Fast Anther Dehiscence Status Recognition System Establishing by Deep Learning to Screen Heat Tolerant Cotton

Zhihao Tan

Huazhong Agriculture University

Jiawei Shi

Huazhong Agriculture University

Rongjie Lv

Huazhong Agriculture University

Qingyuan Li

Wuhan Academy of Agricultural Sciences

Jing Yang

Xinjiang Academy of Agricultural Sciences

Yizan Ma

Huazhong Agriculture University

Yanlong Li

Huazhong Agriculture University

Yuanlong Wu

Huazhong Agriculture University

Rui Zhang

Huazhong Agriculture University

Huanhuan Ma

Huazhong Agriculture University

Yawei Li

Huazhong Agriculture University

Li Zhu

Huazhong Agriculture University

Longfu Zhu

Huazhong Agriculture University

Xianlong Zhang

Huazhong Agriculture University

Jie Kong

Xinjiang Academy of Agricultural Sciences

Wanneng Yang

Huazhong Agriculture University

Ling Min (✉ lingmin@mail.hzau.edu.cn)

Huazhong Agricultural Univeristy <https://orcid.org/0000-0003-4278-0626>

Research

Keywords: Cotton anther, Deep learning, Faster R-CNN, YOLOv5, Model ensemble, High temperature stress

Posted Date: December 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1128527/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Fast anther dehiscence status recognition system establishing by deep learning to
screen heat tolerant cotton**

Zhihao Tan^{1,#}, Jiawei Shi^{1,#}, Rongjie Lv^{1,#}, Qingyuan Li², Jing Yang³, Yizan Ma¹,
Yanlong Li¹, Yuanlong Wu¹, Rui Zhang¹, Huanhuan Ma¹, Yawei Li¹, Li Zhu¹, Longfu
Zhu¹, Xianlong Zhang¹, Jie Kong^{3,*}, Wanneng Yang^{1,*}, and Ling Min^{1,*}

Affiliations

¹ National Key Laboratory of Crop Genetic Improvement, Huazhong
Agricultural University, Wuhan, Hubei 430070, China

² Forestry and Fruit Tree Research Institute, Wuhan Academy of Agricultural
Sciences, Wuhan 430075, China

³ Institute of Economic Crops, Xinjiang Academy of Agricultural Sciences,
Xinjiang 830091, China

[#] These authors contributed equally

^{*} To whom correspondence should be addressed.

Tel: +86-2787283955.

E-mails: lingmin@mail.hzau.edu.cn, ywn@mail.hzau.edu.cn, _____ and
kongjie.258@163.com

Abstract

Background

Cotton is one of the most economically important crops in the world. The fertility of male reproductive organs is a key determinant of cotton yield. The anther dehiscence or indehiscence directly determine the probability of fertilization in cotton. Thus, the rapid and accurate identification of cotton anther dehiscence status is important for judging anther growth status and promoting genetic breeding research. The development of computer vision technology and the advent of big data have prompted the application of deep learning techniques to agricultural phenotype research. Therefore, two deep learning models (Faster R-CNN and YOLOv5) were proposed to detect the number and dehiscence status of anthers.

Result

The single-stage model based on YOLOv5 has higher recognition efficiency and the ability to deploy to the mobile end. Breeding researchers can apply this model to terminals to achieve a more intuitive understanding of cotton anther dehiscence status. Moreover, three improvement strategies of Faster R-CNN model were proposed, the improved model has higher detection accuracy than YOLOv5 model. We have made four improvements to the Faster R-CNN model and after the ensemble of the four models, R2 of “open” reaches 0.8765, R2 of “close” reaches 0.8539, R2 of “all” reaches 0.8481, higher than the prediction result of either model alone, and can completely replace the manual counting method. We can use this model to quickly extract the dehiscence rate of cotton anther under high temperature (HT) condition. In

addition, the percentage of dehiscent anther of randomly selected 30 cotton varieties were observed from cotton population under normal conditions and HT conditions through the ensemble of Faster R-CNN model and manual observation. The result showed HT varying decreased the percentage of dehiscent anther in different cotton lines, consistent with the manual method.

Conclusions

The deep learning technology first time been applied to cotton anther dehiscence status recognition instead of manual method to quickly screen the HT tolerant cotton varieties and can help to explore the key genetic improvement genes in the future, promote cotton breeding and improvement.

Keywords

Cotton anther, Deep learning, Faster R-CNN, YOLOv5, Model ensemble, High temperature stress

Background

Cotton is an economically important crop, and its reproductive development is susceptible to a variety of adverse stresses that affect its yield and quality. The reproductive organs of cotton include stamens and pistils, and stamens are more sensitive to heat stress than female organs (Peet et al., 1998). In many summer crops, reproductive organ abortion caused by high temperatures (HT) is manifested by normal development of the female reproductive system and abnormal development of the male reproductive system, failure to produce functional pollen or failure of the

anthers to achieve dehiscence properly to release pollen. Anther development is a complex process, going from sporogenic cells to anther dehiscence, and has been divided into 14 periods by studying a variety of male sterile mutants (Sanders et al., 1999). Anther dehiscence, the final step in anther development, includes three processes: secondary thickening of the inner wall of the anther chamber, degradation of the septum cells, and dehiscence of the cleft which ultimately allow the release of pollen (Kim et al., 2010). Therefore, anther dehiscence is directly related to the probability of fertilization in cotton. If we can obtain phenotypic data on anther dehiscence quickly and accurately and conduct genome-wide association analysis, we can easily obtain the functional genes related to the anther dehiscence. It is also important to analysis the molecular mechanism of cotton male reproductive organ respond to stresses.

In the past, the acquisition of cotton dehiscent or indehiscent anther number data from the pictures relied mainly on visual observation and manual counting, it is difficult to guarantee the accuracy of visual readings because anther growth is intermingled, resulting in unclear definition of individual anthers, and the background and foreground of anthers are easily confused. Moreover, a larger amount of anther data is needed to judge the anther growth and dehiscence status of individual plants in population under different conditions. However, it is obviously difficult to achieve this accurately and quickly with manual methods.

After 2012, the concept of deep learning was proposed, and deep learning techniques have evolved rapidly in the past few years. The YOLO series,

Faster-RCNN and single shot multibox detector (SSD) are three important deep learning neural network models (Liu et al., 2016). Faster-RCNN mainly crawls preselected boxes and then performs deep learning classification. The image detection process of Faster-RCNN includes crawling region proposal, candidate feature frame extraction, and candidate feature frame classification. The YOLO model more cleverly uses the idea of regression by taking the whole image as input, dividing it into several boxed regions, removing individual boxes with very low relevance by setting specific thresholds, and finally selecting the highest scoring region with a nonmaximum suppression algorithm. Through classification and extraction of image features and end-to-end training of deep learning models, computers can accurately detect specific content in images. Through the building of different datasets and the replacement of deep learning network architectures, researchers can obtain network models that are more suitable for research purposes than previous approaches.

The application of target detection technology to agriculture using machine learning has been very extensive (Barre et al., 2017; Fuentes et al., 2017; Ubbens and Stavness, 2017; Gutierrez A et al., 2019). In maize, a parabolic model has been used to mine the diversity of stem-end meristematic tissues and to find candidate genes that correlate with the transport of phytohormones, cell division, and cell size by GWAS (Yang et al., 2007). In rice, the ratio of spikes to leaves, a new trait of rice, has been extracted using a feature pyramid network mask model that has achieved leaf and spike recognition accuracies of 0.98 and 0.99, respectively (Yang et al., 2020). Ferentinos KP has designed a convolutional neural network model to solve the

problem of early plant disease detection. Through the deep learning method, several model structures have been trained with plant leaf images and have identified the corresponding plant leaf lesions with 99.53% accuracy. The model has become a powerful tool for the early diagnosis and early warning of plant leaf diseases and can be further improved. Therefore, the system can be used in real time in a real cultivation environment ([Ferentinos, 2018](#)). Ubbens JR et al. have designed an open-source deep learning tool called Deep Plant Phenomics for plant phenotypic deep learning. This tool provides pretrained neural networks for several common plant phenotypic tasks including leaf counting, image classification and age regression. Botanists can use the neural networks provided and trained by this platform to train their plant phenotypes ([Ubbens and Stavness, 2017](#)). Nikita Genze et al. have proposed a convolutional neural network-based seed germination status recognition system that can automatically identify seed categories (including maize, rye, and pearl millet) in petri dishes and automatically determine whether the seeds are germinating. The system achieves an average accuracy of 94% on test data and can help seed researchers to better determine seed quality and performance ([Genze et al., 2020](#)). Scientists use hyperspectral imaging technology to collect spectral and image information from maize seeds and combine convolutional neural networks and support vector machines to model and train spectral data sets and image data sets. This model can quickly detect the vigor state of seeds and simultaneously predict their germination status, providing a framework to advance research on seed germination ([Pang et al., 2020](#)). A MobileNetv2-YOLOv3-based model that combines pretraining

methods such as hybrid training and migration learning to improve its generalization for the early identification of tomato leaf spot disease has been proposed (Liu and Wang, 2021). Image processing and machine learning techniques have been used to accurately classify the three stages of plant growth and soil for different germplasms of two species of red clover and alfalfa. The accuracy on test data was shown to be more than 90% (Samiei et al., 2020). However, no reports of machine learning-based anther identification systems in academia, which give us great incentive to build a deep learning-based anther identification system for cotton.

In this study, using YOLOv5 (Redmon et al., 2016; Redmon and Farhadi, 2017; Redmon and Farhadi, 2018; Pang et al., 2020; Xu et al., 2020) and Faster R-CNN (Ren et al., 2017), combined with a variety of data augmentation methods, a cotton anther recognition model based on deep learning was obtained. This model can quickly recognize the batch input cotton anther images, detect the dehiscent and indehiscent anthers, and obtain the phenotypic data. Using this model to detect 30 randomly selected cotton varieties, it was found that high temperature (HT) could significantly reduce the anther dehiscent rate, which can be used as a basis for screening HT tolerant germplasms and help to locate HT tolerant genes.

Results

YOLOv5 model design

YOLOv5 is a typical one-stage detection model, which increases the detection speed by 50% compared with the previous generation YOLOv4, and its model size is only

1/10 of that of the YOLOv4. The adaptive anchor frame calculation and the use of focus structure enhanced the accuracy of the model for small target recognition. At the same time, the model has four network models with different depths, and the best balance between detection accuracy and recognition speed can be found. It is very common for cotton anthers to block each other in the image, and the obscured anthers are easily ignored in the final output of the prediction box. To screen the prediction box, usually used the NMS or soft-NMS algorithm. The idea of the NMS algorithm is that for a certain category X having N candidate boxes, the candidate boxes are sorted by their confidence, and the highest confidence box A is selected. The other candidate boxes B_i ($i=1, 2, 3\dots$) are compared with the highest confidence box A, and an IoU threshold is set. If its IoU is higher than this threshold, the candidate box B1 is discarded. Then the candidate box B2's IoU is compared with that of the highest confidence box A. After several iterations, only prediction boxes that have an IoU lower than the set IoU value are retained. Although this method can prevent the same target from being repeatedly selected by multiple prediction boxes, it cannot prevent overlapping or occluded targets from being ignored.

The idea of Soft-NMS is that M is the current highest scoring box and B_i is the pending box. The larger the IoU of B_i and M, the more the score S_i of B_i drops, rather than having the score go directly to zero as in NMS. This method can effectively retain anther images that overlap and ensure the accuracy of identification results. The linear weighting formula for Soft-NMS can be expressed as:

$$S_i = \begin{cases} S_i & IOU(M, b_i) < N_t \\ S_i(1 - IOU(M, b_i)) & IOU(M, b_i) \geq N_t \end{cases}$$

Thus, to screen the prediction box, when using the NMS algorithm, only the anther images with the highest confidence are retained. Therefore, we used YOLOv5 with the soft-NMS algorithm (Bodla et al., 2017) to screen the prediction box.

Faster R-CNN model design

Faster R-CNN is a classical two-stage object detection network. The network model structure is mainly composed of four parts: feature extraction, region proposal, classification, roi pooling and its comprehensive performance has been greatly improved, especially for the detection accuracy of small targets. The cotton anther belongs to the range of small target detection in the whole image, so we trained the Faster R-CNN model to identify the anther dehiscence state have a better detection effect.

Conv layers is a classical CNN network target detection method, mainly includes three layers of conv, pooling, relu, usually uses to extract the feature maps of the input image. The extracted feature maps will be called by subsequent region proposal networks and classification networks. In the convlayers structure, contains 13 conv layers, 13 relu layers, 4 pooling layers. The Faster R-CNN has a very ingenious detail in the convlayers, it does augmentation treatment on all convolutional layers, fills a layer in the outer layer of the input matrix, so that the matrix is larger than it was, and the images that have been treated in this way are deconvoluted again, and after the convolution operation, the image is kept consistent with the size of the input image. The matrix size is unchanged when the image goes through the conv layer and relu layer, and will change to 1/2 of the original size after going through the pooling layer,

so that when going through the convlayers structure, the size of the input matrix changes to $1/16$ of the original size, thus the resulting feature maps can all correspond one-to-one with the original graph.

Conventional detection methods usually use a sliding window or the selective search method to acquire detection frames, whereas Faster R-CNN discards traditional methods and directly generates detection frames using region proposal networks, which greatly enhances the detection frame generation speed. The region proposal network structure is actually divided into two processes, the first process by softmax classification anchors, to obtain foreground and background (detection target is foreground), the second process is used to calculate the bounding box regression offset for anchors to obtain the exact proposal. Finally, the proposal layer is responsible for integrating foreground anchors and bounding box regression offset to obtain proposals, while simultaneously removing proposals too small and beyond the boundary. The entire Faster R-CNN network arrives to proposal layer, completing detection targets, the next two structures are mainly image recognition.

For the traditional CNN network, the input image of the model must be fixed size, and the output of the model must be a fixed vector or matrix. In practical applications, there are two solutions for images of different sizes: cut the picture to a fixed size or warp the image to a fixed size. However, these solutions will either cause the loss of image information, or lead to changes in the shape information of the image. Therefore, the structure roi pooling is proposed in Faster R-CNN to solve the problem of different image size. Roi pooling is mainly responsible for collecting feature maps

and proposal boxes, calculating proposal feature maps, and sending it to the subsequent identification layer. First of all, proposal is mapped to the same scale as feature maps, and then the vertical and horizontal directions of each proposal are divided into seven parts, so that the output of different sizes of proposal is 7*7, realizing fixed-length output.

Classification using the obtained proposal feature maps, the structure calculates which category each proposal belongs to through full connect layers and softmax, and outputs the probability vector; at the same time, the position offset of each proposal is obtained again by bounding box regression, which is used to return a more accurate target detection box.

The loss function of the object detection network of Faster R-CNN is shown in Formula:

$$L_{reg}(t_i, t_i^*) = \sum_{i \in \{x, y, w, h\}} smooth_{L1}(t_i - t_i^*)$$

$$smooth_{L1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

In the above-mentioned formula, i represents anchors index; t represents predict bounding box; t* represents ground true box corresponding to positive anchor; xy, w, h represents center point coordinates of box, width, height, respectively.

Data augmentation

Auto augment

This approach creates a search space for data-enhanced policies in which a policy contains many sub-policies and randomly selects one sub-policy for each image in a

small batch data set. Each sub strategy consists of two operations, each of which is an image processing function similar to flatting, rotation, or shearing, and the probability and magnitude of applying those functions, using a search algorithm directly on the data set to find the best data augmentation strategy.

Random Resize

Random Resize scales the new image to the same pixel size as the original image by randomly clipping the original image in the data set according to the random aspect ratio.

Random Flip

Random Flip is a common way of data augmentation, which generates new data set samples by randomly flipping the original image of the data set up and down or left and right.

Mixup

Mixup is a data augmentation method for mixing two samples and label data at their corresponding ratios and then generating new sample and label data. Suppose x_1 is a sample of batch one, y_1 is the label corresponding to the sample of batch one; x_2 is the sample of batch two, y_2 is the sample corresponding label of batch two, x_{mix} and y_{mix} is the newly generated sample and corresponding label, respectively. λ is the mixing coefficient resulting from the hyperparametric α and β conducted beta distributions.

The principal formula of the mixup method can be expressed as:

$$x_{mix} = \lambda x_1 + (1 - \lambda) x_2$$

$$y_{mix} = \lambda y_1 + (1 - \lambda) y_2$$

265 $\lambda \sim \text{Beta}(\alpha, \beta)$ $\alpha, \beta \in [0, +\infty]$

266 According to the study, we know that as the hyperparameters α and β increase, the
267 error and generalization ability of the network training will increase. When the beta
268 distribution of the mixing coefficient λ is $\alpha=\beta=0$, the network recovers to the ERM
269 (Empirical Risk Minimization) principle to minimize the training data average error;
270 the beta distribution of the mixing coefficient λ has the best generalization ability and
271 robustness. This method can make full use of all the pixel information, but at the same
272 time also introduces some unnecessary pseudo-pixel information.

273 **Cutmix**

274 Cutmix (Yun et al., 2019) is by cut some regions in the sample and randomly fill in
275 the pixel values of other samples in the data set, and at the same time distribute the
276 final classification results according to a certain proportion. Compared with mixup,
277 cutmix can prevent the occurrence of non-pixel information in the training process.
278 Filling the pixel information of other regions with the missing area of cut can further
279 enhance the positioning ability of the model. At the same time, this method will not
280 increase the training and reasoning burden of the model.

281 **GridMask**

282 By generating a mask with the same resolution as the original image, GridMask
283 multiplies the mask with the original image to get a new image. The pixel value of the
284 new image in the fixed area is 0, which is essentially a regularization method.
285 Compared with directly changing the network structure, GridMask only needs to be
286 augmented when the image is input.

Normalized

We usually use this method after data augmentation, normalizing the pixel value of the image and scaling the pixel value to $[0, 1]$ can prevent the attributes of the large value interval from excessively dominating the attributes of the decimal value interval, and at the same time avoid the numerical complexity in the process of calculation.

In this study, the data augmentation process was shown in the [Figure 1](#).

Model training

In this study, comparative experiments and control variables between YOLOv5 and Faster-RCNN models were used, and various data demonstration methods such as mixing and mixed cutting were used to train for sample imbalance, so as to verify the performance of different models and training methods on the same evaluation index validation set. Firstly, the self-made data set was segmented and analyzed, and VOC format was used to store training set, test set and verification set. Secondly, the model was trained according to whether the data demonstration algorithm was added. Finally, cosine strategy was used to periodically attenuate the learning rate. The training ends when the average loss remains stable. In this study, the training process of Faster R-CNN model was shown in the [Figure 2](#).

The models obtained by different training strategies were tested on the test set and the prediction results of multiple models were obtained. The results of the four groups of comparison experiments indicated that the proposed Faster R-CNN neural network with data augmentation and FPN structure and Multi-Scale could effectively detect

the dehiscence and indehiscence in cotton anther images. Compared with other methods, this method has significant advantages in recognition accuracy. The recognition effect was shown in the [Figure 3](#). The final result was obtained by the prediction results ensembles of multiple models.

Model comparison

Comparison of detection results of Faster R-CNN and YOLOv5

Faster R-CNN and YOLOv5 are used to train the same training set, the test results are compared on the same test set, and the correlation between the test results and the accurate value of manual labeling is analyzed. YOLOv5 using Darknet53 as the backbone network is a typical single-stage model, while Faster R-CNN using Res101 as the backbone network is a standard two-stage model. Obviously, YOLOv5 is more advantageous in detection speed. A comparison of the two models was shown in [Figure 4a](#). Through training and validation, we found that $mAP@0.5:0.95$ of YOLOv5 was 0.485, while $mAP@0.5:0.95$ of Faster R-CNN was 0.478. In $mAP@0.5:0.95$, YOLOv5 was 0.007 higher than Faster R-CNN. In terms of the evaluation index of R^2 in the validation set, the Faster R-CNN was 0.8712 in the category of "open" and 0.8373 in the category of "close", and 0.82 in the category of "all", which were 0.2523, 0.2619, and 0.3104 higher than YOLOv5, respectively. This may be due to the interference of location information. Although YOLOv5 has a slightly higher $mAP@0.5:0.95$, R^2 is far lower than Faster R-CNN ([Table S1](#)). Since quantitative accuracy is our primary research goal, we decided to further optimize the two-stage model Faster R-CNN.

Comparison of detection results with or without FPN (Feature Pyramid Networks)

To further improve the detection effect of the Faster R-CNN model, the FPN structure was added into the Faster R-CNN model. A comparison of the two models is shown in [Figure 4b](#). The mAP@0.5:0.95 of Faster R-CNN with data augmentation was 0.48. For the R^2 of the correlation of test value with the real value, Faster R-CNN with FPN structure have 0.8676, 0.8403 and 0.812 in the category of "open", "close", and "all". Compared with that without FPN structure, mAP@0.5:0.95 of the improved model increased by 0.002 ([Figure 5, model 1 and 3](#)), R^2 of "close" class increased by 0.003, and R^2 of "open" class and "all" decreased slightly ([Table S2](#)).

Comparison of detection results with or without data augmentation

The traditional Faster R-CNN model was taken without data augmentation. To avoid the effect of the sample imbalance, many kinds of data augmentation methods were added to the basic model, such as mixup, cutmix. The model with and without data augmentation were trained and tested on the same data set, and these detection results and correlations with the real value of manual labeling were compared. A comparison of the two models is shown in [Figure 4c](#). We found that the mAP@0.5:0.95 of Faster R-CNN with data augmentation was 0.494, which was 0.016 higher than that of Faster R-CNN without data augmentation ([Figure 5, model 1 and 2](#)). For the R^2 of the correlation of test value with the real value, Faster R-CNN with data augmentation were 0.8579, 0.8401 and 0.8235 in the category of "open", "close", and "all", respectively. The R^2 in the category of "close" and "all" of Faster R-CNN

with data augmentation were 0.0028 and 0.0035 higher than that of Faster R-CNN without augmentation. While R^2 in the "open" category of Faster R-CNN with data augmentation was 0.0133 lower than that of Faster R-CNN without data augmentation. Overall, the evaluation showed that the performance of Faster R-CNN with data augmentation is higher than that of Faster R-CNN without data augmentation (Table S3).

Comparison of detection results with or without Multi-Scale

To test whether the multi-scale training can improve the accuracy of the quantity of dehiscent anther, we added Multi-Scale on the basis of the traditional Faster R-CNN model. The specific content is to obtain the image pyramid at different scales, and then extract the features of different scales for each layer of images to obtain the feature map. Finally, the features of each scale are individually predicted. A comparison of the two models was shown in Figure 4d. The result showed the mAP@0.5:0.95 of model was improved by 0.003 after Multi-Scale training (Figure 5, model 4 and 2). R^2 in the "open" and "close" categories fell by 0.0092 and 0.0007, respectively. R^2 in the "all" category went up 0.0086. Thus, Multi-Scale training has a certain effect on our research goal of cotton anther identification (Table S4).

In this study, the change curves of each model in mAP@0.5:0.95 during the training process were shown in Figure 5. The peak value of traditional Faster-CNN mAP@0.5:0.95 curve was the lowest, while the peak value of Faster R-CNN model with data augmentation, Multi-Scale training and FPN structure was the highest. The loss curve of each model during the training process was shown in Figure 5. At the

end of the training, the loss curve of the four models has tended to be stable.

Screening of HT tolerant cotton germplasms based on cotton anther phenotype data obtained using integrated Faster R-CNN model

In order to select high temperature (HT) tolerant cotton germplasms, the anther images of different cotton lines were obtained under normal temperature (NT) and HT. Then we counted the dehiscence status of anthers from 30 different cotton lines by manual observation and machine recognition, and the statistical results were shown in Table 1. The results of manual observation showed that the average dehiscence rate of cotton anthers treated with NT and HT were 84.35% and 35.46%, respectively. The results of machine recognition showed that the average dehiscence rates of cotton anthers treated with NT and HT were 83.81% and 35.08%, respectively. First of all, we can believe that for the acquisition of the phenotypic data of cotton anther dehiscence rate, the result of machine recognition has been extremely accurate, and the recognition speed is fast, which is not affected by artificial subjective factors, save manpower and material resources, there are obvious advantages compared with manual observation. Secondly, there is a great difference in the anther dehiscence rate of the same cotton variety between HT and NT conditions. The result showed that HT greatly reduced the cotton anther dehiscence rate ([Table 1](#)), and then affected the pollination process, resulting in a reduction in cotton yield. Finally, by observing 30 cotton lines, we found that the anther dehiscence rate of S003 and S004 was still more than 85% under HT stress, which was significantly improved compared with other lines ([Table 1](#)). In addition, we screened cotton lines with HT tolerance in large

quantities through machine recognition, and obtained more than 35 HT tolerant cotton lines. These HT tolerant germplasms will be used in cotton HT tolerance breeding.

Discussion

Through analysis, we found that the mAP@0.5:0.95 value of the model increased significantly after adding data augmentation and FPN structure and Multiscale, but the change of R² was not significantly positively correlated with mAP@0.5:0.95. In order to obtain the most accurate data in the application, four models were trained, as shown in Figure 5 and tested them on the same batch of test sets. The recognition results obtained were integrated by the following formula:

$$\text{result}_{\text{open}} = \frac{\sum_{i=1}^4 \text{model}_i^{\text{open}}}{4}$$

$$\text{result}_{\text{close}} = \frac{\sum_{i=1}^4 \text{model}_i^{\text{close}}}{4}$$

Among those, i represents the number of the model in Figure 5. $\text{Model}_i^{\text{open}}$ represents the number of dehiscent cotton anthers identified by model_i in the verification set. $\text{Model}_i^{\text{close}}$ represents the number of indehiscent cotton anthers identified by model_i in the verification set.

After the comparison with the real value, it is found that when the model is integrated, the detection result after ensemble effectively compensates for the error, and the correlation between the detection result and the real value will increase. After the ensemble of the four models, R² of “open” reaches 0.8765, R² of “close” reaches

0.8539, R^2 of “all” reaches 0.8481, higher than the prediction result of either model alone. Therefore, when accurate data is needed, we can choose to integrate the detection results of the four models, so that the detection data obtained is the most reliable. Of course, directly using Faster R-CNN model with FPN structure and data augmentation and multi-scale has higher robustness and higher accuracy.

It is well known that anther is the male organ of plant, anther abortion will directly lead to male sterility and reduce yield. Our previous studies could be preliminarily concluded that HT stress can reduce cotton yield by inhibiting cotton male fertility. HT mainly decreased pollen viability, the anther growth number, and the percentage of dehiscent anther, caused the decreases of male fertility in cotton (Min et al., 2014; Ma et al., 2018). Furthermore, with the development of sequencing technology, a large amount of cotton germplasm resequencing data and transcriptome variation data have been obtained (Wang et al., 2017; Ma et al., 2021). However, no enhancing male reproductive organs HT tolerance genes have been cloned. The main reason is that it is difficult to obtain phenotypes of reproductive organs. Thus, in this study, we built the trained augmentation Faster R-CNN rapid identification system of cotton anther phenotype, can quickly investigate the anther phenotype, and can use to location of the genes affecting cotton anther dehiscence under HT by combining the genome-wide association study and whole transcriptome association study. This will effectively promote cotton HT tolerance breeding and ensure cotton safe production under the trend of global warming.

Conclusions and future directions

Conclusions

1. A cotton anther phenotype recognition system based on deep learning is proposed for the first time, which can help researchers to quickly investigate the anther phenotype of cotton and locate the genes that respond to the influence of stress on cotton anther for breeding improvement.

2. A lightweight cotton anther dehiscence detection model based on YOLOv5 is proposed, which can be easily implanted into embedded devices or mobile devices.

3. Through the change of the accuracy and correlation of Faster R-CNN after the improvement of the data augmentation method, the feasibility and superiority of the improved method were verified.

4. After the ensemble of the four models, R^2 of “open” reaches 0.8765, R^2 of “close” reaches 0.8539, R^2 of “all” reaches 0.8481, higher than the prediction result of either model alone, and can completely replace the manual counting method. It provides new technical support for cotton reproductive development and HT tolerance breeding.

Future directions

In this study, YOLOv5 and Faster R-CNN were applied to identify the dehiscence status of cotton anther, and achieved fast and accurate identification. But there are still some areas where there is room for improvement:

1. We examined the dehiscence of cotton anthers, but other phenotypes such as the growth position of anthers and the distance between anthers and stigmas are also

important for the cotton fertility under HT. Other phenotypic characteristics of cotton anther can be collected by using a comprehensive platform that integrates multiple data points to analyze cotton reproductive development.

2. The cotton anther dehiscence recognition model trained in this study should be further developed and applied to other computer platforms or servers to facilitate cotton reproductive development researchers to use the model to quickly obtain anther dehiscence data.

3. In this study, the experience of deep learning model training for cotton anther dehiscence can be applied to other plant anther state detection. It is one of the directions to further enrich the research content to further construct cotton anther state recognition models of various crops based on deep learning.

Materials and methods

Material growing and dataset acquisition

In total, 510 cotton lines from natural populations were planted in 2016-2019 in experimental cotton fields at Huazhong Agricultural University, Wuhan, Hubei (113.41E, 29.58N), Turpan, Xinjiang (89.19E, 42.91N), and Alar, Xinjiang (81.29E, 40.54N). At Wuhan, the field was planted at a density of 27,000 plants per hectare with each row including more than 12 individuals. At Alar and Turpan, Xinjiang, the fields were set up with two streets and planted at a density of 195,000 plants per hectare. More than 30 individuals of each line were arranged in rows. Cotton anther images were collected each year at each location three days after the onset of normal

temperatures and after high temperatures during bloom.

A Canon 70 d HD digital camera was used throughout the acquisition of a research image set. To prevent the negative interference of background with the subsequent machine recognition effort, a black curtain was used as the photo background for the experiments. In the actual image collection process, it was found that the cotton anthers were surrounded by cotton petals, and the anthers growing at the root of the style were not easily captured by the camera, so it was not conducive to the accurate collection of data by taking the pictures directly. Thus, it was necessary to preprocess the cotton flowers before acquiring the pictures by stripping the cotton petals and fixing the anther sides. To prevent overfitting and overcome insufficient training data, the same anthers were included in multiple distant near-field images (Figure 6). Finally, a total of 38,895 high-definition RGB whole anther images were acquired.

Morphologically, dehiscent anthers are rough and grainy because pollen is released and adheres to anther edges, while indehiscent anthers have smooth edges, because no pollen is released. Therefore, the obtained cotton anther images were annotated using “Labelimg” image annotation software, as shown in Figure 7. The image boundary of each visible cotton anther is within an annotation box that reduces the influence of background on model training and is labeled “open” and “close” to distinguish dehiscent and indehiscent anthers, respectively. A total of 2,845 images were annotated one by one. The images were used as the data set and were randomly divided into a training set and validation set in the ratio of 7:3 (Table S5).

Experimental operation environment

The hardware environment used in this study is in [Table S6](#), and on this basis, the training environment is python, open-cv, cuda, etc., and the frameworks used in this study are paddle and pytorch.

Metrics were used to evaluate the proposed method

In this study, we used mAP@0.5:0.95, as well as MAD (Mean Absolute Deviation) and R^2 as the evaluation indicators of the model. The indicators are explained as follows :

mAP@0.5:0.95 is the process of increasing intersection over union (IoU) from 0.5 to 0.95 according to the span of 0.05. The mAP corresponding to each IoU is added to obtain the average value of mAP in this process. The formula is expressed as follows:

$$P = \frac{T_P}{P_N}$$

$$R = \frac{T_P}{T_N}$$

$$AP = \int_0^1 P(R) dR$$

In the above formula, T_P is the correct number of a category identified by the model, P_N is the total number of categories identified by the model, and T_N is the true number of a category. Averaging the AP values of all categories is called mAP.

We took the absolute value of the absolute error between the measured value and the real value and then calculated the average value, called it MAD. Due to the deviation is absolute value, the positive and negative will not be offset, thus the mean absolute error can reflect the actual situation of predicted value deviation. The smaller

the value, the closer the prediction of model is to reality.

The main purpose of this study is to develop a deep learning model that can quickly and accurately identify anther dehiscence and explore the influence of high temperature stress on cotton anther dehiscence. In the model identification phase, we identify the location of the cotton anther without strict requirements, but we demand model to recognize the number with the result of artificial observation is the closest to, so the anther number by artificial observation was used as accurate value of validation set, with correlation coefficient between predicted values and the accurate value as the main evaluation index of the model.

List of abbreviations

AUPRCs: areas under precision/recall curves

ERM: empirical risk minimization

FPN: feature pyramid networks

HT: high temperature

lncRNAs: long noncoding RNAs

MAD: mean absolute deviation

NLP: natural language processing

NT: normal temperature (NT)

RPN: region proposal network

SSD: single shot multibox detector

SVM: support vector machine

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The data set offers the dehiscence of cotton anthers under high temperature stress, so data is temporarily unavailable. As the experiment progresses, our data will gradually improve the available state.

Competing interests

The authors declare that they have no competing interests.

Funding

Project supported by the Fundamental Research Funds for the Central Universities (2021ZKPY019), the National Natural Science Foundation of China (32072024), the Platform Construction of Genetic Improvement and Molecular Design Breeding for Xinjiang Island Cotton (2020172-2).

Authors' contributions

Z.T. and R.L carried out the experiment with J.S using the Faster R-CNN and YOLOv5, Z.T, R.L, and L.M. wrote the main manuscript text, and H.M, L.Z, Y.L, J.Y, Y.W, Y.M, R.Z, and Q.L. obtained and labeled the anther pictures, L.M, X.Z, W.Y, L. Z, and J.K. designed and supervised the research, L.M, X.Z, J.K, and W.Y revised the manuscript. L.M agrees to serve as the author responsible for contact and ensures

communication. All authors reviewed the manuscript.

Acknowledgements

Appreciations are given to the editors and reviewer.

References

Barre P, Stover BC, Muller KF, Steinhage V (2017) LeafNet: A computer vision system for automatic plant species identification. *Ecological Informatics* 40: 50-56

Bodla N, Singh B, Chellappa R, Davis LS (2017) Soft-NMS: improving object detection with one line of code. 2017 Ieee International Conference on Computer Vision (Iccv): 5562-5570

Ferentinos KP (2018) Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture* 145: 311-318

Fuentes A, Yoon S, Kim SC, Park DS (2017) A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors* 17: 153

Genze N, Bharti R, Grieb M, Schultheiss SJ, Grimm DGJPM (2020) Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops. 16: 1-11

Gutierrez A, Ansuategi A, Susperregi L, C Tubío, Lena L (2019) A Benchmarking of Learning Strategies for Pest Detection and Identification on Tomato Plants for Autonomous Scouting Robots Using Internal Databases. *Journal of Sensors* 2019: 1-15

Kim SG, Lee S, Kim YS, Yun DJ, Woo JC, Park CM (2010) Activation tagging of an

592 Arabidopsis SHI-RELATED SEQUENCE gene produces abnormal anther
 593 dehiscence and floral development. *Plant Molecular Biology* 74: 337-351
 594 Liu J, Wang XW (2021) Early recognition of tomato gray leaf spot disease based on
 595 MobileNetv2-YOLOv3 model (vol 16, 83, 2020). *Plant Methods* 17: 16
 596 Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) SSD:
 597 single shot multiBox detector. *Computer Vision - Eccv 2016, Pt I* 9905: 21-37
 598 Ma YZ, Min L, Wang JD, Li YY, Wu YL, Hu Q, Ding YH, Wang MJ, Liang YJ, Gong
 599 ZL, Xie S, Su XJ, Wang CZ, Zhao YL, Fang QD, Li YL, Chi HB, Chen M,
 600 Khan AH, Lindsey K, Zhu LF, Li XY, Zhang XL (2021) A combination of
 601 genome-wide and transcriptome-wide association studies reveals genetic
 602 elements leading to male sterility during high temperature stress in cotton.
 603 *New Phytologist* 231: 165-181
 604 Ma YZ, Min L, Wang MJ, Wang CZ, Zhao YL, Li YY, Fang QD, Wu YL, Xie S, Ding
 605 YH, Su XJ, Hu Q, Zhang QH, Li XY, Zhang XL (2018) Disrupted genome
 606 methylation in response to high temperature has distinct affects on microspore
 607 abortion and anther indehiscence. *Plant Cell* 30: 1387-1403
 608 Min L, Li YY, Hu Q, Zhu LF, Gao WH, Wu YL, Ding YH, Liu SM, Yang XY, Zhang
 609 XL (2014) Sugar and auxin signaling pathways respond to High-Temperature
 610 stress during anther development as revealed by transcript profiling analysis in
 611 cotton. *Plant Physiology* 164: 1293-1308
 612 Pang L, Liu H, Chen Y, Miao J (2020) Real-time Concealed Object Detection from
 613 Passive Millimeter Wave Images Based on the YOLOv3 Algorithm.

SENSORS-BASEL 20: 1678

- Pang L, Men S, Yan L, Xiao J (2020) Rapid vitality estimation and prediction of corn seeds based on spectra and images using deep learning and hyperspectral imaging techniques. *IEEE Access* 8: 123026-123036
- Peet M, Sato S, Gardner R (1998) Comparing heat stress effects on male-fertile and male-sterile tomatoes. *Plant cell* 21: 225-231
- Redmon J, Divvala S, Girshick R, Farhadi A (2016) You Only Look Once: unified, real-Time object detection. 2016 *Ieee Conference on Computer Vision and Pattern Recognition (Cvpr)*: 779-788
- Redmon J, Farhadi A (2017) YOLO9000: Better, Faster, Stronger. 30th *Ieee Conference on Computer Vision and Pattern Recognition (Cvpr 2017)*: 6517-6525
- Redmon J, Farhadi A (2018) Yolov3: An incremental improvement. *arXiv preprint arXiv: 02767*
- Ren SQ, He KM, Girshick R, Sun J (2017) Faster R-CNN: Towards real-time object detection with region proposal networks. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 39: 1137-1149
- Samiei S, Rasti P, Vu JL, Buitink J, Rousseau D (2020) Deep learning-based detection of seedling development. *Plant Methods* 16: 11
- Sanders PM, Bui AQ, Weterings K, McIntire KN, Hsu YC, Lee PY, Truong MT, Beals TP, Goldberg RB (1999) Anther developmental defects in *Arabidopsis thaliana* male-sterile mutants. *Sexual Plant Reproduction* 11: 297-322

Ubbens JR, Stavness I (2017) Deep Plant Phenomics: A deep learning platform for complex plant phenotyping tasks. *Frontiers in Plant Science* 8: 1190

Wang MJ, Tu LL, Lin M, Lin ZX, Wang PC, Yang QY, Ye ZX, Shen C, Li JY, Zhang L, Zhou XL, Nie XH, Li ZH, Guo K, Ma YZ, Huang C, Jin SX, Zhu LF, Yang XY, Min L, Yuan DJ, Zhang QH, Lindsey K, Zhang XL (2017) Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nature Genetics* 49: 579

Xu ZF, Jia RS, Sun HM, Liu QM, Cui Z (2020) Light-YOLOv3: fast method for detecting green mangoes in complex scenes using picking robots. *Applied Intelligence* 50: 4670-4687

Yang CY, Xu ZY, Song J, Conner K, Barrena GV, Wilson ZA (2007) Arabidopsis MYB26/MALE STERILE35 regulates secondary thickening in the endothecium and is essential for anther dehiscence. *Plant Cell* 19: 534-548

Yang ZF, Gao S, Xiao F, Li GH, Ding YF, Guo QH, Paul MJ, Liu ZH (2020) Leaf to panicle ratio (LPR): a new physiological trait indicative of source and sink relation in japonica rice based on deep learning. *Plant Methods* 16: 15

Yun S, Han D, Oh SJ, Chun S, Choe J, Yoo Y (2019) CutMix: Regularization strategy to train strong classifiers with localizable features. *2019 IEEE/CVF International Conference on Computer Vision (ICCV 2019)*: 6022-6031

Figure legends

Figure 1: Data augmentation

The above images showed the effect of different data augmentation methods on the

same cotton anther image.

Figure 2: Model ensembles

Integrated flow chart of cotton anther recognition model ensembles.

Figure 3: Cotton anther identification effect graph

a: The purple box marks an indehiscent cotton anther, and the pink box marks a dehiscent cotton anther.

b: The blue box marks an indehiscent cotton anther, and the gray box marks a dehiscent cotton anther.

c: The pink box marks an indehiscent cotton anther, and the green box marks a dehiscent cotton anther.

d: The gray box marks an indehiscent cotton anther, and the red box marks a dehiscent cotton anther. In each test, the colors of the prediction boxes with different labels were randomly generated.

Figure 4: Comparison of different models

a: Comparison of YOLOv5 and Faster R-CNN. YOLOv5 model has higher recognition speed, Faster R-CNN model has higher detection accuracy.

b: Comparison of with or without FPN (Feature Pyramid Networks)

The mAP@0.5:0.95 of the improved model increased by 0.002, R^2 of "close" class increased by 0.003, and R^2 of "open" class and "all" decreased slightly.

c: Comparison of with or without data augmentation. The improved model has a slight decline in the number of R^2 in the open category and an improvement in other evaluation indicators.

d: Comparison of with or without data Multi-Scale

The result showed the $mAP@0.5:0.95$ of model was improved by 0.003 after Multi-Scale training. R^2 in the "open" and "close" categories fell by 0.0092 and 0.0007, respectively. R^2 in the "all" category went up 0.0086. "open" and "close" represent dehiscent and indehiscent anther, respectively.

Figure 5: $mAP@0.5:0.95$ curves and LOSS curves

a: $mAP@0.5:0.95$ curves. **b:** LOSS curves. **Model 1** is the Faster R-CNN with FPN structure. **Model 2** is the Faster R-CNN with data augmentation and FPN structure. **Model 3** is the traditional Faster R-CNN. **Model 4** is the Faster R-CNN with Multi-Scale and data augmentation and FPN structure. **Epoch:** All the data were sent into the network to complete a process of forward calculation and back propagation. **$mAP@0.5:0.95$** is the process of increasing IoU from 0.5 to 0.95 according to the span of 0.05. The mAP corresponding to each IoU was added to obtain the average value of mAP in this process.

Figure 6: Data acquisition.

a: The image dataset captures the platform scene. **b:** Image of cotton anther. **c:** The surface of dehiscent cotton anther (open) is rough from the image. **d:** The surface of indehiscent cotton anther (close) is smooth from the image.

Figure 7: Image labeling

The above figures are manually marked cotton anther images using the "Labeling" software. Green boxes represent indehiscent anthers and red boxes represent dehiscent anthers. When the image labeling was finished, we corresponded the location

information of the image with the name of the image one by one and saved it in VOC format. **a:** All the anthers are indehiscent. **b:** All the anthers are dehiscent. **c:** Dehiscent anthers account for the majority. **d:** Indehiscent anthers account for the majority.

Table

Table 1: Screening of HT tolerant cotton germplasms using ensemble Faster R-CNN model

Open: the average number of dehiscent anthers from different flowers, $n \geq 4$; Close: the average number of indehiscent anthers from different flowers, $n \geq 4$; Dehiscent rate: the number of dehiscent anthers of each flower / (the number of dehiscent anthers of each flower + the number of indehiscent anthers of each flower), $n \geq 4$.

Supplementary information

Additional file 1: Table S1. Comparison of YOLOv5 and Faster R-CNN.

Additional file 2: Table S2. Comparison of FPN.

Additional file 3: Table S3. Comparison of data augmentation.

Additional file 4: Table S4. Comparison of Multi-Scale.

Additional file 5: Table S5. Dataset.

Additional file 6: Table S6. Experimental configuration.

Figures

Figure 1

Data augmentation The above images showed the effect of different data augmentation methods on the same cotton anther image.

Figure 2

Model ensembles Integrated flow chart of cotton anther recognition model ensembles.

Figure 3

Cotton anther identification effect graph a: The purple box marks an indehiscent cotton anther, and the pink box marks a dehiscent cotton anther. b: The blue box marks an indehiscent cotton anther, and the gray box marks a dehiscent cotton anther. c: The pink box marks an indehiscent cotton anther, and the green box marks a dehiscent cotton anther. d: The gray box marks an indehiscent cotton anther, and the red box marks a dehiscent cotton anther. In each test, the colors of the prediction boxes with different labels were randomly generated.

Figure 4

Comparison of different models a: Comparison of YOLOv5 and Faster R-CNN. YOLOv5 model has higher recognition speed, Faster R-CNN model has higher detection accuracy. b: Comparison of with or without FPN (Feature Pyramid Networks) The mAP@0.5:0.95 of the improved model increased by 0.002, R2 of "close" class increased by 0.003, and R2 of "open" class and "all" decreased slightly. c: Comparison of with or without data augmentation. The improved model has a slight decline in the number of R2 in the open category and an improvement in other evaluation indicators. d: Comparison of with or without data Multi-Scale The result showed the mAP@0.5:0.95 of model was improved by 0.003 after Multi-Scale training. R2 in the "open" and "close" categories fell by 0.0092 and 0.0007, respectively. R2 in the "all" category went up 0.0086. "open" and "close" represent dehiscent and indehiscent anther, respectively.

Figure 5

mAP@0.5:0.95 curves and LOSS curves a: mAP@0.5:0.95 curves. b: LOSS curves. Model 1 is the Faster R-CNN with FPN structure. Model 2 is the Faster R-CNN with data augmentation and FPN structure. Model 3 is the traditional Faster R-CNN. Model 4 is the Faster R-CNN with Multi-Scale and data augmentation and FPN structure. Epoch: All the data were sent into the network to complete a process of forward calculation and back propagation. mAP@0.5:0.95 is the process of increasing IoU from 0.5 to 0.95 according to the span of 0.05. The mAP corresponding to each IoU was added to obtain the average value of mAP in this process.

Figure 6

Data acquisition. a: The image dataset captures the platform scene. b: Image of cotton anther. c: The surface of dehiscient cotton anther (open) is rough from the image. d: The surface of indehiscient cotton anther (close) is smooth from the image.

Figure 7

Image labeling The above figures are manually marked cotton anther images using the "Labelimg" software. Green boxes represent indehiscient anthers and red boxes represent dehiscient anthers. When the image labeling was finished, we corresponded the location information of the image with the name of the image one by one and saved it in VOC format. a: All the anthers are indehiscient. b: All the anthers are dehiscient. c: Dehiscient anthers account for the majority. d: Inde hiscient anthers account for the majority.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [table1.xlsx](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)