Analysis of key factors affecting maize tassels detection and construction of shared dataset based on UAV

Xuli Zan
College of Land Science and Technology, China Agricultural University
Xiang Gao
College of Land Science and Technology, China Agricultural University
Diyou Liu
Aerospace Information Research Institute, Chinese Academy of Sciences
Wei Liu
College of Computer and Information Engineering Xiamen University of Technology
Ziyao Xing
College of Land Science and Technology, China Agricultural University
Xiaodong Zhang
College of Land Science and Technology, China Agricultural University
Zhe Liu (liuz@cau.edu.cn)
College of Land Science and Technology, China Agricultural University
Suchuang Di
Beijing Water Science and Technology Institute
Yuanyuan Zhao
College of Land Science and Technology, China Agricultural University
Shaoming Li
College of Land Science and Technology, China Agricultural University

---

Research Article

Keywords: maize tassels, efficient sample usage, spatial resolution, images acquisition time, tasseling date

Posted Date: July 24th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3175093/v1

Additional Declarations: No competing interests reported.

Analysis of key factors affecting maize tassels detection and construction of shared dataset based on UAV

Xuli Zan[1,3], Xiang Gao[1], Diyou Liu[4, 1], Wei Liu[5], Ziyao Xing[1], Xiaodong Zhang[1, 2], Zhe Liu[1, 2*], Suchuang Di[3], Yuanyuan Zhao[1, 2], Shaoming Li[1, 2]

[1] College of Land Science and Technology, China Agricultural University, Beijing, 100193, China, [2] Key Laboratory of Remote Sensing for Agri-Hazards, Ministry of Agriculture and Rural Affairs, Beijing, 100083, China, [3] Beijing Water Science and Technology Institute, Beijing, 100048, China, [4] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 101408, China and [5] College of Computer and Information Engineering，Xiamen University of Technology, Xiamen, 361024, China

** For correspondence. E-mail address: liuz@cau.edu.cn*

Background Rapid and accurate detection of tassels is of great significance for maize breeding, seed production and the acquisition of key growth stage. To liberate manpower and improve the efficiency of production management, many automatic detection methods with acceptable accuracy have been proposed. However, images acquisition parameters of these methods were quite different, so they cannot provide an operable standard for practical applications. In this study, based on multi-temporal unmanned aerial vehicle (UAV) RGB images with maize flowering stage, we created UAV Maize Tassel Detection (UAVMTD) dataset, and used Faster R-CNN to answer what are the key factors affecting detection accuracy from two aspects of efficient use of samples and data acquisition standards. Based on the detection results, we estimated tasseling date of different plots and analyzed varieties' differences.

Results The results show that model performance would not be greatly affected before the amount of training data changed by orders of magnitude, but it can be improved effectively by adjusting sub-images' sizes, and the final model was selected with AP@0.5IOU was 0.916; images obtained at 12 pm were more suitable for tassels detection, AP@0.5IOU, recall and precision were 3%, 2% and 6% higher than that at 8 am; optimal spatial resolution was around 1cm for tassels detection by considering the recognition effect and data acquisition efficiency.

Conclusions This study analyzed key factors affecting maize tassels detection and provided a reasonable reference for future applications, which is helpful to screen out varieties from large-scale breeding materials.

Keywords: maize tassels; efficient sample usage; spatial resolution; images acquisition time; tasseling date

1 Background

Maize (Zea mays L.) tassels' size and shape are related to the amount of pollen [1] and light interception of upper leaves, thus can affect the grain yield and quality [1-3]. Tasseling stage of maize is one of the crucial growth stages and also an important agronomic trait in maize breeding [4], which indicates the transition from vegetative growth to generative growth and helps in taking potential actions, such as implementing agronomic inputs to improve the crop yield. Except that, detasseling is a routine management measure in seed maize production field to ensure the purity and quality of hybrid seed [5-6]. In China, detasseling process is mainly dependent on manual and there is no effective method to inspect detasseling quality. Therefore, rapid and accurate detect maize tassels and estimate tasseling stage have practical value for seed maize production, varieties selection and production management.

Traditional crop phenotype acquisition relies on a large number of repeated manual measurements [7-8], in order to obtain flowering time of different varieties, breeders even have to observe and record in the field every day during tasseling season. This work is labor-intensive, and moreover, the results are subjective due to inconsistent standards among observers. In recent years, UAVs equipped with various sensors has been regarded as an effective tool for crop science [9-13] due to its advantages of low construction cost, flexibility and high spatio-temporal resolution [14-16]. At present, researches on maize tassel detection is mainly include classical machine learning based on multi-step combination, end-to-end objective detection algorithm and semi-supervised [17] or unsupervised methods [18] to solve large sample dependence.

Classical machine learning methods are mainly composed of proposal regions extraction and further precise detection, which is a process from coarse to fine. For example, Ferhat Kurtulmus et al. [5] realized tassel position detection with a correct rate of 81.6% by combining SVM classifier and morphological operation, in which SVM was used for image binarization, and a number of morphological operations were implemented to determine tassel position. Lu et al. [19] proposed a series of operations such as threshold segmentation and self-adaptive filtering in saliency color space to obtain the potential tassel regions, and used SVM classifier for further detection. Similarly, MAO Zheng-Chong et al. [20] and Tang et al. [21] successfully separated the region of tassels through segmentation algorithm based on HSI color space. In our previous research [6], we also developed random forest classifier (RF) and morphological method to extract proposal regions of tassels. The result showed that our method could find tassel regions with different sizes and shapes, but the error rate was high, therefore, these proposal regions were then input into VGG16 network to improve detection accuracy. This kind of classical machine learning algorithm involves many steps, the process is tedious, and there are also problems that algorithm parameters are difficult to transfer.

In order to improve the efficiency of feature space process and recognition performance, deep learning is gradually introduced into the detection of crop key organs in recent years. At present, there are relatively few researches on maize tassels detection based on deep learning, but some similar works, such as sorghum head detection [22], wheat ear [23-24] and rice panicle detection [18,25] have achieved high accuracy, which can provide reference for our research. However, it should be noted that the detection of maize tassels is more difficult than sorghum and wheat, because tassels are tiny in size, and their color, morphological characteristics have little differences from background [4]. As for maize tassels detection based on deep learning, the existing researches are as follows. Lu Hao et al. [26] proposed a deep convolutional neural network (CNN)-based local count regression framework to count maize tassels. The team then compared the robustness, detection speed and performance of several state-of-the-art object detection methods and object counting method, and proposed that object detection methods (especially Faster R-CNN) were better in performance and information acquisition [27], so Faster R-CNN is often used as tassel detection model or baseline. For instance, Yunling Liu et al. [28] constructed Faster R-CNN with ResNet and VGGNet, and applied them to UAV images at the flight altitude of 15m. They founded that ResNet had better performance, with detection accuracy of 89.96%. Aziza Alzadjali [4] can recognize UAV images at 20m flight altitude with F1-score up to 95.9%, 2% lower than that of Faster R-CNN. It can be seen that deep learning algorithm has great potential in maize tassel recognition, however, the existing researches were committed to use advanced algorithms in pursuit of higher detection accuracy, and not enough attention to the practicality of algorithms.

UAV remote sensing is the most effective way to grasp the development situation of maize tassels in breeding and seed production fields in time. However, it can be found from the evolution process of maize tassel detection methods that all kinds of existing researches pursue higher recognition accuracy by applying or constructing various models, instead of paying attention to the actual operability of algorithms; and also, didn't answer how should data be collected and processed in different application scenarios, and how should expensive manual labels be used exhaustively. That is, there is a lack of research on the key factors affecting detection performance; in addition, it is difficult to compare model due to large differences in acquisition parameters of UAV images (acquisition time, flight altitude, and sensor, etc.) in different studies. To address this issue, we constructed diversity sample dataset UAVMTD, and used the best two stage detection method Faster R-CNN to detect maize tassels of different varieties and tasseling stages. On this basis, the influence of sample size, spatial resolution and image acquisition time on model performance were analyzed to provide reference for the optimal flight plan.

2 Materials and Methods

2.1 Data Acquisition

The field trial in this paper were located in Hebi City, Henan Province, and Tongliao City, Inner Mongolia Autonomous Region, China. The former field conducted variety comparison experiment, and the latter was common production experiment, whose management scheme was consistent with the local planting management level. We obtained UAV images in Hebi of tasseling stage in 2019 and 2020, specific location of the field and maize varieties in these two years have changed (Fig. 1), but the agronomic management measures such as planting density, row spacing and plant spacing remained consistent.

In 2019, the trial area was about 34.5 m from north to south and 72 m from east to west, including 170 breeding plots with a size of 2.4m × 5m. The planting density of these plots was 67500 plants/ha, and the row spacing was 0.6m. A total of 167 maize varieties with different genetic backgrounds were sown on June 15, 2019, among which Zhengdan958 was repeated four times as the control variety (CK). In 2020, the trial area was about 40 m from north to south and 51 m from east to west, with 123 maize varieties were sown on June 11, 2020 (Fig. 1A). In 2021, three main varieties named Zhengyuanyu, Zhengdan958 and Xianyu335 were sown in May10, 2021. The planting density of each variety varied from 30000 plants/ha to 135000 plants/ ha, with a gradient of 15000 plants/ha (Fig. 1B).

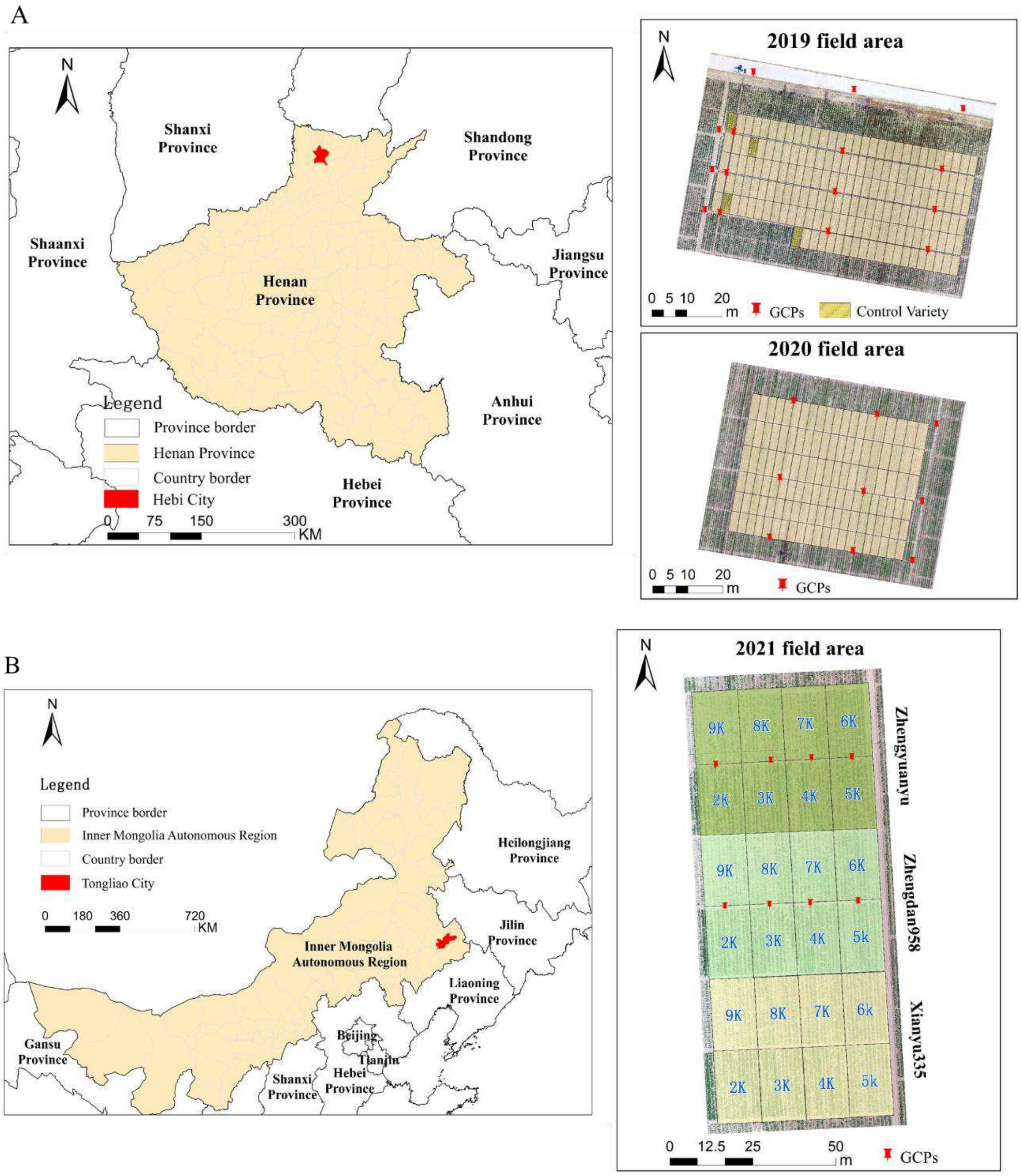

Fig. 1. The layout of experimental field. (A) Field area in Hebi City; (B) Field area in Tongliao City.

In order to improve the geometric accuracy of UAV images, ground control points

(GCPs) were deployed in the field, and the corresponding position information with centimeter-level accuracy was measured through Real-time kinematic (RTK) technology (Figure 1, i80, Shanghai Huace Navigation Technology Ltd., Shanghai, China). The visible images were acquired using DJI ZENMUSE X4s (resolution: 5472*3648) and DJI ZENMUSE X5s (resolution: 5280*3956) cameras mounted on the DJI Inspire 2 drone (DJI-Innovations, Inc., Shenzhen, China). We collected 5 groups of data from August 7 to August 12 2019, 2 groups of data at 8:00, 12:00 on August 9 2020, and 7 groups of data from July 20 to July 27, 2021, all with a flight altitude of 20m. The details of the data are shown in the following table (Table 1).

Table 1. Image data details

Year	Date	Time	Visible sensor	Flight altitude
2019	0807	15:00-16:00	DJI ZENMUSE X4s	20m
2019	0808	09:00-10:00	DJI ZENMUSE X4s	20m
2019	0809	10:00-11:00/13:00-14:00	DJI ZENMUSE X4s	20m
2019	0812	16:00-17:00	DJI ZENMUSE X5s	20m
2020	0809	08:00-09:00/12:00-13:00	DJI ZENMUSE X5s	20m
2021	0720	08:00-09:00/16:00-17:00	DJI ZENMUSE X5s	20m
2021	0722	08:00-09:00/12:00-13:00 /16:00-17:00	DJI ZENMUSE X5s	20m
2021	0725	10:00-11:00	DJI ZENMUSE X5s	20m
2021	0727	11:00-12:00	DJI ZENMUSE X5s	20m

2.2 Sample dataset construction

2.2.1 Construction of UAVMTD

Considering the different variety, planting density, production pattern and light conditions during image acquisition, 16 images were randomly selected from each group of UAV data in 2019, 2020 was 10 images from each group, while in 2021 6 images were selected, and a total of 142 images were selected to label tassel samples. In order to reduce the influence of camera lens distortion, all these images were cropped into 1/4 of original size. In addition, the selected images in 2019 and 2021 should be at least 2 shots apart (according to the overlap was 85%) so as to ensure the diversity of sample dataset and the effectiveness of model evaluation. In 2020, it should be noticed that selected images belong to the same spatial position in these two periods (8:00 am and 12:00 pm on August 9, 2020).

The LabelImg graphical image annotation tool (Fig. 2) was used to draw the bounding boxes around each tassel, and the location information was saved in xml file associated with each image. We eventually labeled 28182 tassel samples, which constituted the complete dataset of this study, called UAVMTD. In order to facilitate the use of the data set, we generated a standard description document for multiple sets of data collected on the same day. The description document was organized according to the three modules of "Basic Information", "Field Information" and " Flight Parameter", and users could read various external parameters and the growth development of maize tassels (Details are available online at: https://github.com/Xulizzz/UAVMTD).

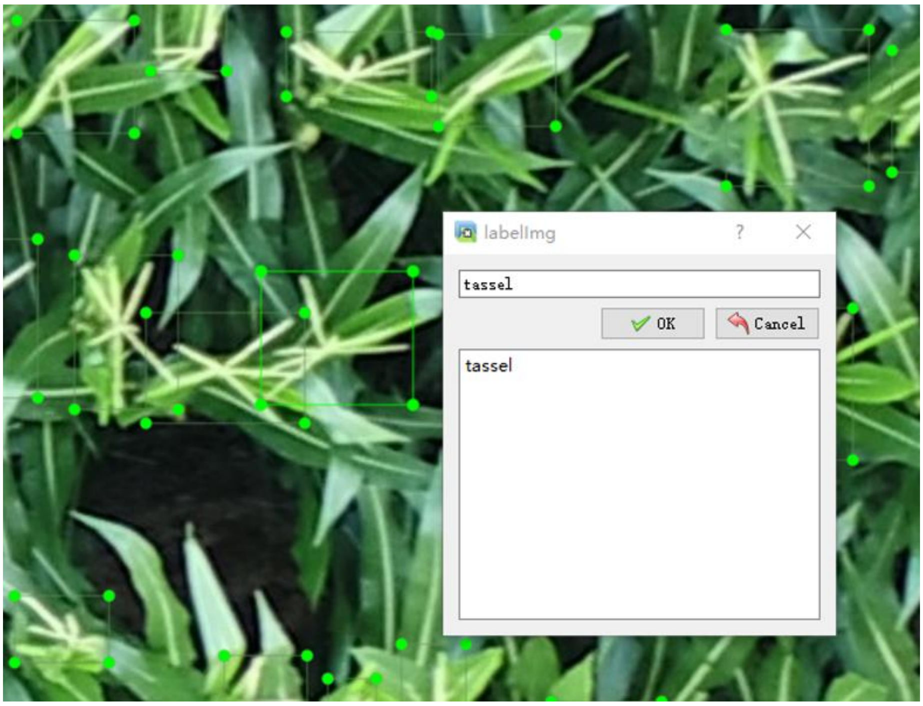

Fig. 2. Example of bounding boxes drawn

2.2.2 Sample dataset re-construction for different factor analysis

UAVMTD was divided into two parts according to images' collection years, samples in 2019 was used for model training, and the other part which contains images acquired at different production pattern and different times was used for the influence analysis of detection performance.

According to the ratio of training set: validation set: test set = 6:2:2, the corresponding amount of data was randomly extracted from labeled images in 2019 for 3 times. In order that the trained model can adapt to tassel images collected under different conditions, the 5 groups of data was evenly distributed in training, validation and test set. Due to the limitation of GPU memory, we cropped original images into multiple sub-images while keeping 50% overlap [23]. Considering that the default input size of the Faster R-CNN is 600 pixels, and to compare the effects of different sizes of sub-images on the model detection accuracy, we set the sizes of sub-images to 600 × 600 and 300 × 300 pixels. In addition, for the subsequent analysis of the impact of sample sizes on recognition accuracy, we constructed 12 different datasets by controlling the number of labeled images (50, 40, 30, 20, 10, 5) participating in model training (Table 2). It should be noted that the validation and test sets in these 12 datasets did not change.

Table 2. Details of different dataset

Dataset number	The size of sub-image	Spatial resolution	Original images for train/val/test	Sub- images for train/val
1	300×300	0.5cm	50/15/15	5850/1755
2	300×300	0.5cm	40/15/15	4680/1755
3	300×300	0.5cm	30/15/15	3510/1755
4	300×300	0.5cm	20/15/15	2340/1755
5	300×300	0.5cm	10/15/15	1170/1755
6	300×300	0.5cm	5/15/15	585/1755
7	600×600	0.5cm	50/15/15	1200/360
8	600×600	0.5cm	40/15/15	960/360
9	600×600	0.5cm	30/15/15	720/360
10	600×600	0.5cm	20/15/15	480/360
11	600×600	0.5cm	10/15/15	240/360
12	600×600	0.5cm	5/15/15	120/360

When analyzing the influence of spatial resolution on detection performance, we resample the existing dataset to simulate images acquired at different flight altitudes, which can save the time of manual annotation. Spatial resolutions of these new datasets were 1.5, 2 and 3 times of the original resolution respectively, and two schemes were adopted to set the size of sub-images during model training: one is consistent with the original dataset; the other is to adjust the size so that the number of sub-images participating in training is consistent (see 3.2.2 for specific information).

2.3 Maize tassel detection based on Faster R-CNN

In this paper, the Faster R-CNN network was selected as the tassel detection framework, because its performance was generally better than others [29]. This network is a typical two-stage object detection model proposed by Ross Girshick's team in 2015 [30], and won many first places in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) and COCO Contest in 2015. Faster R-CNN is mainly composed of Region Proposal Network (RPN) and classification part (Fig. 3).

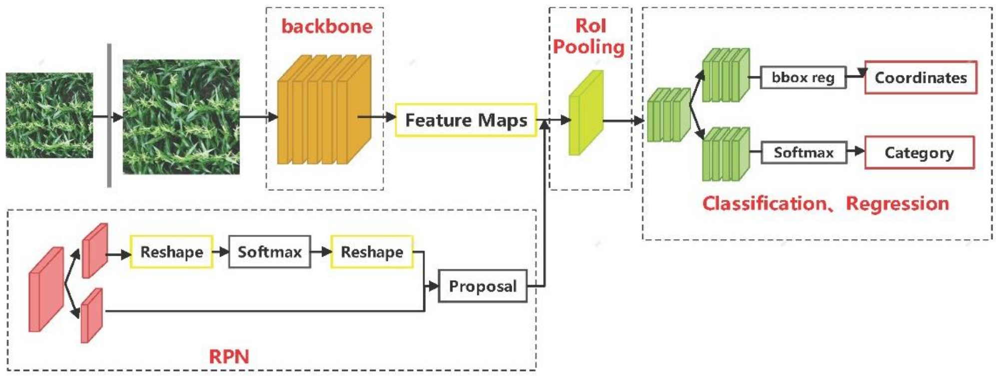

Fig. 3. Architecture of tassel detection model based on Faster R-CNN

Compared with the sliding window and image pyramid method used by OpenCV Adaboost and the selective search method used in R-CNN, the efficiency of bounding boxes generation through RPN network was greatly improved in Faster R-CNN. RPN generate a set of anchors with specific sizes and aspect ratios that centered on each pixel of the feature map output by convolution neural network, then assign whether the anchor is positive or negative and give prediction score. The RPN will eventually return top-scoring anchors as candidate bounding boxes. And then, combined with the feature map, the category of these candidate bounding boxes is calculated by a fully connected layer, and the bounding box regression is carried out to obtain accurate target boxes. In this paper, the anchor parameters were following the default setting with 9 anchors in 3 sizes and 3 aspect ratios.

We chose restnet50 as backbone, and used the weight parameters from PASCAL VOC's pre-trained model in order to achieve better detection results and reduce the running time. To prevent overfitting due to limited samples and improve the model's generalization, data augmentations with rotate, flip and mirror were used. The iterations and batch size of the model were 100 and 1, respectively, and initial learning rate was 0.001. In order to make the model parameters more in line with the data of our study, we fine-tuned model after 50 iterations, at this time the learning rate was set to 0.0001. In the whole training process, learning rate decay was adopted. Our method was implemented on the Ubuntu18.04 operation system with Inter Xeon E5-2678 and NVIDIA GeForce RTX 2080 GPU.

2.4 Model evaluation

To evaluate the influence of different factors on model performance, precision, recall rate, F1-score and AP@0.5IOU were selected [31]. Precision, recall and F1 will change with the threshold of IOU (Intersection over Union), which is the intersection ratio between the predicted bounding box and the labelled bounding box. Here, we set the IOU threshold to 0.5, and the corresponding calculation equation is as follows:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$

A predicted bounding box is considered as TP (True Positive) if IOU between it and labelled bounding box is greater than threshold. Otherwise, the predicted bounding box is considered as FP (False Positive). If the IOU between the labelled bounding box and the prediction bounding box is smaller than the threshold, the prediction bounding box is FN (False Negative). Precision refers to the proportion of TP among positive results returned by our model, with a ratio of 1.0 being perfect; while recall is the proportion of TP in all target objects. F1-score indicates the average of the precision and recall. The metrics were calculated as follows:

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

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

$$F1 = \frac{2 \times Precition \times Recall}{Precition + Recall}$$

The AP@0.5IOU was used to quantify the detection performances, which varies between 0 and 1. It refers to the area under the PR curve formed by precision, recall rate and coordinate axis under different values of IOU.

3 Results

3.1 Influence of sample usage on detection performance

In order to avoid the uncertainty in model training process and evaluate model performance more accurately under different datasets, we split dataset (a total of 80 images) into training, validation and test set through 3 random extractions. In addition, 12 datasets corresponding to Table 2 were constructed under these extractions, and model was repeatedly trained 3 times on each dataset. The AP@0.5IOU evaluation results obtained under multiple experiments are shown in Fig. 4. It can be found that the fluctuation of repeated experiments under the same size of training set is small, indicating good consistency of the results.

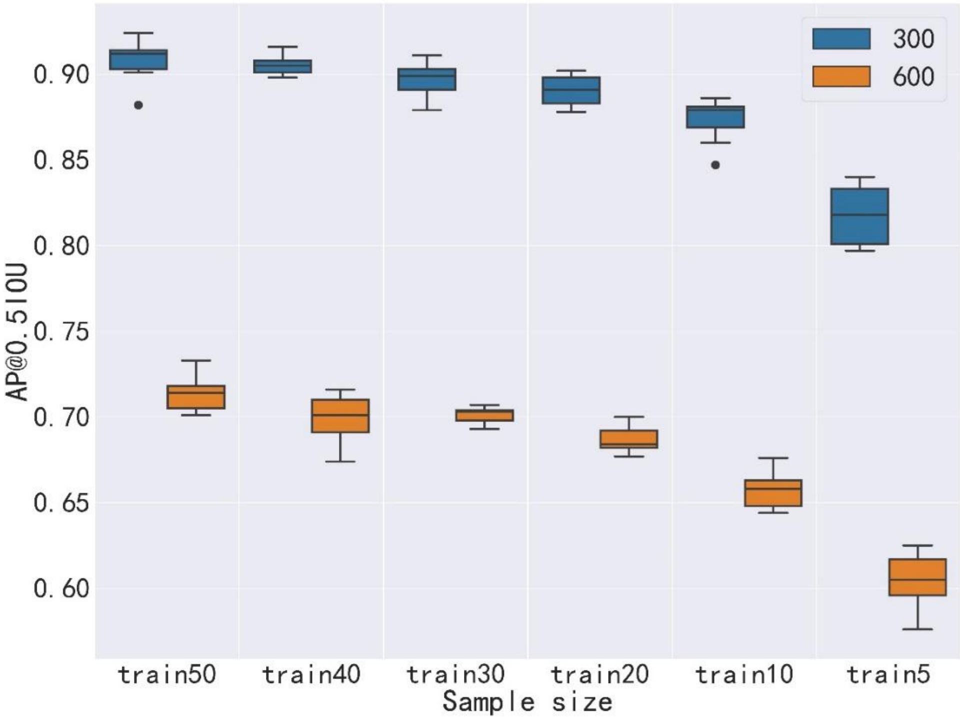

Fig. 4. Boxplots of AP@0.5IOU under multiple experiments.

3.1.1 Accuracy difference caused by labeled images' number

Fig. 5 shows the model performance under different number of labeled images and different sizes of sub-images (result is the average of multiple repeated experiments), and the corresponding specific data is shown in Table 3. It can be found that with the decrease of labeled images, each evaluation indicator showed a downward trend, but this trend was not obvious before the number decreased to 5 (the number of labeled images in Dataset6 and Dataset12 changed by orders of magnitude compared with

Dataset1 and Dataset7). Indicating that the increase of sample size with additional labeled images has little effect on performance. This phenomenon can be explained by the lower complexity of our detection problem as compared to multi-object task and the use of pre-trained model.

Moreover, with the decrease of labeled images, the precision of tassel detection did not change much, while recall changed relatively greatly. The variation range of precision were 0.1% — 3.3% and1.8% — 4.7%, when the size of sub-images was 300 × 300, 600 × 600 respectively; while recall varied from 0.2% — 9.4% and 0.2% — 8.8%, respectively. This indicating that the increase of labeled images makes the model more capable to deal with complex scenes.

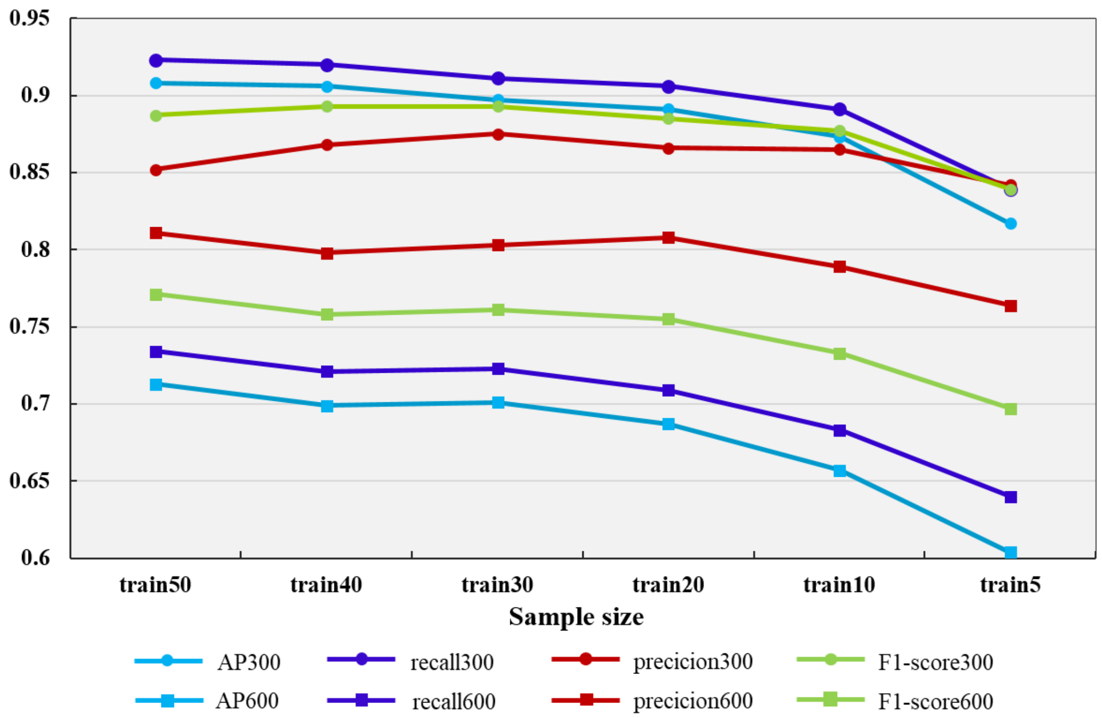

Fig. 5. The influence of sample size and sizes of sub-images on the model performance

Table 3. Performance of the Faster R-CNN models over different training set.

Dataset number	Sub- images for train/val	Mean AP@0.5IOU	Mean recall	Mean precision	Mean F1
1(#30050)	5850/1755	0.908	0.923	0.852	0.887
2(#30040)	4680/1755	0.906	0.920	0.868	0.893
3(#30030)	3510/1755	0.897	0.911	0.875	0.893
4(#30020)	2340/1755	0.891	0.906	0.866	0.885
5(#30010)	1170/1755	0.873	0.891	0.865	0.877
6(#30005)	585/1755	0.817	0.839	0.842	0.839
7(#60050)	1200/360	0.843	0.868	0.816	0.841
8(#60040)	960/360	0.837	0.864	0.798	0.829
9(#60030)	720/360	0.833	0.86	0.797	0.827
10(#60020)	480/360	0.817	0.845	0.791	0.816
11(#60010)	240/360	0.792	0.824	0.783	0.802
12(#60005)	120/360	0.74	0.782	0.769	0.775

3.1.2 Accuracy difference caused by sub-images' size

Comparing two groups of different line graphs in Fig. 5, when the labeled images participating in training were same, the 300 × 300 size of sub-images was always superior to the performance of 600 × 600, and this advantage was even greater than the change in sample size. For example, Dataset1 compared to Dataset7, Dataset2 compared to Dataset8, etc. They are all models trained with the same number of labeled images and different sub-images' size, the performance of the former was better than latter significantly, and the difference of AP@0.5IOU between them was 6%.

In addition, the advantage of changing the size of sub-images is even more significant than that of labeled images' number. For instance, the number of labeled images participating in training process of Dataset 7 was 5 times of Dataset 5, but the performance of the latter was better than former significantly. This difference was caused by the increase of training size due to the smaller sub-images' size, and the training set was further enriched through data augmentations. Therefore, in practical application, on the premise of ensuring detection accuracy, we can find the best size of sub-images to save the cost of manual annotation. In the discussion section, we specifically discussed the influence of clipping sizes on model detection performance.

3.1.3 Detection performance analysis

Based on various evaluation indicators in Table 3, the best model trained by Dataset2 was finally selected as the tassel detection model in this paper, its AP@0.5IOU was 0.916 and corresponding loss curve was shown in Fig. 6. The rapid decline in the loss value near iteration 50 was due to fine-tuning.

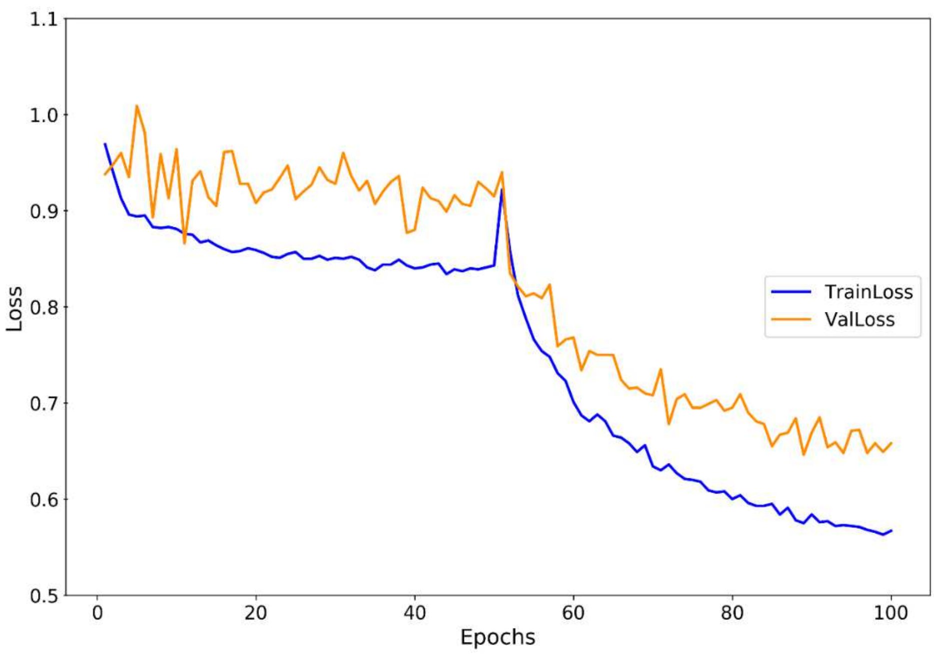

Fig. 6. Loss curve of optimal model

Observing the detection results of multi-temporal UAV images in tasseling stage, the detection effect of late tasseling stage was better than that of early tasseling stage (Fig. 7), which is consistent with our previous study [6]. This phenomenon was mainly caused by (1) the high flexibility of top leaves leads to the large inclination angle, and were easy to be misidentified as tassels, (2) tassels at early stage are difficult to be labeled without omissions and accurately, therefore, some tassels detected by trained model did not have corresponding ground truth, and also subtle difference between ground truth and detection result would be magnified (intersection positions of blue and yellow boxes in Fig. 7A). This problem also exposes that traditional evaluation metrics cannot meet the requirement of crop object recognition.

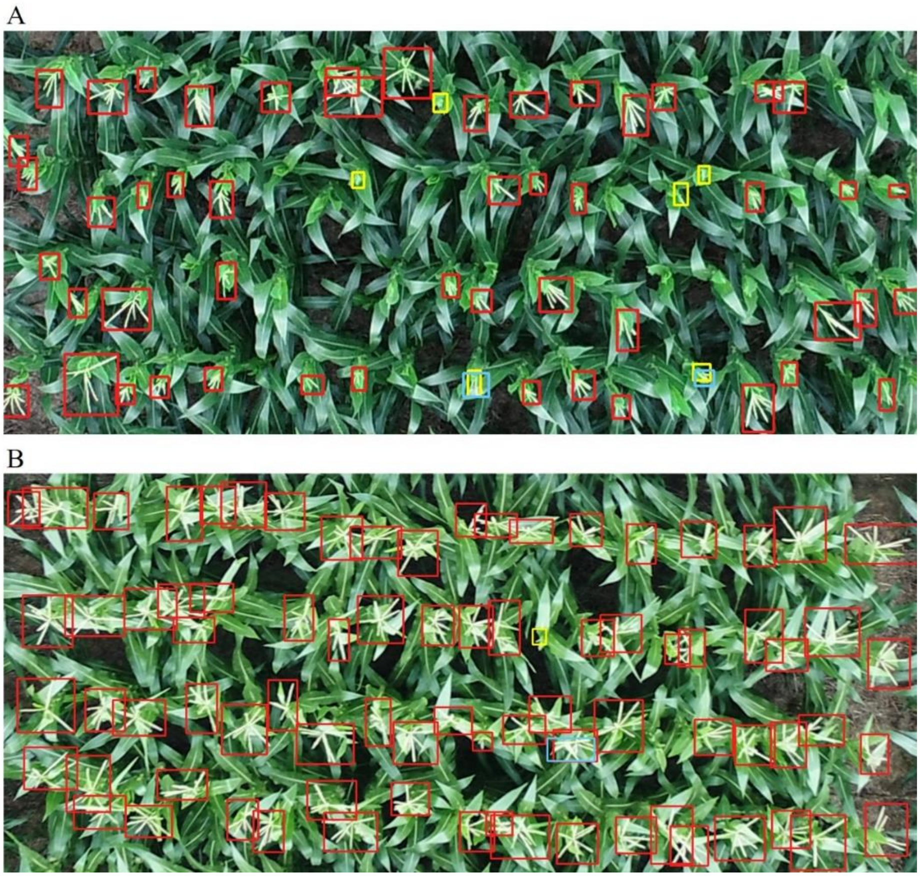

Fig. 7. Detection results of tassels under different tasseling stages. (A) Early tasseling stage; (B) Late tasseling stage. The red, blue, and yellow boxes in the figure represent correct detection, absent detection, and incorrect

detection, respectively.

3.2 Influence of flight plan on detection performance

In the current research on maize tassels detection, the flight altitude and acquisition time of UAV images are quite different, and there is no fixed reference standard of data acquisition. We compared the detection performance of UAV images with different acquisition time (changes in light conditions), different spatial resolutions, and further put forward optimal flight plan of maize tassels detection.

3.2.1 Influence of acquisition time on detection performance

We compared the detection effect of UAV images acquired at different time periods on August 9, 2020 through the best model analyzed in Section 3.1 (samples under different illumination conditions have been taken into account during model training).

Fig. 8 shows the detection effect of tassels at different acquisition time periods, columns 1 and 3 in figure are manual labeled samples on UAV images obtained at 8:00 am and 12:00 pm, which were displayed in green bounding boxes; while columns 2 and 4 are the corresponding detection results, in which red, blue and yellow represent correct detection, absent detection and incorrect detection tassels, respectively. Among the detection results at 12:00 pm, the number of incorrect detection tassels was significantly less than that of 8:00 am, which was consistent with the accuracy evaluation results in Table 4: AP@0.5IOU, recall and precision at 12:00 pm were 3%, 2% and 6% higher than those at 8:00 am. This was because the light intensity received by sensor from different angle were inconsistent caused by small solar elevation angle at 8:00 am, resulting in a large number of light spots similar to maize tassels on the image.

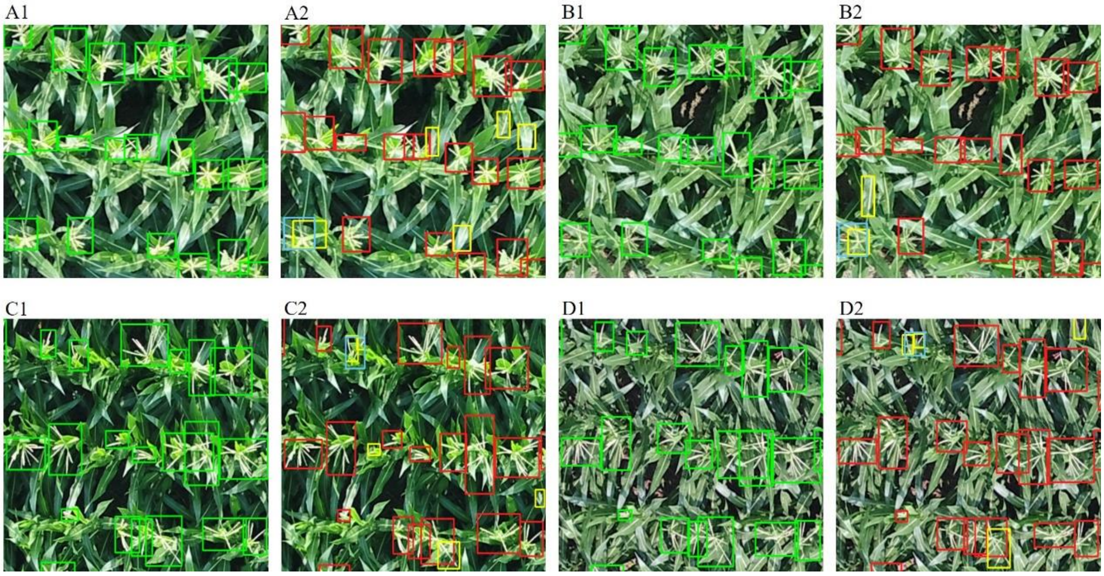

Fig. 8. Tassel detection results under different acquisition time periods. A1, B1 are manual labeled samples on UAV images obtained at 8:00 am and 12:00 pm in the same spatial position, A2, B2 are corresponding detection results. C1, C2, D1 and D2 have the same meaning. The red, blue, and yellow boxes have the same meanings as Fig. 7.

Table 4. Detection performance under different acquisition time periods.

Date	Time	AP@0.5IOU	Recall	Precision	F1
2020/08/09	08:00 am	0.907	0.922	0.807	0.861
2020/08/09	12:00 pm	0.937	0.948	0.866	0.905

3.2.2 Influence of spatial resolution on detection performance

In order to evaluate the impact of spatial resolution on model accuracy, the training set corresponding to the current best performing model was selected to construct training sets with different spatial resolutions as descried in Section 2.22 (spatial resolutions were 1.5, 2 and 3 times of the original resolution respectively). Then, detection model was retrained on datasets with different spatial resolutions. Two schemes were adopted for the size of sub-images, as shown in Table 5: (a) the size of sub-images was consistent with Dataset 2, so the amount of data participating in training process decreased with the change of spatial resolution (Fig. 9 A, B, C); (b) the size of sub-images was adjusted so that the number of data participating in training was kept consistent with Dataset 2 (Fig. 9 D, E, F, we have stretched these pictures to the same size for better display, and their real sizes were 200×200, 150×150 and 100×100 pixels , respectively).

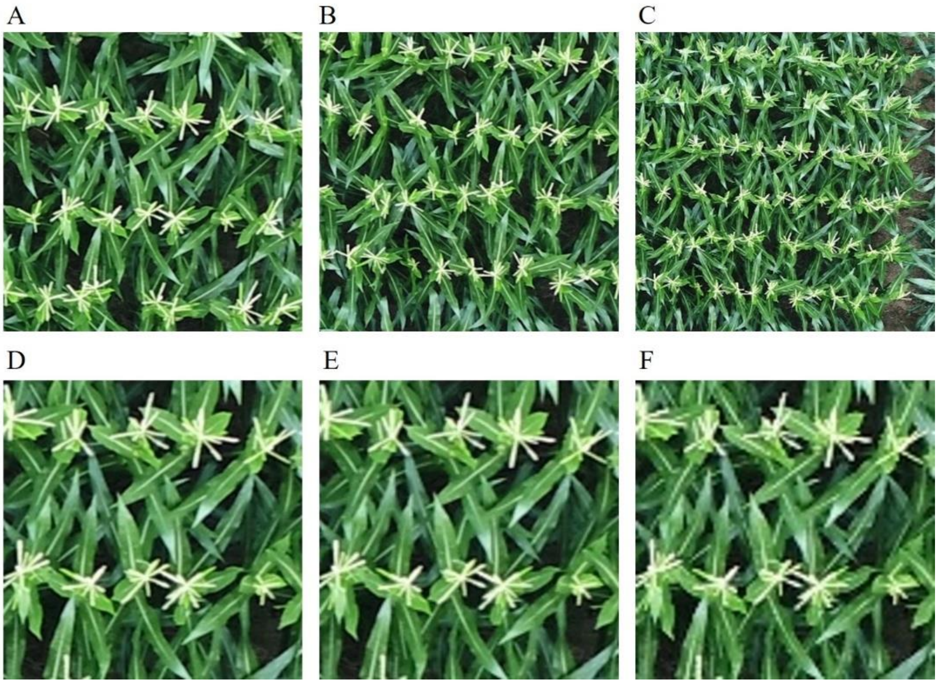

Fig. 9. Images under different spatial resolutions. A, B and C are 1.5, 2, 3 times of the original resolution respectively, following the first clipping scheme; B, C and D are also 1.5, 2, 3 times of the original resolution, but following the second clipping scheme.

The model performance with different spatial resolutions obtained by these two schemes is shown in Table 5. When the spatial resolution was decreased and the size of sub-images remains unchanged (2A, 2B, 2C in Table 5), the amount of training data decreased sharply (gradually dropping from 4680 to 320), resulting in the decline of model performance, and AP dropped from the original 0.906 to 0.390. Detection results showed that the number of absent tassels gradually increased with the decrease of spatial resolution (Fig. 10 A, B, C); However, while the spatial resolution was decreased, and the size of sub-images was correspondingly changed to ensure that the amount of training data remains unchanged (2D, 2E, 2F in Table 5), the reduction of spatial resolution to a certain extent would not change the model performance. The recognition effect was also good when the resolution was 1cm with an estimated flight altitude was 35m. Only when the resolution was reduced to tassels become blurred (for example, Dataset 2F and Fig. 9F in this experiment, which corresponds to an estimated flight altitude of 50 m) would there be a large difference in performance, with AP decreased by 5%. Comparing results under the same resolution, like 2A and 2D, 2B and 2E, as well as 2C and 2F, the model effect of the second scheme was always better than that of the first, due to the smaller size of sub-images and increased number of training data, which was consistent with Section 3.1.2.

Table 5. Model performance under different spatial resolution data sets.

Dataset number	The size of sub-images	Spatial resolution	Flight altitude	Sub- images for train	Mean AP@0.5IOU
2	300×300	0.5cm	20m	4680	0.906
2A	300×300	0.75cm	25m	1920	0.866
2B	300×300	1cm	35m	960	0.752
2C	300×300	1.5cm	50m	320	0.390
2D	200×200	0.75cm	25m	4680	0.906

2E	150×150	1cm	35m	4680	0.897
2F	100×100	1.5cm	50m	4680	0.853

The flight altitudes were calculated according to the sensor performance we used in this paper.

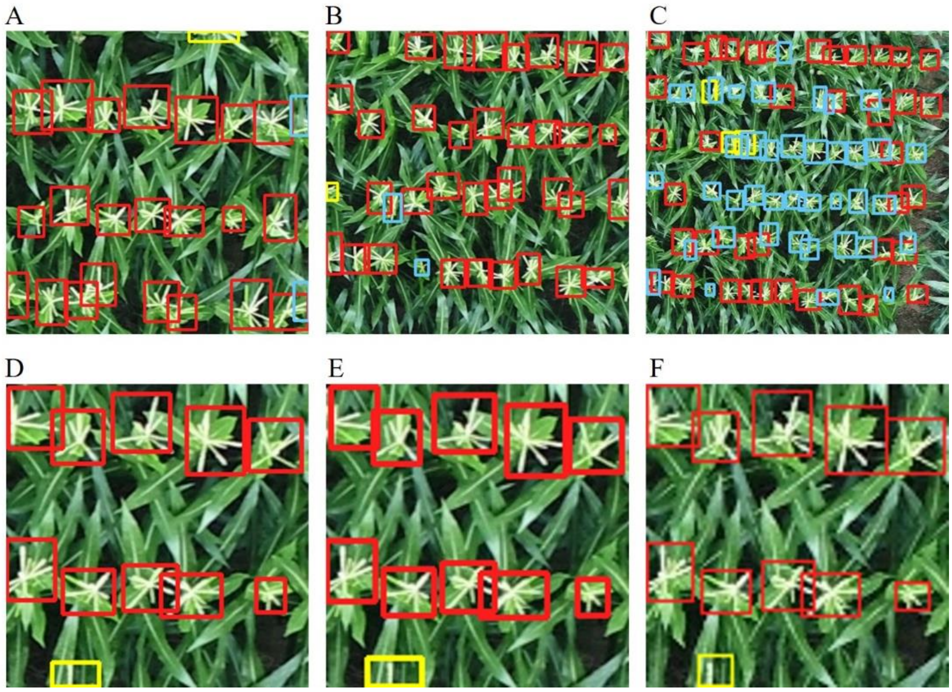

Fig. 10. Detection results under different spatial resolutions (A) (B) (C) (D) (E) and (F) have the same meanings with Fig. 9. The red, blue, and yellow boxes in the figure represent correct detection, absent detection, and incorrect detection, respectively.

Combined with the above experiments, it can be known that the reduction of spatial resolution will lead to the deterioration of detection performance, but as long as the image quality is not too poor, this difference can be made up in the data processing stage by decreasing the size of sub-images to increase the amount of training set. Therefore, in future applications, we can increase flight altitude to acquire phenotypic information more efficiently. For example, with DJI ZENMUSE X4s camera, the flight altitude can be increased to 35 m, which could obtain more images of 0.067 ha per minute compared with 20m.

3.3 Differences in variety characteristics during tasseling

Tasseling stage for population is the day when 50% plants have tassel. Therefore, relying on manpower to obtain tasseling date requires breeder to patrol each plot every day, which is time-consuming. To address this issue, A. Kumar, M [32] (Kumar et al., 2020) used Faster R-CNN model to detect the growth of tassels and estimate tasseling stages based on UAV images of different stages.

Based on the detection results of time series UAV images, combined with the actual number of plants in the plot, which can be calculated according to planting density, row spacing and plant spacing, we described the variation of tasseling rate in 9 plots (3 of these plots were planted with check variety Zhengdan958) over continuous observation (Fig. 11). With half of actual plants number as an indication, tasseling date of each plot was roughly obtained in combination with tasseling rate (Table 6). However, as the study area was in the rainy season during tasseling period, two days of data were missing in images acquisition process, so the results displayed in italics may have one-day error in Table 6.

Although all varieties were sown on the same day, there were differences in tasseling date and stage among varieties. Only Plot1 and Plot4 had the same tasseling date as the check variety (August 8), and the rest were later than that. Except that, the tasseling duration of Plot 6 was longer than others, and tasseling date was late, which reflected the poor tasseling uniformity of this variety. As check variety, Zhengdan958 showed same tasseling rate curves of CK1 and CK2, indicating that this variety is stable. The reason for the big difference between CK3 and these two plots is that the serious lack of seedlings in the early stage. Although it is made up by seedling supplementation, but growth stage in this plot was inconsistent, so the tasseling process lasted a long time.

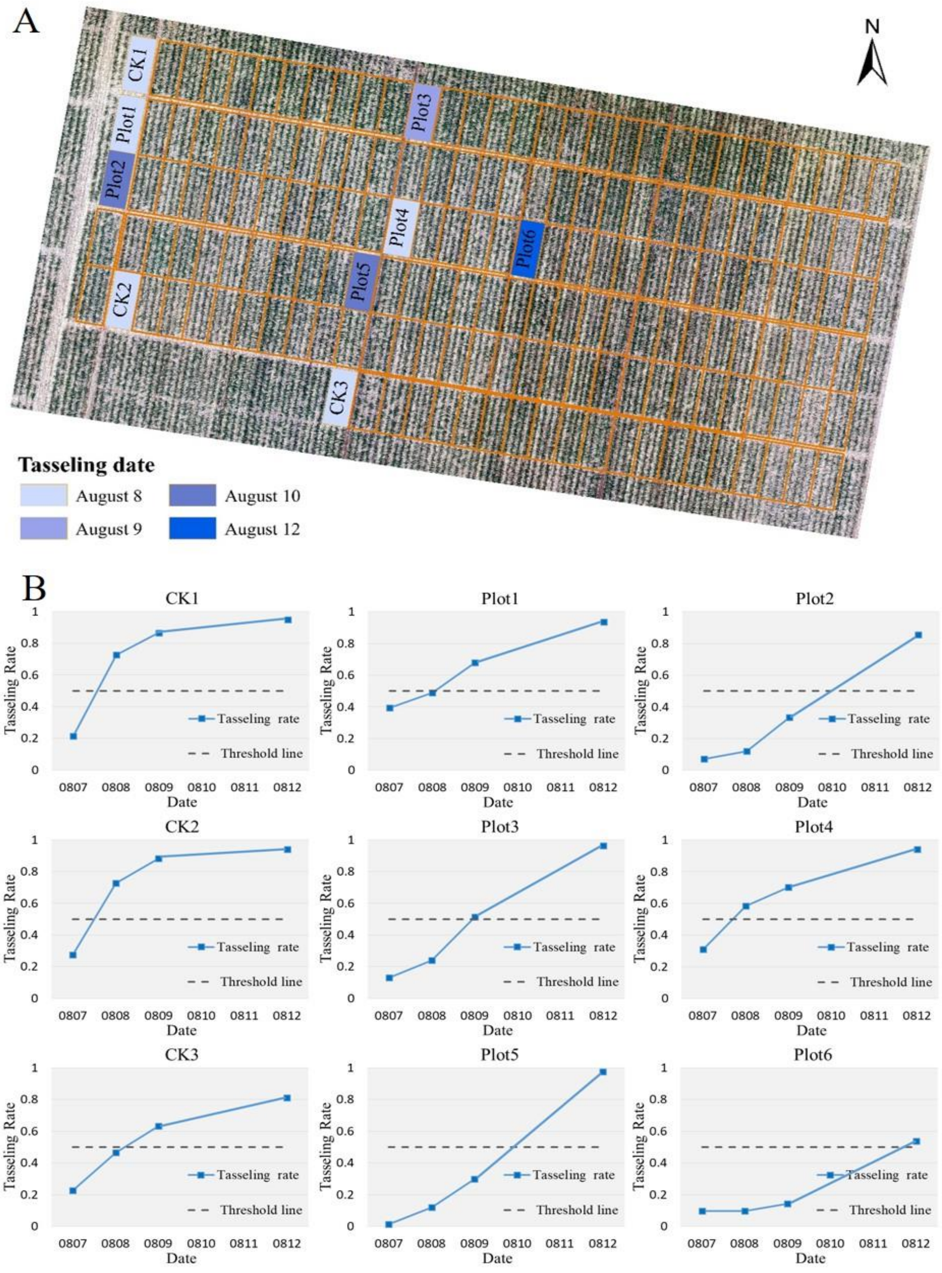

Figure 11. Tasseling rate of different varieties. CK1—CK3 are the results of repeated experiments of Zhengdan958 as a check variety.

Table 6. Tasseling rate and stage of plots.

Plot	Tasseling rate on August 7	Tasseling rate on August 8	Tasseling rate on August 9	Tasseling rate on August 12	Tasseling stage
CK1	0.214	0.726	0.869	0.952	August 8
CK2	0.274	0.726	0.881	0.940	August 8
CK3	0.226	0.464	0.631	0.810	August 8
Plot1	0.393	0.488	0.679	0.940	August 8
Plot2	0.071	0.119	0.333	0.857	*August 10*
Plot3	0.131	0.238	0.512	0.964	August 9
Plot4	0.310	0.583	0.702	0.940	August 8
Plot5	0.012	0.119	0.298	0.976	*August 10*
Plot6	0.095	0.095	0.143	0.536	August 12

4 Discussion

The results obtained in Section 3.12, the 300 × 300 size of sub-images always outperforms the 600 × 600. In order to explore whether this advantage always exists with the change of sub-images' size, we constructed several training sets with different sizes on the basis of Dataset8 in Table 2, and analyzed the influence of different sub-images' sizes with the number of original labeled images was the same. Referring to the width and height distributions of labeled samples (Fig. 12), tassels' sizes were mostly within the range of (11, 100), so the clipping sizes of sub-images include 600 × 600, 500 × 500, 400 × 400, 300 × 300, and 200 × 200.

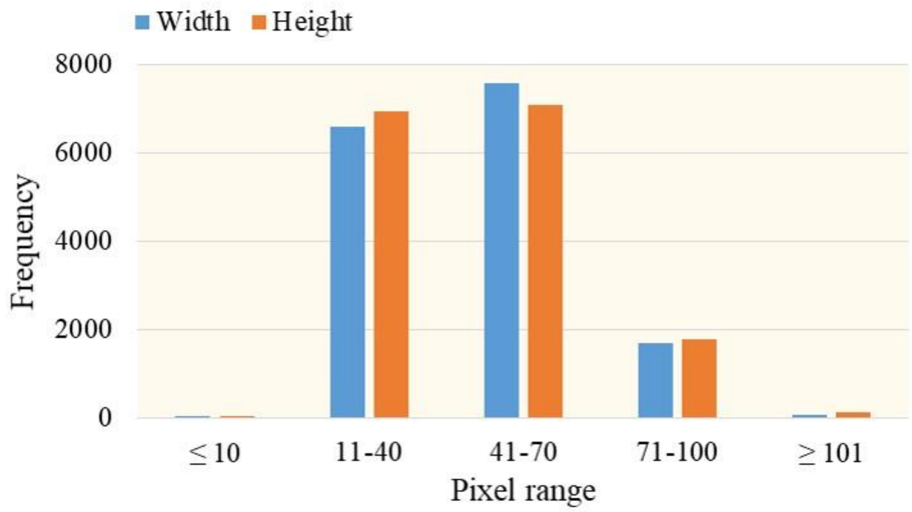

Fig. 12. The distributions of pixel width and height of tassel samples.

With the size of sub-images changes, the amount of data involved in model training process has changed greatly and evaluation indicators increased gradually (except 200 × 200). In general, when the amount of training set reaches about 3000, considerable detection result can be achieved, which is consistent with the results obtained in Table 3. Moreover, the improvement of performance with the reduction of sub-images' size will not always exist. For example, the size of 200 × 200 in this paper was too small and tassels were divided into incomplete parts, which was disadvantageous for model learning, and also these incomplete parts led to the decline of precision when evaluated model performance. Therefore, when selecting the size of sub-images, it should be considered comprehensively in combination with the detection object's size and the amount of training set.

Table 7. Model performance under different sizes of sub-images.

Dataset number	The size of sub-image	Sub- images for train/val	Mean AP@0.5IOU	Mean Recall	Mean Precision	Mean F1
8	600×600	960/360	0.699	0.721	0.798	0.758
8.1	500×500	1400/525	0.821	0.84	0.828	0.834
8.2	400×400	2800/810	0.885	0.903	0.85	0.87
2	300×300	4680/1755	0.906	0.920	0.868	0.893
8.3	200×200	11200/4200	0.884	0.905	0.818	0.859

It is worth pointed out that all the studies carried out in this paper were based on manual labeled samples. Due to individual differences among sample markers, all samples were cross-checked, and our interest objects are small in size and large in number, so it taken a lot of time to prepare labeled dataset. In addition, we founded that the detection result of trained Faster R-CNN model was better than manual labeling that without cross-checked, because many tassels were missed by us during labeling. These above challenges force us to propose a more reliable and efficient sample labeling method, so we released the dataset used in this study, hoping to reduce workload of researchers engaged in similar work. In the future, we will explore unsupervised method combined with manual verification to generate a large number of samples [17] and constantly update this dataset.

5 Conclusion

The aim of this paper is to confirm what are the key factors affecting detection accuracy and provide reference for future practical application. Meanwhile, we also estimated tasseling date of different plots and analyzed differences in variety characteristics. For this, we created UAVMTD dataset from 142 UAV RGB images acquired during 3 years, and it has assembled 28182 manually labeled maize tassels containing different varieties, planting density, production pattern and different light conditions. UAVMTD is available online at: https://github.com/Xulizzz/UAVMTD. The main conclusions are as follows:

(1) The model trained based on our diversity dataset UAVMTD can adapt to different scenes, and meet the need of tasseling date extraction, which is helpful to screen maize tassels with specific characteristics;

(2) The decrease of labeled images participating in model training would lead to the degradation of various evaluation indicators, but this trend was not obvious before the data changed by orders of magnitude; Combining the size of detection object and the amount of training set to determine sub-images' size can effectively improve detection accuracy;

(3) The detection result on images acquired at 12 pm was better than that at 8 am, with AP@0.5IOU, recall and precision were 3%, 2% and 6% higher respectively. In the future, we can try to collect UAV images at noon to improve the detection effect;

(4) To a certain extent, the coarser spatial resolution did not lead to poor recognition performance. When the resolution was 1cm, maize tassels could also be detected well by adjusting the size of sub-images.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The dataset analyzed are available at: https://github.com/Xulizzz/UAVMTD

Competing 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.

Funding

This work was supported by the National Key Research and Development Program of China and Shandong Province, China（2021YFB3901300）.

Authors' information

College of Land Science and Technology, China Agricultural University, Beijing, 100193, China

Xuli Zan, Xiang Gao, Diyou Liu, Ziyao Xing, Xiaodong Zhang, Zhe Liu, Yuanyuan Zhao, Shaoming Li

Key Laboratory of Remote Sensing for Agri-Hazards, Ministry of Agriculture and Rural Affairs, Beijing, 100083, China

Xiaodong Zhang, Zhe Liu, Yuanyuan Zhao, Shaoming Li

Beijing Water Science and Technology Institute, Beijing, 100048, China

Xuli Zan, Suchuang Di

Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 101408, China

Diyou Liu

College of Computer and Information Engineering，Xiamen University of Technology, Xiamen, 361024, China

Wei Liu

Corresponding author

Correspondence to Zhe Liu

Authors' contributions

XZ (Xiaodong Zhang), SL and ZL conceived the original research; XZ (Xuli Zan), XG and WL carried out field area investigation and data acquisition; XZ (Xuli Zan) ,XG and DL proposed methodology; XZ (Xuli Zan), ZX, SD and YZ analyzed initial data; ZL managed sample dataset; XZ (Xuli Zan) wrote the original draft. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Gage JL, Miller ND, Spalding EP, Kaeppler SM, de Leon N. TIPS: a system for automated image-based phenotyping of maize tassels. Plant Methods. 2017;13:1–12.
2. Hunter RB, Daynard TB, Hume DJ, Tanner JW, Curtis JD, Kannenberg LW. Effect of tassel removal on grain yield of corn (Zea mays L.) 1. Crop Sci. 1969;9:405–6.
3. Gao Z, Sun L, Ren J-H, Liang X-G, Shen S, Lin S, et al. Detasseling increases kernel number in

maize under shade stress. Agric For Meteorol. 2020;280:107811.
4. Alzadjali A, Alali MH, Sivakumar ANV, Deogun JS, Scott S, Schnable JC, et al. Maize Tassel Detection From UAV Imagery Using Deep Learning. Front Robot AI. 2021;8.
5. Kurtulmus F, Kavdir I. Detecting corn tassels using computer vision and support vector machines. Expert Syst Appl. 2014;41:7390–7.
6. Zan X, Zhang X, Xing Z, Liu W, Zhang X, Su W, et al. Automatic detection of maize tassels from UAV images by combining random forest classifier and VGG16. Remote Sens. 2020;12:3049.
7. Han L, Yang G, Dai H, Xu B, Yang H, Feng H, et al. Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. Plant Methods. 2019;15:1–19.
8. Ribera J, He F, Chen Y, Habib AF, Delp EJ. Estimating phenotypic traits from UAV based RGB imagery. arXiv Prepr arXiv180700498. 2018.
9. Yue J, Yang G, Li C, Li Z, Wang Y, Feng H, et al. Estimation of winter wheat above-ground biomass using unmanned aerial vehicle-based snapshot hyperspectral sensor and crop height improved models. Remote Sens. 2017;9:708.
10. Zhou L, Gu X, Cheng S, Yang G, Shu M, Sun Q. Analysis of plant height changes of lodged maize using UAV-LiDAR data. Agriculture. 2020;10:146.
11. Apolo-Apolo OE, Martinez-Guanter J, Egea G, Raja P, P and Pérez-Ruiz M. Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV. Eur J Agron. 2020;115:126030.
12. Ma J, Li Y, Chen Y, Du K, Zheng F, Zhang L, et al. Estimating above ground biomass of winter wheat at early growth stages using digital images and deep convolutional neural network. Eur J Agron. 2019;103:117–29.
13. Tao H, Feng H, Xu L, Miao M, Long H, Yue J, et al. Estimation of crop growth parameters using UAV-based hyperspectral remote sensing data. Sensors. 2020;20:1296.
14. Watanabe K, Guo W, Arai K, Takanashi H, Kajiya-Kanegae H, Kobayashi M, et al. High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. Front Plant Sci. 2017;8:421.
15. Yang G, Liu J, Zhao C, Li Z, Huang Y, Yu H, et al. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. Front Plant Sci. 2017;8:1111.
16. Jang G, Kim J, Yu J-K, Kim H-J, Kim Y, Kim D-W, et al. Cost-effective unmanned aerial vehicle (UAV) platform for field plant breeding application. Remote Sens. 2020;12:998.
17. Kumar A, Desai SV, Balasubramanian VN, Rajalakshmi P, Guo W, Naik BB, et al. Efficient Maize Tassel-Detection Method using UAV based remote sensing. Remote Sens Appl Soc Environ. 2021;23:100549.
18. Hayat MA, Wu J, Cao Y. Unsupervised Bayesian learning for rice panicle segmentation with UAV images. Plant Methods. 2020;16:1–13.
19. Lu H, Cao Z, Xiao Y, Fang Z, Zhu Y, Xian K. Fine-grained maize tassel trait characterization with multi-view representations. Comput Electron Agric. 2015;118:143–58.
20. Zhengchong M, Yahui S. Algorithm of male tassel recognition based on HSI space. Transducer Microsyst Technol. 2018.
21. Tang W, Zhang Y, Zhang D, Yang W, Li M. Corn tassel detection based on image processing BT - 2012 international workshop on image processing and optical engineering. International Society for Optics and Photonics; 2011. p. 83350J.

22. Guo W, Zheng B, Potgieter AB, Diot J, Watanabe K, Noshita K, et al. Aerial imagery analysis-quantifying appearance and number of sorghum heads for applications in breeding and agronomy. Front Plant Sci. 2018;1544.

23. Madec S, Jin X, Lu H, De Solan B, Liu S, Duyme F, et al. Ear density estimation from high resolution RGB imagery using deep learning technique. Agric For Meteorol. 2019;264:225–34.

24. Velumani K, Madec S, de Solan B, Lopez-Lozano R, Gillet J, Labrosse J, et al. An automatic method based on daily in situ images and deep learning to date wheat heading stage. F Crop Res. 2020;252:107793.

25. Zhou C, Ye H, Hu J, Shi X, Hua S, Yue J, et al. Automated counting of rice panicle by applying deep learning model to images from unmanned aerial vehicle platform. Sensors. 2019;19:3106.

26. 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;13:1–17.

27. Zou H, Lu H, Li Y, Liu L, Cao Z. Maize tassels detection: a benchmark of the state of the art. Plant Methods. 2020;16:1–15.

28. Liu Y, Cen C, Che Y, Ke R, Ma Y, Ma Y. Detection of maize tassels from UAV RGB imagery with faster R-CNN. Remote Sens. 2020;12:338.

29. Pound MP, Atkinson JA, Townsend AJ, Wilson MH, Griffiths M, Jackson AS, et al. Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. Gigascience. 2017;6:gix083.

30. 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.

31. Davis J, Goadrich M. The relationship between Precision-Recall and ROC curves. Proc 23rd Int Conf Mach Learn. 2006. p. 233–40.

32. Kumar A, Taparia M, Rajalakshmi P, Guo W, Naik B, Marathi B, et al. UAV based remote sensing for tassel detection and growth stage estimation of maize crop using multispectral images. IGARSS 2020-2020 IEEE Int Geosci Remote Sens Symp. IEEE; 2020. p. 1588–91.