Research on Intelligent recognition system of Cotton apical Bud based on Deep Learning

Jianliang Li¹,a, Xinlei Zhi²,b, Yingying Wang³,c, Qingzheng Cao⁴,d

¹ Beijing Zhenxing Institute of Metrology and testing, Beijing, China)
² Beijing Zhenxing Institute of Metrology and testing, Beijing, China)
³ Beijing Zhenxing Institute of Metrology and testing, Beijing, China)
⁴ Beijing Zhenxing Institute of Metrology and testing, Beijing, China)
bemail: 941988028@qq.com,cemail: ljl60046@163.com,demail: 407132036@qq.com
*aemail: ktgxzx2020@163.com

Abstract: Cotton topping is a key link in the whole cotton production process, which can eliminate the top growth advantage and promote cotton production and income. The automation of cotton topping can greatly reduce the topping time and labor intensity, and the high-speed and accurate identification of cotton top buds is the prerequisite and basis for automatic topping. This article designs a cotton top bud intelligent identification system based on YOLOv3 network, which can detect cotton top buds in visible light images in real time, and provide visual information for the subsequent realization of cotton top bud position measurement and mechanical control. A computer workstation (RTX2070s) was used to identify 15049 cotton top bud images under different weather and illumination conditions. The results show that the average time of single frame image recognition is controlled within 100 milliseconds, and the top bud recognition rate reaches 96%, which creates good conditions for the development of automatic cotton topping equipment and has broad application prospects.

1. Introduction
Cotton is an important part of my country’s economic crops and the main raw material of my country’s textile industry [1]. In the process of cotton growth, cotton topping is needed, that is, when the cotton grows to a certain height or period, the top core of the cotton is cut off to increase cotton yield. At present, cotton topping mainly relies on manual completion, with high labor intensity and low efficiency. Therefore, under the current labor shortage and high labor cost, it has important practical significance and broad development prospects to realize automatic cotton topping quickly, accurately and efficiently.

Fast and accurate cotton top bud identification is the key to automatic cotton topping [2]. The recognition of cotton top buds under complex conditions is affected by many factors: the top buds and branches and leaves overlap each other and shade each other; the light conditions in the natural environment are complex and changeable; the cotton top buds have different morphologies and are less distinguishable from branches and leaves, etc. These bring many difficulties for the rapid and accurate identification of cotton top buds.

Cotton top bud recognition essentially belongs to target detection. At present, target detection methods are mainly divided into traditional methods and deep learning methods. Traditional methods
are greatly affected by the external environment and imaging scenes. Although corresponding improvements are proposed for different applications, the recognition performance cannot meet the actual use requirements under conditions such as changes in illumination and complex backgrounds [3-7]. In recent years, deep learning methods have begun to be applied to the recognition of various agricultural products due to their high recognition accuracy and better adaptability to complex environments, such as the research on green citrus visual detection technology in natural environments[8], based on deep full convolutional neural The field rice ear segmentation of the network[9], the convolutional neural network combining the hole convolution and the global pooling to identify crop seedlings and weeds[10], etc.

The YOLO target recognition algorithm transforms the target recognition problem into a regression problem, and only uses a deep convolutional network to quickly and accurately identify the target [11]. YOLOv3 has made some adaptive improvements on the basis of YOLOv2, including multi-scale recognition, multi-label classification, etc., and uses the DarkNet-53 network improved based on the Res Net network as the feature extractor, which greatly improves the ability to recognize small targets [12]. Combining the advantages of the YOLO series network, this article proposes a cotton top bud identification method based on YOLOv3, which can realize the rapid and accurate identification of cotton top buds.

2. YOLOv3 Network Model
The YOLO network model is a one-stage method [13]. YOLO divides the input image into S×S grids. If the center of an object falls within a certain grid, the corresponding grid is responsible for detecting the object. YOLOv1 uses a non-maximum suppression algorithm for prediction, which is not effective in detecting small targets and the positioning is not accurate enough. The YOLOv2 proposed by Redmon et al. [14] absorbed the idea of RPN in Faster R-CNN. The output layer uses a convolutional layer instead of the fully connected layer of YOLOv1. The convolutional layer uses Batch Normalization, and the k-means algorithm is proposed, but YOLOv2 is the small target detection effect is average. YOLOv3 has made some improvements on the basis of YOLOv2, deepened the network, proposed multi-label classification prediction, used logistic regression to regress the box confidence, and proposed cross-scale prediction, using similar FPN (feature pyramid networks). The fusion method is to perform position and category prediction on feature maps of multiple scales, and the detection effect of small targets is significantly improved. The Logistic classifier is used to replace the Softmax classifier. YOLOv3 uses the deep residual network to extract image features and realizes multi-scale prediction, obtaining the best balance between recognition accuracy and speed [15]. YOLOv3 target detection loss function formula (1) is shown, and the structure diagram is shown in Figure 1.

\[
\text{loss}(\text{object}) = \sum_{i=0}^{K} \sum_{j=0}^{M} I_{ij} (2 - w_i * h_i) (-(x_i * \log(h_i) - (1 - x_i) * \log(1 - h_i))) + \\
\sum_{i=0}^{K} \sum_{j=0}^{M} I_{ij} (2 - w_i * h_i) (-(y_i * \log(h_i) - (1 - y_i) * \log(1 - h_i))) + \\
\sum_{i=0}^{K} \sum_{j=0}^{M} I_{ij} [(w_i - h_i)^2 + (h_i - h_i)^2] - \\
\sum_{i=0}^{K} \sum_{j=0}^{M} I_{ij} [C_i \log(C_i) + (1 - C_i) \log(1 - C_i)] - \\
\sum_{i=0}^{K} \sum_{j=0}^{M} I_{ij} [I_{i}] - \\
\sum_{i=0}^{K} \sum_{j=0}^{M} I_{ij} [C_i] - \\
\sum_{i=0}^{K} \sum_{j=0}^{M} I_{ij} [1 - C_i]
\]

(1)

In YOLOv3, Loss is divided into three parts: The first is the error caused by the target frame position \(x, y, w, h\) (upper left corner and length and width), that is, the box brings Loss. And the Loss brought by box is divided into BCE Loss brought by \(x, y\) and MSE Loss brought by \(w, h\); the second is the error brought by target confidence, which is Loss brought by obj (BCE Loss); The third is the error
caused by the category, which is the Loss caused by the class (the number of BCE Loss in the category).

3. Data Set Production

3.1 Test Data Collection

The collection location of the cotton top bud image is located in the cotton test field in Binzhou City, Shandong Province. The collection time is from July to August. The visible light binocular camera is used to collect the visible light image at a distance of about 400mm directly above the cotton top bud, and the frame rate is 30 frames/s. The image size is 1280×720 pixels. The images include various natural environments and lighting conditions such as sunny, cloudy, strong light, low light, backlight, face the light, etc., to meet the diverse needs of sample data.

Perform data amplification on the original data set, and use methods such as flipping, zooming, rotating, cropping, translation, adding noise and color dithering (including adjusting image brightness, saturation, and contrast) to augment the collected images, and finally 23168 images were obtained, the original image and the enhanced image are shown in Figure 2 and Figure 3 respectively. After removing the fuzzy images caused by camera movement and invalid images without cotton top buds, a data set containing 19,914 images is obtained.
3.2 Data Set Preparation

Use the Labeling tool to label the cotton top buds with the above data set. Considering the correspondence between labels and data and ensuring the uniform distribution of the data set, the data set is randomly divided into training set and test set according to the ratio of 25% and 75%. The training set contains 4865 labeled samples with borders, and the test set contains borders. The number of labeled samples is 15049. Store the final data set in the format of the PASCAL VOC data set, and then subdivide the test set into 2 parts according to the different growth stages of the top buds of cotton, that is, the morphological characteristics of the top buds are more obvious (collected in July), which is recorded as test set A; The morphological characteristics of the terminal buds were not obvious (collected in August) and recorded as test set B. Among them, the test set A contains 12603 image samples, and the test set B contains 2446 image samples. The final data set is shown in Table 1.

| Data set | Training set | Test set | Total quantity |
|----------|--------------|----------|---------------|
|          | A            | B        |               |
| Number of images | 4862 | 12603 | 2446 | 19914 |

4. Model Training

4.1 Test Platform

The operating system used for training is win10; the test framework is tensorflow. Darknet-53; the processor is i7-8700T@2.40GHz+2×8GB DDR4+500G SDD, six cores and twelve threads; the graphics card is RTX 2070S, with 8GB of video memory; Use the parallel computing framework of CUDA 9.0 version with the deep neural network dependency library of CUDNN 7.0 version;

The running scene is Visual Studio 2015 Community and OpenCV3.40. The physical map is shown in Figure 4.
4.2 Network Training

This article uses the YOLOv3 network model for experimentation, a total of 244 hours of training time, 31376 batches of iterations, the average training time of each iteration is 22 seconds, and the total number of training sheets is 2008964. Pre-training is not used to train the network model, combined with the back-propagation algorithm to modify the model parameters, so that the loss function is continuously reduced. When the average loss is less than 0.01, and the multi-iteration loss value is less than 0.065, the training is stopped.

The model parameters are set to 64 samples per batch, the initial learning rate is 0.001, and the momentum factor is 0.9. After 5000 iterations of training, the learning rate is reduced by 10 times. The model saves weights every 100 times.

After the training is completed, the iterative loss value is read from the log file and the curve is drawn. The result is shown in Figure 5. It can be seen from Figure 3 that the loss value of the first 21000 iterations decreases rapidly, and then the loss value stabilizes in a small range, below 0.065, the model accuracy and generalization performance have reached the requirements.

Use F1 score (F1 score) and AP value (average precision) to evaluate the model trained by the loss function. The F1 value calculation formula and AP value calculation formula are as follows:

\[
P = \frac{TP}{(TP + FP)}
\]

\[
R = \frac{TP}{(TP + FN)}
\]
Where \( P \) is the accuracy rate, \( R \) is the recall rate, \( TP \) is the number of real positive samples, \( FP \) is the number of false positive samples, and \( FN \) is the number of false negative samples.

5. Analysis of Test Results

This article uses the YOLOv3 network as the basic network and does not use the pre-training model. There are 15049 test samples, including backlight, face the light, strong light, low light, different growth stages of top buds, and different types of top buds. The test results show that the average recognition time of each image is 98ms; for test set A, the recognition rate is 96.81%; for test set B, the recognition rate is 92.89%; the comprehensive recognition rate is 96.18%. The test result image shows that the system can correctly identify the cotton top buds in the image and meet the cotton topping requirements. The statistics of the recognition results are shown in Table 2.

| Time            | Test set A | Test set B | Total  |
|-----------------|------------|------------|--------|
| Test sample     | 12603      | 2446       | 15049  |
| Correct identification | 12202   | 2272       | 14474  |
| Recognition rate| 96.81%     | 92.89%     | 96.18% |

Part of the recognition results are shown in the figure below. The red box in the figure is the cotton top bud mark identified by the system. The cotton top bud shape in July is more prominent, which is conducive to identification. The recognition rate is higher than that in August. Affected by shadows, the recognition rate decreases compared with that under face the light conditions.

\[
F1 = \frac{2PR}{P+R} \\
AP = \int_0^1 P(R) dR
\]
Train the YOLOv3 network, and compare and experiment under different test sets. As shown in Table 3, the F1 value of the YOLOv3 model is above 97% and the AP value is above 93% when the cotton top bud identification features are more obvious, and when the cotton top bud identification features are not obvious, the F1 value it can also reach more than 91%. In a natural environment, its F1 value can reach 96%, and its AP value can reach 93%. The analysis of the results of the above-mentioned comparative experiments shows that the YOLOv3 neural network can effectively identify the top buds of cotton in the natural environment, and the recognition accuracy and recognition speed have significant advantages.

6. Conclusion
This article presents a method for identifying cotton top buds based on YOLOv3 network. The test results show that in a real natural environment, the system recognition time is less than 100ms, and the recognition success rate can reach over 96%. It has the ability to quickly and accurately identify cotton top buds, which can meet the cotton topping needs and provide correct visual information for subsequent cotton top bud position measurement and mechanical control.

References
[1] Zhang M, Chen CY . Security analysis of cotton industry chain in China. [J]. Journal of Tianjin University of Technology, 2011, 30(2): 77-81.
[2] Han YJ, Shi L, Cao LL, et al. Research status and Prospect of cotton terminal bud identification and location technology. China Agricultural Chemical Journal, 2018, 11: 39-11.
[3] Tian N, Yang XW, Shan DL, et al. Current situation and Prospect of digital agriculture in China[J]. China Agricultural Chemical Journal, 2019, 40(4): 210-213.
[4] Wu X, Zhang Y, Li LH, et al. Path extraction method of vision guided AGV under complex illumination [J/OL]. Journal of Agricultural Machinery, 2017, 48(10): 15-24.
[5] LI C, SONG D, TONG R, et al. Illumination-aware faster. R-CNN for robust multispectral pedestrian detection. [J]. Pattern Recognition, 2019, 85: 161-171.
[6] WU X, SUN C, ZOU T, et al. Intelligent path recognition against image noises for vision guidance of automated guided vehicles in a complex workspace[J]. Applied Sciences, 2019, 9(19): 4108.
[7] Shi ZL, Yang Y, Ding YQ, et al. Research on the development status of cotton mechanized topping technology in Xinjiang. [J]. Anhui Agricultural Science, 2014, 42(35): 12746-12748, 12745.
[8] Xiong JT, Liu Z, Tang LY, Lin R, et al. Research on visual inspection technology of green citrus in natural environment[J]. Journal of Agricultural Machinery, 2018, 49(4): 45-52.
[9] Duan LF, Xiong X, Liu Q, et al. Segmentation of rice spikes in field based on deep total convolution neural network. [J]. Journal of Agricultural Engineering, 2018, 34(12): 202-209.
[10] Sun J, He XF, Tan WJ, et al. Identification of crop seedlings and weeds using convolution neural network based on hole convolution and global pooling. [J]. Journal of Agricultural Engineering, 2018, 34(11): 159-165.
[11] Lan WB, Dang JW, Wang YP, et al. Pedestrian detection based on YOLO network model. Proceedings of the 2018 IEEE International Conference on Mechatronics and Automation. Changchun, China. 2018. 1547-1551.
[12] Chen ZB, Ye DY, Zhu CX, et al. Target recognition method based on improved YOLOv3. Application of Computer System, 2020, 29(1): 49-58.
[13] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 779-788.
[14] Redmon J, Farhadi A. YOLO9000: Better, faster, stronger[C] //IEEE Conference on Computer Visionand Pattern Recognition. Honolulu: IEEE. 2017: 7263-7271.
[15] Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv: 1804.02767, 2018.