
Construct Hybrid Windows Dataset
Size:1008*756
Size:2016*1512
Size:4032*3024
Manual annotation
HW dataset 3927 pictures with annotation
Alternating training
Train panicle detection model
conv relu pooling
conv relu pooling
...
conv relu pooling
Feature learning network
Region Proposal Network
Classification & Regression Network
Fuse broken panicles
Image cutting
Sub_window1 Size:1008*756
Sub_window2 Size:1008*756
Sub_window16 Size:1008*756
Input
Building model
Spike detection modle
Output
Sub_result1 Size:1008*756
Sub_result2 Size:1008*756
Sub_window16 Size:1008*756
Spike fusion

Fig. 2 The schematic layout of the MHW-PD for the robust detection and counting of rice panicles

2.2.1 Selection of the feature learning network

Feature learning is the technique that iteratively abstracts the semantic and position information of the target from the image data and converts them into feature maps. The extracted features are dependent on the layer property and thus the receptive field of a layer can be given by equation (1) (Ren et al., 2018).

$$S_{R}(t) = (S_{RF}(t-1) - 1)N_{s}(t) + S_{f}(t) \qquad (1)$$

Where $S_{RF}(t)$ and $N_{s}(t)$ are the receptive field size and the step size of the t^{th} convolution layer, and $S_{f}(t)$ is the size of filter of the t^{th} convolution layer. The ideal dimension of the receptive field is a delicate balance between clutter noise and the integrity of the extracted feature. In the present Faster-RCNN experiment, the relationship between the receptive field of the feature learning network and the object/target has been set as in equation (2):

$$\frac{S_{RF}(t)}{S_{ob}(h_{obj},w_{obj})} \approx 1 \qquad (2)$$

Where $S_{obj}(h_{obj},w_{obj})$ represents the size of the object to be detected, and h_{obj} and w_{obj} respectively represent the length and width of the minimum circumscribed rectangle of the target to be detected. According to equation (2), the ideal dimension of the receptive field is ideally to be about the same as that of the targets (i.e. the rice panicles). According to equation (1), the dimensions of the receptive field of the last convolutional layer of the most popular networks, such as the Alex-Net(Krizhevsky et al., 2012), ZF-Net(Zeiler et al., 2014), VGG16-Net(Simonyan et al., 2014) and Google-Net (Szegedy et al., 2015) are tabulated in Table 1. The average sizes (length × width) of rice panicles in the image data that have been selected for this work is about 260×180 pixels. Thus the VGG16 network which features a receptive field of 212×212 may present a closer match to the average panicle dimensions of the data that utilized in this work than other networks. Therefore, the VGG16 network and the classification layer have been selected as the feature learning network in this work.

Table 1. Tabulated the receptive field of different nets for the 800×600 pixels input image

Net name	Reception field of the last layers	S_{RF}/S_{obj}
ZF-Net	139×139	0.41
Alex-Net	195×195	0.81
VGG16-Net	**212×212**	**0.96**
Google-Net	224×224	1.07

2.2.2 Design of the Multi-scale Hybrid Window (MHW) Structure

Targets are generally regarded as small when they are less than 32×32 pixels or when their length and width are smaller than a tenth of that of the image where they are contained. The construction of a multi-scale hybrid window by partitioning a picture into sub-images will tend to enhance the proportions of the object features with respected to the background within the sub-image, especially when the objects are small. The richer of the target feature will enhance the discrimination ability of the RPN to identify/propose the anchors to be foreground or background thereby improving the detection efficiency. The design of the MHW structure involves the considerations of: i) the various sizes of hybrid windows needed for a given input image, ii) the number of window layers and iii) the selection of layers that are the most suitable to the ranges of various input image sizes.

The largest hybrid window that can theoretically be constructed in each layer of the n-layer feature learning network can be given by equation (3):

$$\begin{cases} A_{H(t)} = \dfrac{H + S(t) + 2 * S_p(t)}{N_s(t)} + 1 \\ A_{W(t)} = \dfrac{W + S(t) + 2 * S_p(t)}{N_s(t)} + 1 \end{cases} \quad t = 1,2,\cdots n \qquad (3)$$

Where $A_{H(t)}$ and $A_{W(t)}$ represent the length and width of the t^{th} feature map of the feature learning network respectively, H and ◆ represent the length and width of the original raw image respectively, and n is maximum number of layers in the feature learning network. $N_s(t)$ is the step size of the t^{th} convolution layer, and S_f (t) is the size of the filter of t^{th} convolution layer , and $S_p(t)$ is the expansion of the t^{th} convolution layer. The optimal input image size is given in equation (4):

$$\begin{cases} h_{in} = \frac{h_{obj}}{T_1} & 0.1 < T_1 < 1 \\ w_{in} = \frac{w_{obj}}{T_2} & 0.1 < T_2 < 1 \end{cases} \qquad (4)$$

where h_{in} and w_{in} represent the length and width of the optimum input image dimensions; h_{obj} and w_{obj} represent the length and width of the smallest rectangle of the object to be detected respectively; T_1 and T_2 represent the ratio of the length and width of the object respected to the dimensions of the input image respectively. The optimal dimensions of the multi-scale hybrid window structure can then be deduced as shown in equation (5):

$$\begin{cases} h_{HW}(i) = A_{H(t)} & A_{H(t)} \in (h_{min}, h_{max}) \\ w_{HW}(i) = A_{W(t)} & A_{W(t)} \in (w_{min}, w_{max}) \end{cases} \quad i = 1,2,...,p \; \& \; t = 1,2,...,n \; (5)$$

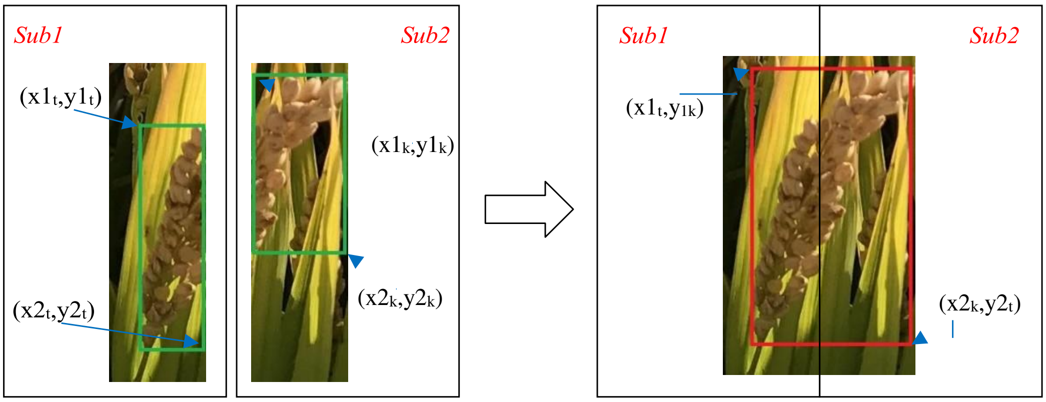
When there are p layers of multi-scale hybrid windows, $h_{HW}(i)$ and $w_{HW}(i)$ represent the optimal length and width of the i^{th} layer respectively; (h_{min}, h_{max}) and (w_{min}, w_{max}) represent the possible range of the optimal length and width of the input images that will produce the best learning and classification performances.

2.2.3 MHW fusion

One of the drawbacks for partitioning the input image into sub-images is the panicle may be unintentionally cut into several parts in different sub-images. To eliminate the repeated counting of the same panicle that resides in various sub-images during the prediction stage, a fusion algorithm is designed to detect the occurrence of the panicle that has been subdivided into parts. A simple way to correct this unintentional partition of the target object is to check the vicinity of all the predicted boxes. A simple spatial distance monitor algorithm has been implemented to check the vicinity of all the predicted location boxes: if two predicted boxes are adjacent or very close to each other while their sum of size (height× length) is close to the average panicle size, e.g. when they are say <10 pixels apart and sum is between 130×90 pixels and 390×270 pixels (from 1/2 to the 3/2 of the average panicle size), the boxes pairs will be merged into one by adopting the largest vertices of the corner coordinate as illustrate in Table 2 and Figure 3.

Table 2. The Mini-Code of the Fusion Algorithm for recombining dissected rice panicles

```
Input: $(x1_n, y1_n, x2_n, y2_n)$: the coordinates of the left upper and right lower vertices of the panicle detected in sub-windows
Output: $(x1'_m, y1'_m, x2'_m, y2'_m)$: the coordinates of prediction boxes fused
For($k = 1; k \leq n; k{+}{+}$)
    For($t = 1; t \leq n; t{+}{+}$)
If$(|x1_k - x2_t| < 10 \,\&\&\, |y1_k - y2_t| < 2h) \;||\; (|y1_k - y2_t| < 10 \,\&\&\, |x1_k - x2_t| < 2w) ||$
$(90 < (|y1_k - y2_k| + |y1_t - y2_t|) < 270) || (130 < (|x1_k - x2_k| + |x1_t - x2_t|) < 390)$
$(x1'_m, 1'_m, x2'_m, y2'_m) = (min(x1_k, x1_t), min(y1_k, y1_t), max(x2_k, x2_t), max(y2_k, y2_t))$
m++;
```



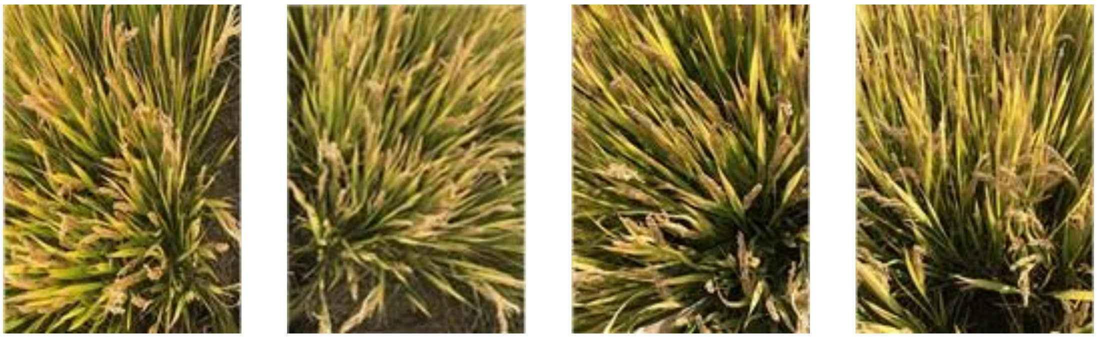
Result before fusing Result after fusing

Fig. 3 Illustrated the fusion of vertically dissected rice panicles

3 Construction of dataset and model

3.1 Image data acquisition

The rice variety chosen is ‘Nanjing46’ and all images were acquired in Nanjing, Jiangsu Province, China. The field consisted of a widely cultivated rice variety with planting scheme of 3-5 seedlings per hole and 30×12 cm spacing between plants. The imaging was performed using random viewing angles at objective distances of ~60 cm towards the rice plant using a Canon EOS 70D camera with resolutions of 4032×3024 pixels. The images contain various numbers of small-sized panicles ranging from 50-90 per image, which have shown the complex interaction relationship between different rice plants. As shown in Figure 4, there were 141 images and 126 images acquired under normal (9:00 am) and strong (2:00 pm) illumination conditions respectively. The picture of the rice panicle appears in yellow color, and the full image is filled with large number of light greenish rice leaves together with shadows due to the oblique illumination angle and partially due to the leaf occlusions. The average dimensions (length × width) of panicles in the image data is about 260×180 pixels after selecting 200 independent panicles randomly and calculating the average size (length× width) of their minimum circumscribed rectangles.



(a) Normal illumination (b) Intense illumination

Fig. 4 Sample of have been taken under different viewing angles and illumination conditions

3.2 Multi-scale hybrid window dataset construction

3.2.1 Calculate the structure of the MHW

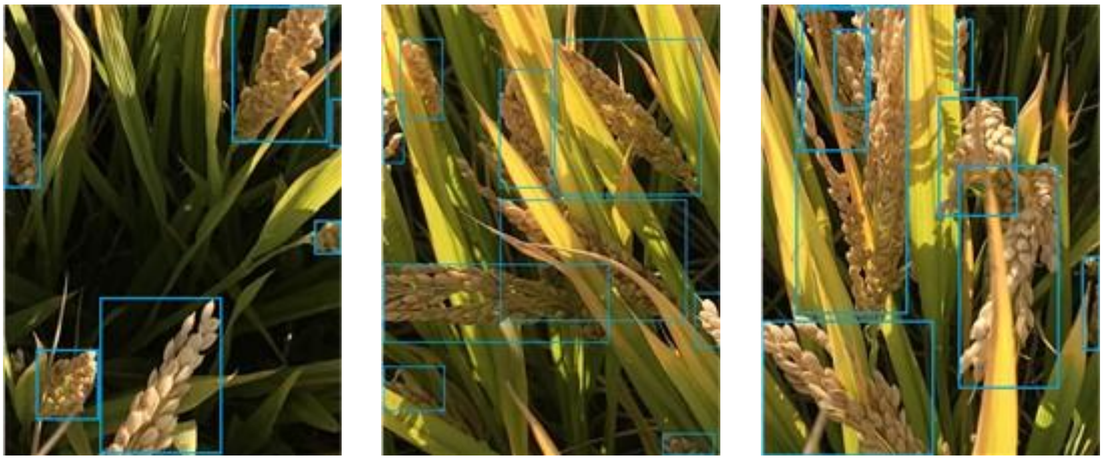
The average size of rice panicle in the data set is about 260×180 pixels which is less than one-tenth of the image size with occupancy about 0.4% of the full picture. This gives the most appropriate dimensions of the input images ranging between 260×180 pixels and 2600×1800 pixels as according to equation 4. As mentioned in section 2.2.1, the VGG16 network has been chosen because it is more effective to learn the features of objects particularly those with physical dimensions like that in our data set. The optimal dimensions of each layer of the multi-scale hybrid window can be assessed through equation 5, which gives the topmost 3 layers to be ideally having 2016×1512 pixels, 1008×756 pixels and 504×378 pixels respectively. Although theoretically the more of the network layers the richer that the features can be learned, however, it is a balance between performance and computational complexity. When the layer with input images of sizes 504×378 pixels, it contains utmost only a few rice panicles which may not be economical in view of the amount of the extra computational and labeling workload involved. Hence, only the two extra topmost layers have been utilized in this work.

3.2.2 Formation of the MHW dataset

Among the 267 rice pictures collected, 130 of those (~50%) were randomly selected as the training set, and 57 pictures (~20%) were used as the validation set and the remaining 80 pictures (~30%) was used as the test data set. There is no data overlap among the training, validation and test sets. For the model training, we only construct the MHW dataset for the training set and the validation set. Conventional subsampling using a fixed scheme for altering image dimensions(Ghiasi et al., 2016) may not be desirable when the problem in question consists of targets in various sizes. Here, for each image in the training and validation data set, the raw image at 4032×3024 pixels resolution (hereafter referred as R1) is divided along the length and width in 4 and 2 equal parts respectively to form a four and sixteen units of sub-images respectively. Then these 4 sub-images at 2016×1512 pixels resolution (hereafter referred as R2), and 16 at 1008×756 pixels resolution (hereafter referred as R3) together with the raw image are collectively termed as multi-scale hybrid windows (MHW). Alternative MHW partition schemes which select different layers to train the model (such as R1 & R2, R2 & R3) have also been utilized in the experiment.

3.2.3 Target labeling schemes

The labeling of MHW images for training and validation dataset has been performed manually by recording the coordinates of the minimum circumscribed rectangle of the panicle, using the annotation software named ‘LabelImg’. In the case of panicles that have been partitioned into several parts, all parts are labeled as independent rice panicles. In the case of the rice panicles that are occluded by leaves, only the exposed parts are labeled as independent panicles. For panicles that are overlapping to each other, the front panicles are labeled as independent target while the rear part will be marked only if they are visible. Figure 5 shows some examples of annotation schemes that have been adopted in this work.



(a) Independent panicles **(b) Panicles covered by leaves** **(c) Overlapping panicles**

Fig. 5 Examples of manual annotations of panicles

3.3 Configuration of test dataset for experiments

The remaining 80 raw pictures at resolution of 4032×3024 pixels (i.e. at ‘R1’) in the section 3.2.2 was termed as the ‘Dataset_test’ in this paper. Each image in the Dataset_test was then partitioned equally into 16 sub-images giving a total of 1280 pictures at 1008×756 pixels (i.e. at ‘R3’), which is collectively referred as

‘Dataset_test_1’. The number of panicles in the picture of Dataset_test_1 ranges from 0-20. By merging two of the adjacent neighboring sub-images of the 16 partitioned images of the raw pictures produces 4×80 of new images at resolution of 2016×1512 (i.e. at ‘R2’). All these sub-images were then sorted into another two data sets (Dataset_test_2 and Dataset_test_3) as according to the number of panicles in the imagery as illustrated in Table 3. These 3 data sets provide a range of different number (and hence different sizes) of panicles as targets for the classifiers to detect (and count) under various degrees of background cluttering.

Images of rice panicles collected in real fields are normally exhibit blurring and discoloring due to the complicated environment in the rice field. Imaging such complex scene by using limited depth of view optical systems under various illumination geometries, will result in some objects that are out-of-focus and/or discolored due to the variable irradiance and also targets at various depth across the scene. As mentioned image data had been collected at two different solar irradiances: one at 9 am (thereafter referred as ‘normal’ illumination) and also at 2 pm (thereafter referred as ‘intense’ illumination). Another data set, termed as the ‘Dataset_test_4’ which is organized in four categories of a) in-focus & normal illumination, b) in-focus & intense illumination, c) blurry & normal illumination and d) blurry & intense illumination.

Table 3. Description of the datasets that have been employed in this study

Name of the Datasets	Composition of Dataset		
	Category	Size of Image Pictures in Dataset	Number of Pictures in Dataset

Dataset_test	Original test images	4032×3024	80
Dataset_test_1	Cut in 16 equal parts	1008×756	1280
Dataset_test_2	0~10(panicle number in sub-window image)	1008×756	205
	11~20(panicle number in sub-window image)	1008×756	108
	21~30(panicle number in sub-window image)	1008×1512	70
	31~40(panicle number in sub-window image)	1008×1512	41
Dataset_test_3	41~50(panicle number in image)	4032×3024	22
	51~60(panicle number in image)	4032×3024	22
	61~70(panicle number in image)	4032×3024	16
	71~80(panicle number in image)	4032×3024	9
	81~90(panicle number in image)	4032×3024	7
Dataset_test_4	In-focused & Normal illumination	1008×756	67
	In-focused & Intense illumination	1008×756	72
	Blurry & Normal illumination	1008×756	62
	Blurry & Intense illumination	1008×756	74

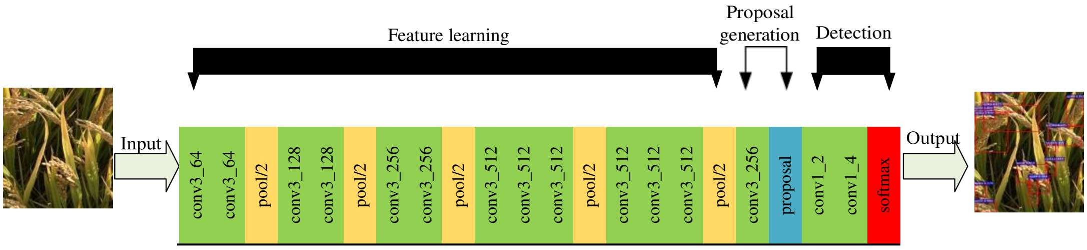
3.4 Construct the automatic rice panicle counting model

3.4.1 Computational hardware and platform

All processing performed in this work was carried out by the AMAX's PSC-HB1X deep learning workstation which consisted of an Intel(R) E5-2600 v3 CPU with clock speed of 2.1GHZ, 128GB DRAM, 1TB hard disk and with a GeForce GTX Titan X graphics card. The operating environment was Ubuntu 16.0.4, Caffe, Python 2.7.

3.4.2 Model training

The proposed MHW-PD network consists of three parts: the feature learning network, the candidate region generation network and the detection network (Figure 6). The feature learning network utilizes the VGG16 network but without its classification layer. The region generation network traverses the feature map (stride=1) with a 3×3 convolution kernel and a 9 candidate region with three aspect ratios of 1:1, 2:1 and 1:2 to indicate the high probability of target (panicle) presence is generated by the proposal layer. The detection network uses a convolution operation with a convolution kernel size of 1×1 and a sliding step size of 1 to achieve full connectivity.



Note:Conv3_64 represents convolutional layers with 3*3 convolution kernel and 64 channels, and so on; pool/2 represents max pooling layer with 2*2 pooling window; Proposal represents a predictive box with a fixed size; and Softmax represents a Trainable softmax classifier.

Fig. 6 Schematic Structural configuration of the proposed MHW-PD network

The VGG16 network is trained through the optimization of the loss function using the stochastic gradient descent (SGD) method for the identification of panicles, and the location of the targets are obtained through the regression model. We set the batch-size and iteration steps to 128 and 80000 respectively, and the learning rate changes from 0.001 to 0.0001 after iteration steps reaches 50000. The loss function consists of contributions from the classification and regression loss as shown in equation (6):

$$(\{P_i\},\{t_i\})=\frac{1}{N_{cls}}\sum_i L_{cls}(P_i,P_i^*)+\lambda\frac{1}{N_{reg}}\sum_i P_i^* L_{reg}(t_i,t_i^*) \qquad (6)$$

Where the N_{cls} represents the mini-batch size of training, N_{reg} represents the generated number of candidate regions, i is the anchor number, the weighting parameter λ is set as λ=10. The P_i is the probability of the anchor point being as target, and when the anchor point is predicted as positive the corresponding P_i^* value is given as 1 and otherwise it is 0 if the anchor is negative. t_i and t_i^* represent the coordinates of the upper left and lower right vertex of the predicted bouncing box respectively. L_{cls} and L_{reg} are the logarithmic and robust regression loss respectively:

$$L_{cls}(P_i,P^*_i) = -\log\left[P^*_i P_i + (1-P^*_i)(1-P_i)\right] \quad (1)$$

$$L_{reg}(t_i,t^*_i) = \begin{cases} 0.5(t_i - t^*_i)^2 & |t_i - t^*_i| < 1 \\ |t_i - t^*_i| - 0.5 & |t_i - t^*_i| \geqslant 1 \end{cases} \quad (8)$$

3.5 Performance assessment indexes

The counting accuracy and the false detection rate have been utilized as the performance indexes in this work. The counting accuracy (P_c) refers to the ratio of detecting the correct number of panicles to the actual number of panicles; while the false detection rate (P_e) is the ratio of the detection error (false positive) to the actual number of panicles (ground truth) in the imagery data set:

$$P_c = N_{cor}/N_{real} \quad (9)$$

$$P_e = N_{err}/N_{real} \quad (10)$$

Where N_{cor} and N_{err} are the correct (true positive) and wrong (false positive) number of panicles detected by the model respectively, and N_{real} represents the actual number of panicles in the test sample.

Prior to the accuracy assessment, the repeated counting of the same panicle from the MHW partitioned pictures is firstly evaluated. This is achieved through the assessment of the repetition ratio (P_{rep}) as shown in the equations (11), (12) and (13):

$$P_{rep} = \frac{N_{rep}}{\sum_{i=1}^{k} N_{subi}} \quad (11)$$

$$N_{rep} = \sum^{k} N_{subi} - N_{cor} \quad (12)$$

$i = 1$

$$P_{rrep} = \frac{\sum_{i=1}^{k} N_{subi} - N_{rep}}{N_{terp}} \quad (13)$$

where $\mathrm{N_{rep}}$ represents the number of the repeated panicles that has been removed by the fusion algorithm; N_{subi} is number of the detected panicle in the i^{th} sub-window; k is the total number of the sub-windows in the picture; N_{cor} represents the number of panicles detected after image fusion; P_{rrep} is the de-duplication rate and N_{terp} is the number of the panicles that have been counted repeatedly.

4 Results

4.1 Parameters that affect the performances of classifier

Based on the hardware mentioned in section 3.4.1, it cost about 0.102s to test a sub image for our model. In addition, to testify how the performance of the classifier is affected by the receptive field of the network, the number of layers in the hybrid windows and the effectiveness of the proposed MHW image partitioning method, two different ways of sample preparations have been utilized:

A．MHW partitioning method (see section 3.2)

B．Down-sampling method (DS):

a. Each image in the training and validation data sets (i.e. the Dataset_test) is down-sampled by a factor of 2 from the raw resolution of R1 into R2, which is then down-sampled again into R3. The down sampling was done through Laplacian filtering method (Ghiasi et al., 2016).

b. This method does not exploit any window partitioning.

The experiment was performed using one to three layers of the MHW, two different networks (ZF and VGG16) which had receptive fields to target size ratio (S_{RF} /S_{obj}) of 0.4 and 0.96 respectively (see Table 1), and data prepared with (i.e. the MHW method) and without window partitioning processing (i.e. the DS method). The averaged counting accuracy P_c over 3 experimental runs using pictures of dataset_test_1 is shown in Table 4.

Table 4. Average panicle detection results under various network configurations

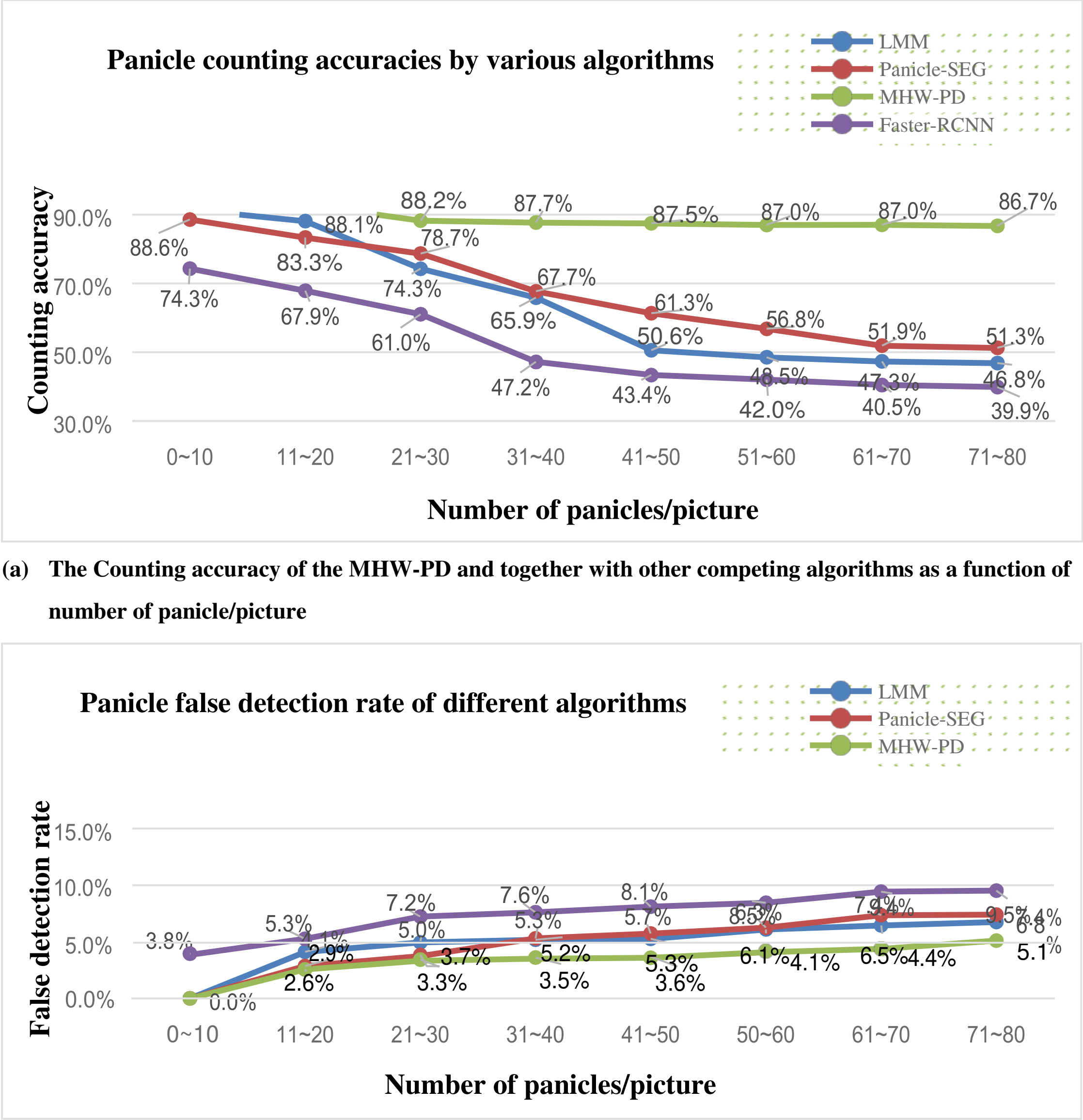
Number of MHW layers	**Resolution of MHW layer**	**P_c/% (Average ± STD)**			
		Down sampling (DS)		**MHW**	
		ZF	**VGG16**	**ZF**	**VGG16**
1	4032×3024	31.0% ± 0.37%	34.7% ± 0.37%	37.4% ± 1.12%	38.1% ± 0.56%
1	2016×1512	38.7% ± 0.96%	42.3% ± 0.37%	45.2% ± 0.37%	47.7% ± 0.56%
1	1008×756	50.2% ± 0.55%	53.5% ± 0.56%	58.4% ± 0.37%	61.2% ± 0.56%
2	4032×3024 2016×1512	41.6% ± 1.10%	44.7% ± 1.12%	47.9% ± 0.56%	50.2% ± 0.55%
2	4032×3024 1008×756	53.5% ± 0.56%	56.5% ± 1.17%	63.0% ± 0.92%	66.7% ± 0.56%
2	2016×1512 1008×756	63.5% ± 0.73%	72.9% ± 0.92%	73.1% ± 0.76%	78.1% ± 0.73%
3	4032×3024 2016×1512 1008×756	74.8% ± 0.37%	78.5% ± 0.36%	83.3% ± 0.92%	87.2% ± 0.37%

Firstly, it is noted that the reduction of the layer resolution from R1 (4032×3024 pixels) to R3 (1008×756 pixels), e.g. when the single layer of MHW of the VGG16 network is used, the panicle counting accuracy is increased from 38.1% to 61.2%. This is an almost 60% better detection when the layer is in lower (i.e. at R3) resolution. This trend of enhancement in panicle counting accuracy is seen regardless whether the data set was prepared with or without window partitioning. Secondly, the detection performance by the VGG16 network is ~5% better than that of the ZF network. This apparent small difference observed from the well matched receptive field of the VGG16 comparing to the very mismatched ZF network, is mainly due to the mixture of panicle densities in the current employed dataset_test_1. The proposed MHW enhances more of detection accuracy when the target sizes are small, i.e. when the densities of panicles are high (see section 4.2). Thirdly, when the image partitioning technique is applied (i.e. the MHW method) there is 14.4% increase in the counting accuracy in comparison to the detection that performed using non-image partitioning technique (i.e. the DS method). This can be seen, e.g. from the 61.2% accuracy given by the single layer of MHW of the VGG16 that uses input data at R3 resolution, in direct comparison to that of 53.5% obtained from the down-sampling (DS) method. Note that this ~14% of performance enhancement by using MHW is not a representative figure because of the mixed panicle densities in the dataset_test_1 that has been employed in this experiment. Fourthly, it is well-known that the increasing number of the MWH layers improves the detection performance in general, which can be seen from Table 4 that there is over 40% increase of panicle counting accuracy when the number of layers is increased from 1 to 3. Despite of using the image data set (i.e. the dataset_test_1) that contains a mixture of different panicle densities, the results presented in this section indicate that the use of multi-scale hybrid windows enhances the feature learning capacity of the network, particularly when the target sizes in the imagery is closely match to the receptive field of the feature extraction network.

4.2 Effectiveness of MHW-PD for the detection of large number of panicles

Followed by the positive results given by the previous section, the experiment here is aimed at assessing how effective is the proposed MHW-PD for the identification of different number (i.e. density) of rice panicles of the scene which is presented by the input imagery data. This section examines the proposed method vigorously by assessing the ability of the proposed MHW-PD method for counting high number of panicles (i.e. small target size), and, to compare its performance with respected to various existing algorithms. Three competing methods: a) the technique that based upon filtering using Laplacian, Median and Maxima (LMM) filters(Fernandez-Gallego et al., 2018); b) the Panicle-Seg(Xiong et al., 2017) which segments rice panicles (i.e. identification) using super-pixel clustering and CNN classification and c) the Faster-RCNN that performs panicles detection without any window partitions; had been utilized here to verify the usefulness of the proposed MHW technique for enhancing the extraction of features particularly those from small targets. Both Dataset_test_2 and Dataset_test_3 had been used as the test data for all classifiers employed in this experiment. All competing classifiers had been trained using the 130 pictures of the training data set which were in R1 resolution (i.e. 4032×3024 pixels), while the proposed MHW-PD was trained using the partitioned images in 3 different scales as described in section 4.4.1. All experiments were based on the VGG16 and they were repeated 3 times. The abilities in terms of the averaged counting accuracies and error detection rates of all classifiers to cope with scenes (i.e.

images) which contain various numbers of panicles are plotted in Figure 7.



(a) The Counting accuracy of the MHW-PD and together with other competing algorithms as a function of number of panicle/picture

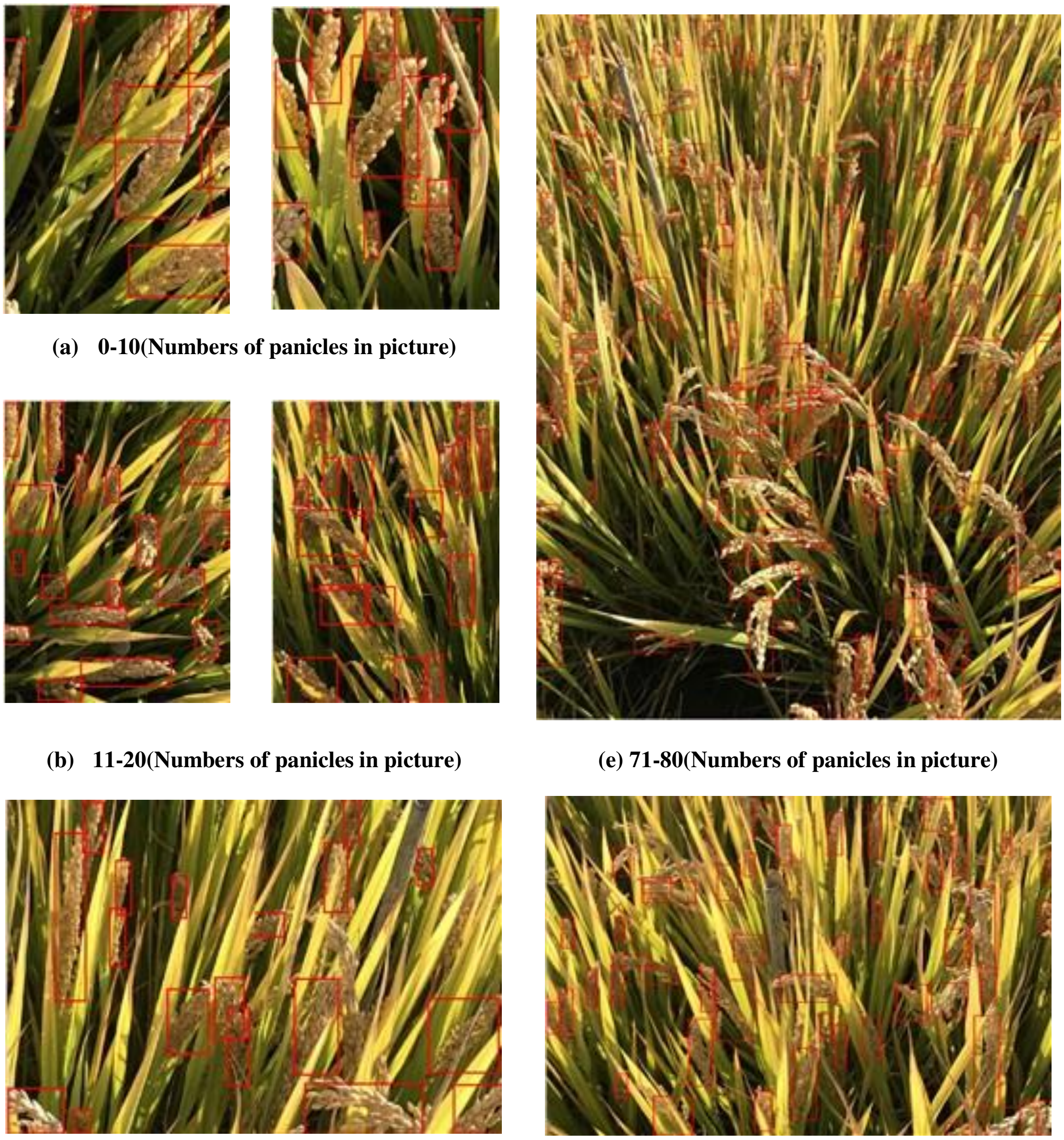
(b) The false detection rate of the MHW-PD and the other competing algorithms as a function of number of panicle/picture

Fig. 7 The Detection results of the MHW-PD and together with other competing algorithms to demonstrate the effectiveness of the proposed method particularly when high numbers of panicles are present in the scene

Figure 7 displays a rather astonished picture which exhibits the robustness of the classifiers to the increasing complexity of the rice field conditions vividly. At a glance there are two rather distinct trends that can be observed: one is the rapid decreasing detection performance, in the order of ~40%, when the number of panicles is increased from ~10 to ~50 in the scene. The other obvious trend is the very robust detection performance, with a slight drop of ~8% even when the panicle number in the scene is increased to 70-80/picture. The latter result is given by the proposed MHW-PD method which utilizes a pre-processing technique with the classification unit invariant to other competing methods (e.g. the Faster-RCNN).

One point to note is the direct comparison between the performances of the proposed MHW-PD with respected to the Faster-RCNN: in both cases the processing networks are essentially the same, however, the panicle classification performances between these two seemingly the same network are completely different. The averaged detection accuracies given by the Faster-RCNN and the MHW-PD for the scenes with panicle number <40 (i.e. when the target sizes are much larger than 260×180 pixels) are 62.6% and 90.8% respectively. This is almost 45% better detection by the MHW-PD when the panicle sizes are relatively large. However, the same two techniques for classifying the scenes with panicle number between 40 and 80 give the averaged accuracies of 41% and 87% respectively. This is over 110% of better detection by the proposed MHW-PD when the panicle sizes are small (i.e. smaller than the average size of 260×180 pixels).

Figure 8 depicts representative classified images of the rice panicle scenes obtained by using the proposed MHW-PD method. The wide range of target sizes, as depicted by the huge variations of areas of the bouncing boxes from large in Figure 8(a) to very small in Figure 8(e), highlights the increasing complexity of the scene which induces higher clutter background and the increasing difficulties to extract the feature of small targets faithfully as that depicted in Figure 8(d) & (e). This result may give another evidence that the detection capability of the propose MHW-PD method is robust against high number (density) of panicles in the rice field.



(a) 0-10(Numbers of panicles in picture)

(b) 11-20(Numbers of panicles in picture)

(e) 71-80(Numbers of panicles in picture)

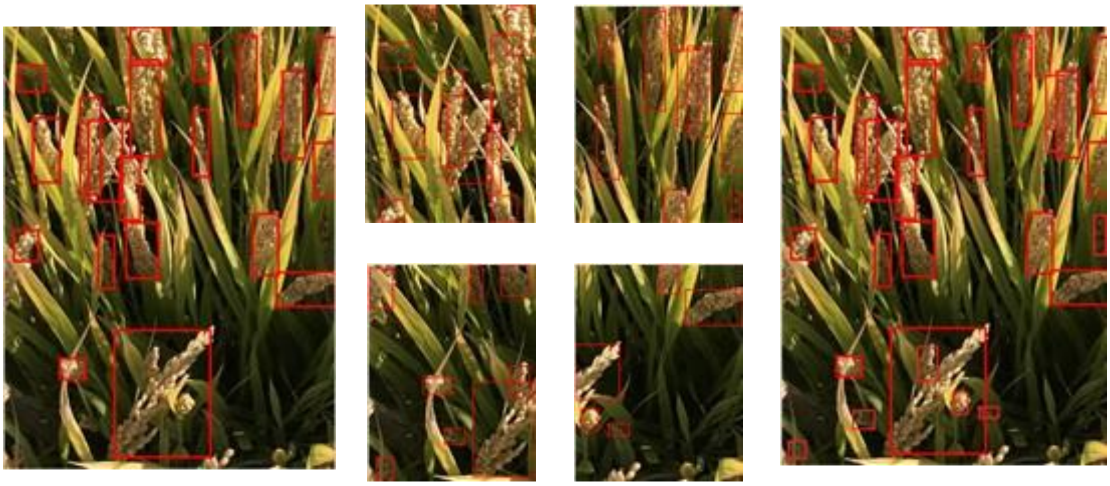
(c) 21-30(Numbers of panicles in picture)

(d) 31-40(Numbers of panicles in picture)

Fig. 8 Sample of pictures to illustrate the effectiveness of the proposed MHW-PD for the detection of various sizes of panicles in the scene

4. 3 Robustness of MHW-PD against numbers of panicles in the scene

This section highlights how the proposed MHW-PD enhances the detection of small target in the imagery data over the conventional classification routine. Here, the 'small' target in this work is referred to the relative size (in pixel unit) of the target object with respected to the pixel dimension of the input images. Figure 9a illustrates the typical classification result produced by the classifier (Faster-RCNN) in which the dimension of the input test image is at R1 resolution (i.e. 4032×3024 pixels). It is seen that some small panicles have been missed out in this classification result. The classification of the same test image after it is partitioned into 4 sub-windows (at R3 resolution) exhibits much better detections as it is illustrated in Figure 9b. After the removal of duplicated counts of dissected panicles at the boundary of sub-windows through the fusion algorithm, the end result as depicted in Figure 9c shows much better detection than that of Figure 9a. At a glance over Figure 9a and Figure 9c, one may notice immediately the distinct difference of the sizes of the panicle bouncing boxes between these two figures: more small bouncing boxes can be spotted from the MHW-PD result (Figure 9c).



(a) Result without cutting **(b) Results of HW after cutting** **(c) Result after fusing**

Fig. 9 Demonstrate the effectiveness of the MHW-PD system

Since the sub-window fusion plays an essential part in the overall performance of the MHW-PD, the robustness of the fusion algorithm over increasing complexity of the scene was investigated here. The experiment was designed to evaluate the detection performance of the algorithm for a range of assorted number of panicles in the data set (Dataset_test_3). The repetition ratio (P_{rep}) is to measure the probability of panicles being counted repeatedly, while the de-duplication rate (P_{rrep}) represents the ability of the fusion algorithm to remove the repeated counts. It can be seen from Figure 10 that P_{rep} is rather constant in the medium density (number) of panicles and it increases slightly at high number of targets in the scene. The P_{rrep} also exhibits rather steady performance at ~95% removal rate when the panicle number <90, but it tends to decrease slightly to ~92% at high end of >100 panicles in the scene. This result may give another support towards the robustness of the proposed MHW-PD system.

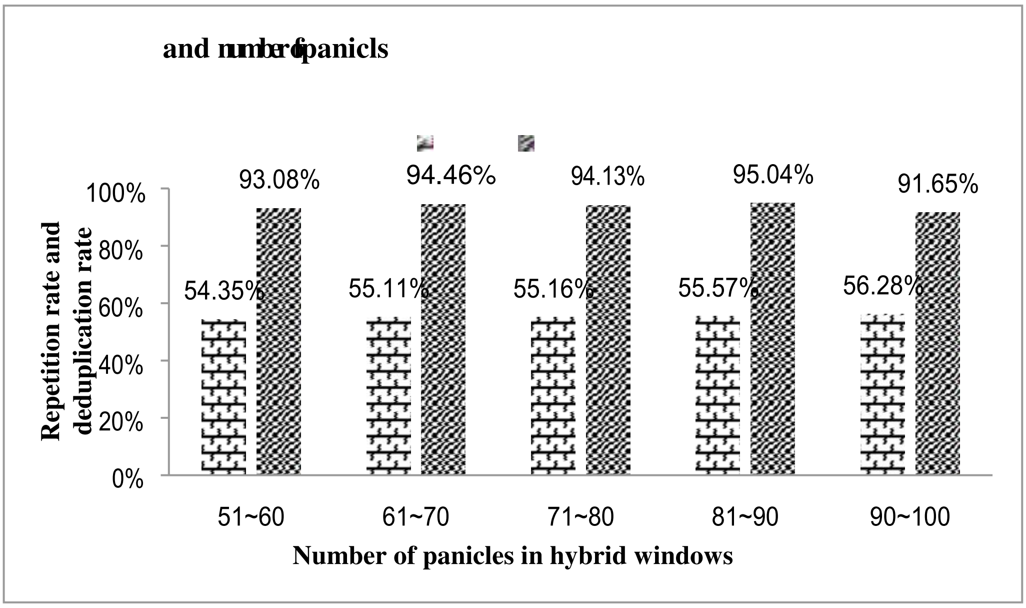
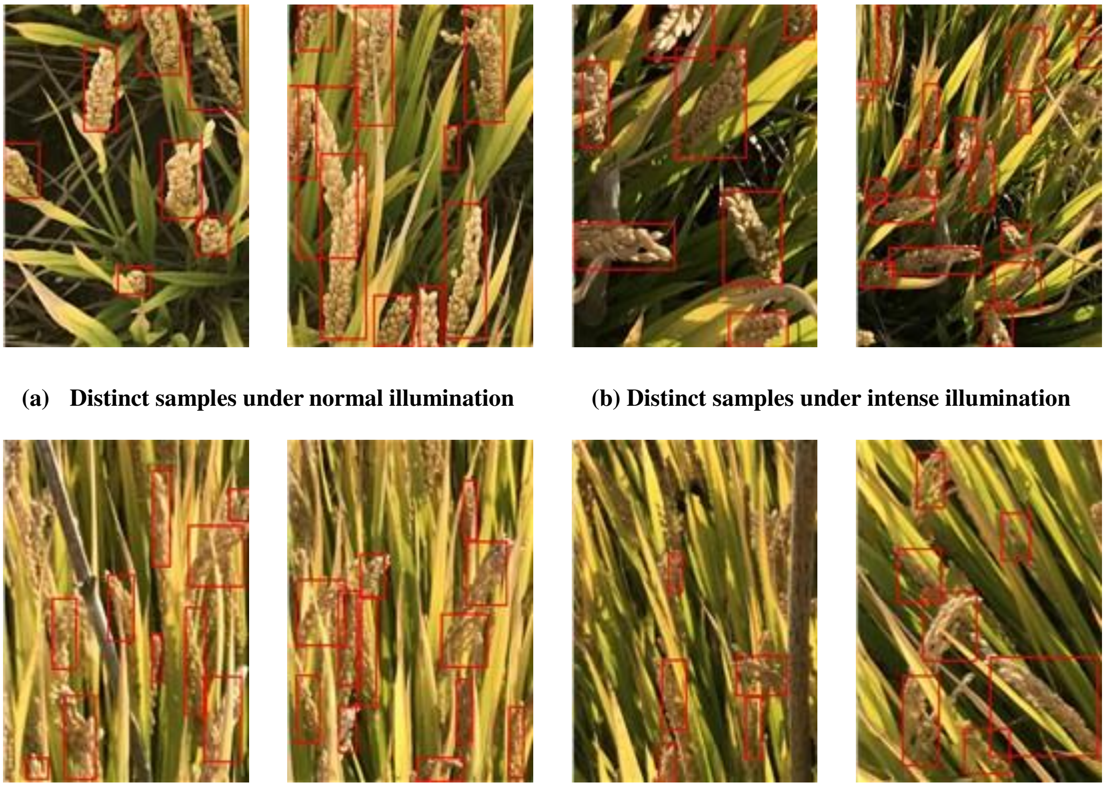


Fig. 10 Highlight the robustness of the P_{rep} and P_{rrep} of the MHW-PD against the number of panicles

4. 4 Robustness of MHW-PD against illumination and imaging artefacts

As shown in figure 8(e), it is observed that the detection results in the top of this image are obviously worsen than the bottom part. During the course of this work, we found that the bottom of images were sharp (in-focused) while the top part were blurry and fuzzy. To understand the robustness of our counting model when the quality of the input images was subjected to various degree of blurriness and shadowing artefacts, the Dataset_test_4 had been used as the test data (see Table 3), which consisted of field images subjected to various degree of blurriness and shadowing and taken under normal (i.e. weak shadowing) and intense (i.e. strong shadowing) illumination conditions. The number of panicles per picture in the Dataset_test_4 was <20. The experiments were run 3 times based on VGG16 to obtain the mean detectio2n accuracy and the associated standard deviation errors. Typical images of the classification outputs from the MHW-PD for the detection of panicles from the dataset_test_4 which contains blurry and strong shadowing pictures are shown in Figure 11. The average counting accuracies and the average false detection rates for the panicle detections of this data set are tabulated in Table 5, which reveals that the hard shadowing imposed by the intense illumination does not affect the detection efficiency significantly. However, there is ~24% drop of detection when the input images for testing are blurry. This may indicate that the fuzziness of the input image does affect the extraction of textural features as expected.



(a) Distinct samples under normal illumination

(b) Distinct samples under intense illumination

(c) Blurred samples under normal illumination

(d) Blurred samples under intense illumination

Fig. 11 To illustrate the Detection of panicles under various illumination and imaging conditions

Table 5. Average detection accuracies for images taken under various illumination and imaging conditions

Quality of input image data	Illumination conditions	P_c/% (Average ± STD)	P_e/% (Average ± STD)
In-focused pictures	Normal (weak) illumination	94.5% ± 0.78%	1.6% ± 0.26%
	Intense (strong) illumination	92.4% ± 0.37%	2.0% ± 0.16%
	Mixture of Normal & Intense illumination	93.4% ± 0.51%	1.8% ± 0.07%
Blurry pictures	Normal (weak) illumination	70.1% ± 0.89%	3.3% ± 0.42%
	Intense (strong) illumination	68.5% ± 1.08%	3.5% ± 0.34%
	Mixture of Normal & Intense illumination	69.3% ± 0.46%	3.4% ± 0.27%

5 Discussions

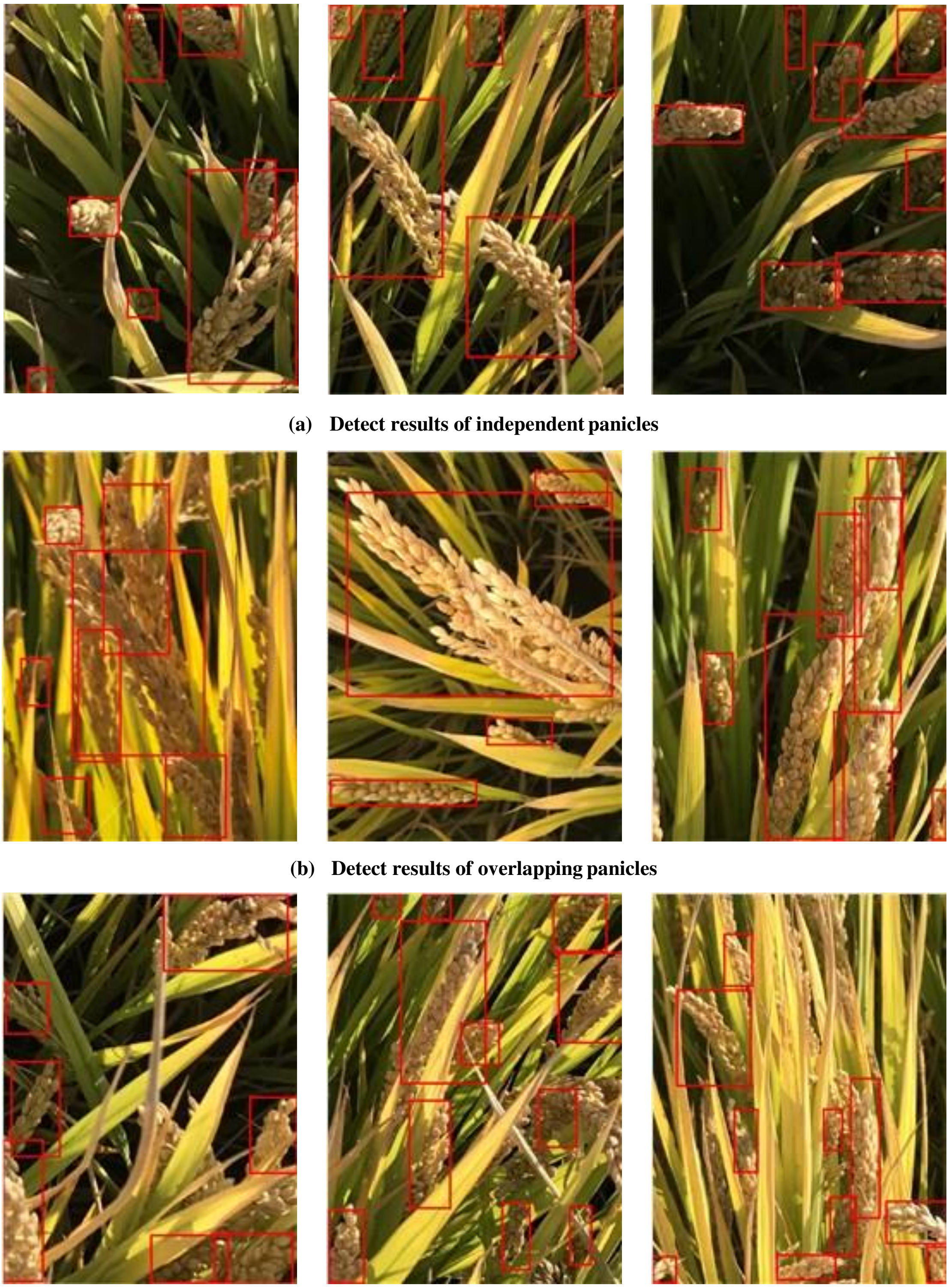
This work has reported a method (MHW-PD) to count the in-field small-sized rice panicle and function robustly independent of the panicle density. Based on the results given by the series of experiments, it is suggested that the dynamic strategies for network selection multi-scale hybrid windows construction tend to enhance the feature learning capacity of the small-sized panicles and eliminate the impact of the increase in the number of rice panicles. Compared to the pure counting method based on thermal imagery (Fernandez et al., 2019), it should be noted that, the individual rice panicle images can be segmented easily since their positions are predicted by MHW-PD. It means more phenotypic traits can be analyzed further in detail, such as the length of panicle, the radian of panicle, the number of panicle grains, the disease spot or the saturation of panicle grains and so on. In addition, the result of 87% is an average accuracy of different clarities, illuminations, occlusions and panicle numbers per image. While most of the current phenotypic studies focus on indoor potted rice, which means more stable imaging conditions (no fuzzy panicles), fewer panicles and less occlusion in the image. Thus, we suppose the MHW-PD can meet the needs of phenotypic researchers to some extent for mining the relationship from traits to genotypes, while there are also some limitations and practical issues we have to consider when the MHW-PD applied in real situations, which may constitute research directions that will be pursued in the future work.

(1) MHW-PD against occlusions. Occlusion has been one of the main factors that affect the performance of panicle counting, which may come from the high plant density and drooping, particularly when the assessment method is based on image recognition technology. In this section, 3 different kinds of occlusions have been studied: a) independent panicle when there is no obstruction, b) occlusion by leaf and c) overlapping panicles. The data set that been utilized in this experiment consisted of

<20 panicles/picture and the training/testing conditions of the MHW-PD network were the same as the previous experiments. Sample pictures of detection results for the identification of panicles in the data set that consists of these 3 types of occlusions are shown in Figure 12, and their averaged detection accuracies are tabulated in table 6. The result has shown quite clear that the detection is strongly affected by occlusions which causes some ~30% degradation of panicle accuracies with respected to the unobstructed base line, when the target panicle is occluded by leaves. Worse still is a ~60% drop in the detection accuracy when panicles in the scene are self-occluded. This large drop in detection efficiency is the inability of the classifier to discriminate the overlapped panicles and in most cases, it misclassifies the agglomerated entity as one panicle (see Figure 12b). The occlusion by leaves is not as severe as that of the self-occlusion as long as the panicle sizes are relatively larger than the leaf blades. However, the detection is seen worse when small panicles are occluded by the leaves or when large part of the panicles are covered by leaves (see Figure 12c). The very limited amount of features is not sufficient enough for the classifier to discriminate the leaf and panicle.

Table 6. Results of images with different occlusions

Types of Occlusions	P_c/%	P_e/%
Independent panicles（114 images）	95.5%	1.2%
Panicles partially covered by leaves（52 images）	62.8%	6.3%
overlapping panicles（46 images）	37.8%	29.4%



(a) Detect results of independent panicles

(b) Detect results of overlapping panicles

(c) Detect results of panicles covered by leaves

Fig. 12 Illustrate the detection by the MHW-PD for the panicles that are subjected to various occlusions

(2) MHW-PD against different imaging heights. Panicle size is the most important factor to consider when we designed the MWH-PD. However, when it comes to the different imaging heights, the main effect is the change of average panicle size. For example, if the images taken at a higher/lower altitude, the number of panicles will rise/fail sharply while the panicle size become smaller/bigger in the single image. Our ideal is selecting feature learning network which can effectively perceive a complete panicle and constructing the multi-scale hybrid windows which can extract the multi-scale panicle features. Therefore, in order to ensure the application effect of the MHW-PD, we have to design different reasonable image acquisition schemes (viewing angles, depth of field, focusing ability and optical aberrations et al.) for different particular imaging heights, which can ensure the panicle size is enough to find a matching feature learning network. At this time, the gap caused by different heights can be filled easily by selecting suitable network and constructing suitable MHW. However, we do not mean the MHW-PD can be applied under any heights because the sizes of the reception fields of the existing network are limited. From this angle, there may be a possibility to extend MHW-PD from the camera images to the high-resolution UAV images in theory, but more issues need to deal with to realize the application. For example, the huge amount of labeling work and some new processing mechanisms for the blur of panicles caused by the propeller wind when the UAV flew at a very low altitude.

(3) MHW-PD against different rice varieties. The shape of panicles has great influence on detection accuracy, which not only comes from the panicles of different rice varieties, but also from the panicles of same variety during different growth periods. In order to realize large-scale promotion application, we have to solve this inevitable problem, while it is very different to construct a universal model. Firstly, collecting images of all rice varieties/growth periods and labeling them costs a lot of money and time. Secondly, universal model means we need count and identify the species at same time. For deep learning networks, the great difficulty to solve this problem lies in how we can realize the feature representation of several rice varieties, which have small difference and even some of the difference is only local. The features can not only represent the rice panicles but also have enough differentiation to support the effective fine-grained classification for those different subspecies and varieties of rice. The problem may become even more difficult for the field scenarios because of the interference of complex field noise. One possible solution we now have tried is to iteratively build single model for every variety or growth period and cascade a multi-discrimination model for counting and identifying.

6 Conclusions

Counting small-sized rice panicles efficiently and accurately by using image based technique has been a challenging task. This paper proposes a new, yet simple method termed as MHW-PD to realize the efficacy of rice panicle counting especially when high number (density) of small-sized rice panicles is involved. The main contribution of this work is to introduce a multi-scale hybrid window (MHW) pre-processing technique for enhancing the richness of the target feature, and then to maximize the feature extraction efficiency of the network through matching the target sizes with the receptive field of the network. Through experimental design and result analysis, the conclusions can be summarized as follows:

(1) The proposed MHW-PD can significantly improve the counting accuracy for the scene where large numbers of panicles in a signal image. The combined effects of selecting the appropriate feature learning network and constructing the optimal hybrid window shown that the average counting accuracy of MHW-PD is 87.2%, which achieves >110% of detection efficiency better than that of the Faster-RCNN for the dense scenes whose number of panicles is between 50 and 80 per image.

(2) The MHW-PD has better stability in counting accuracy for the increasing number of panicle. When the panicle number increases from 10 to 80, the counting accuracy of MHW-PD comes down by 7.6%.

(3) The proposed MHW-PD can be used for infield scenes with hard shadowing imposed by intensified illumination, while the imaging and occlusion artefacts will affect the detection efficiency significantly. There is ~24% drop of detection when the input images for testing are blurry. When the panicles occluded by leaves and self-occluded with panicles crossing each other, the counting accuracy is ~30% and ~60% degradation respected to the unobstructed base line.

Acknowledgements

This paper is supported in part by the National Natural Science Foundation of China (No. 31872847); the Key Program of Science and Technology Infrastructure of Jiangsu Province of China (Modern Agriculture, BE2019383).

REFERENCES

Aich S, Stavness I. Leaf counting with deep convolutional and deconvolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2080-2089.

Alkhudaydi T, Zhou J. SpikeletFCN: Counting Spikelets from Infield Wheat Crop Images Using Fully Convolutional Networks. In: International Conference on Artificial Intelligence and Soft Computing, 2019, pp. 3-13, Springer.

Barré P, Stöver BC, Müller KF, Steinhage V. LeafNet: A computer vision system for automatic plant species identification. Ecological Informatics, 40(2017), pp. 50-56, 10.1016/j.ecoinf.2017.05.005

Cointault F, Guerin D, Guillemin JP, Chopinet B. In-field Triticum aestivum ear counting using colour-texture image analysis. New Zealand Journal of Crop and Horticultural Science, 36(2008), pp. 117-130, Doi 10.1080/01140670809510227

Dobrescu A, Valerio Giuffrida M, Tsaftaris SA. Leveraging multiple datasets for deep leaf counting. In: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2072-2079.

Du Y, Cai Y, Tan C, Li Z, Yang G, Feng H, Dong H. Field wheat ears counting based on superpixel segmentation method. Scientia Agricultura Sinica, 52(2019), pp. 21-33, 10.3864/j.issn.0578-1752.2019.01.003

Duan LF, Huang CL, Chen GX, Xiong LZ, Liu Q, Yang WN. Determination of rice panicle numbers during heading by multi-angle imaging. Crop Journal, 3(2015), pp.

211-219, 10.1016/j.cj.2015.03.002

Fernandez-Gallego JA, Kefauver SC, Gutierrez NA, Nieto-Taladriz MT, Araus JL. Wheat ear counting in-field conditions: high throughput and low-cost approach using RGB images. Plant methods, 14(2018), pp. 22-34, 10.1186/s13007-018-0289-4

Ferrante A, Cartelle J, Savin R, Slafer GA. Yield determination, interplay between major components and yield stability in a traditional and a contemporary wheat across a wide range of environments. Field Crops Research, 203(2017), pp. 114-127, 10.1016/j.fcr.2016.12.028

Ghiasi G, Fowlkes CC. Laplacian pyramid reconstruction and refinement for semantic segmentation. In: European Conference on Computer Vision, 2016, pp. 519-534, Springer.

Girshick R. Fast r-cnn. In: Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440-1448.

Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580-587.

Giuffrida MV, Minervini M, Tsaftaris SA. Learning to count leaves in rosette plants. In: Proceedings of the Computer Vision Problems in Plant Phenotyping (CVPPP), 2016, pp. 1.1-1.13.

Guo W, Fukatsu T, Ninomiya S. Automated characterization of flowering dynamics in rice using field-acquired time-series RGB images. Plant methods, 11(2015), pp. 7-

23, 10.1186/S13007-015-0047-9

Han J, Zhang D, Cheng G, Liu N, Xu D. Advanced deep-learning techniques for salient and category-specific object detection: a survey. IEEE Signal Processing Magazine, 35(2018), pp. 84-100, 10.1109/Msp.2017.2749125

Hasan MM, Chopin JP, Laga H, Miklavcic SJ. Detection and analysis of wheat spikes using Convolutional Neural Networks. Plant methods, 14(2018), pp. 100-113, 10.1186/S13007-018-0366-8

He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37(2015), pp. 1904-1916, 10.1109/TPAMI.2015.2389824

Jin XL, Liu SY, Baret F, Hemerle M, Comar A. Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. Remote Sensing of Environment, 198(2017), pp. 105-114, 10.1016/j.rse.2017.06.007

Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, 2012, pp. 1097-1105.

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, Berg AC. Ssd: Single shot multibox detector. In: European conference on computer vision, 2016, pp. 21-37, Springer.

Lu H, Cao ZG, Xiao Y, Li YA, Zhu YJ. Region-based colour modelling for joint crop and maize tassel segmentation. Biosystems Engineering, 147(2016), pp. 139-150, 10.1016/j.biosystemseng.2016.04.007

Madec S, Jin X, Lu H, De Solan B, Liu S, Duyme F, Heritier E, Baret F. Ear density estimation from high resolution RGB imagery using deep learning technique. Agricultural and forest meteorology, 264(2019), pp. 225-234, 10.1016/j.agrformet.2018.10.013

Maldonado Jr W, Barbosa JC. Automatic green fruit counting in orange trees using digital images. Computers and electronics in agriculture, 127(2016), pp. 572-581, 10.1016/j.compag.2016.07.023

Mussadiq Z, Laszlo B, Helyes L, Gyuricza C. Evaluation and comparison of open source program solutions for automatic seed counting on digital images. Computers and electronics in agriculture, 117(2015), pp. 194-199, 10.1016/j.compag.2015.08.010

Olsen PA, Ramamurthy KN, Ribera J, Chen Y, Thompson AM, Luss R, Tuinstra M, Abe N. Detecting and Counting Panicles in Sorghum Images. In: 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), 2018, pp. 400-409, IEEE.

Pound MP, Atkinson JA, Wells DM, Pridmore TP, French AP. Deep learning for multi-task plant phenotyping. In: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2055-2063.

Qiongyan L, Cai J, Berger B, Okamoto M, Miklavcic SJ. Detecting spikes of wheat plants using neural networks with Laws texture energy. Plant methods, 13(2017), pp. 83-96, 10.1186/s13007-017-0231-1

Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779-788.

Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7263-7271.

Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems, 2015, pp. 91-99.

Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014),

Slafer GA, Savin R, Sadras VO. Coarse and fine regulation of wheat yield components in response to genotype and environment. Field Crops Research, 157(2014), pp. 71-83, 10.1016/j.fcr.2013.12.004

Stein M, Bargoti S, Underwood J. Image Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry. Sensors, 16(2016), pp. 1915-1923, 10.3390/S16111915

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1-9.

Xiong X, Duan L, Liu L, Tu H, Yang P, Wu D, Chen G, Xiong L, Yang W, Liu Q. Panicle-SEG: a robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. Plant methods, 13(2017), pp. 104-119, 10.1186/s13007-017-0254-7

Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: European conference on computer vision, 2014, pp. 818-833, Springer.

Zhou C, Liang D, Yang X, Yang H, Yue J, Yang G. Wheat Ears Counting in Field Conditions Based on Multi-Feature Optimization and TWSVM. Frontiers in plant science, 9(2018), pp. 1024-1040, 10.3389/fpls.2018.01024

Declaration of interests

□The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

□The authors declare the following financial interests / personal relationships which may be considered as potential competing interests:

None

CRediT author statement

Xu can: Conceptualization, Methodology, Software, Data curation, Writing-Original draft preparation. **Jiang Haiyan:** Formal analysis, Supervision, Writing - Review & Editing, Funding acquisition. **Peter Yuen:** Writing - Review & Editing. **Zaki Ahmad Khan:** Validation Writing - Review & Editing. **Chen Yao:** Validation, Data curation.