American Journal of Biochemistry and Biotechnology

Original Research Paper

Research on Cotton Top Bud Target Detection Algorithm Based on Improved RetinaNet

[1]Jikui Zhu, [1]Shijie Lin, [1,2,3]Fengkui Zhang, [1,2,3]Ting Zhang, [1,2,3]Shijie Zhao and [1]Ping Li

[1]College of Mechanical and Electrical Engineering, Tarim University, Alar, Xinjiang, China
[2]Department of Xinjiang Uygur Autonomous Region, Modern Agricultural Engineering Key Laboratory at Universities of Education, Tarim University, Alar, Xinjiang, China
[3]Key Laboratory of Tarim Oasis Agriculture (Tarim University), Ministry of Education, Tarim University, Alar, Xinjiang, China

Article history
Received: 27-11-2023
Revised: 11-03-2024
Accepted: 16-03-2024

Corresponding Author:
Ping Li
College of Mechanical and Electrical Engineering, Tarim University, Alar, Xinjiang, China
Email: lpdyy716@163.com

Abstract: To solve the problems of low accuracy and high miss rate in the recognition of cotton apical buds during mechanical topping, an enhanced method based on the RetinaNet network is proposed for the accurate identification of cotton apical buds under natural light. The traditional RetinaNet algorithm is validated to improve the recall rate and average accuracy of cotton apical bud recognition (mAP@0.5) at 83.61% and 77.64% respectively. Due to the shallow nature of the network, there is still overfitting and the RetinaNet algorithm is improved. This algorithm incorporates R-CBAM and ShuffleViT Block network modules and uses Atrous Spatial Pyramid Pooling (ASPP) to connect the cross-domain feature layer to the feature fusion layer. The results indicate thatcompared with the traditional RetinaNet algorithm, theimprovedRetinaNet algorithm has an average accuracy (mAP@0.5) of 96.25% and a recall rate of 91.10% for cotton apical bud recognition. This indicates that the improved RetinaNet algorithm has optimal recognition performance and high recognition accuracy for cotton apical buds, laying a solid foundation for precise topping operations in cotton cultivation.

Keywords: Cotton Apical Bud, RetinaNet, Target Detection

Introduction

Cotton is an important economic crop and textile raw material in China. To increase the cotton yield, it is necessary to control the height and branch growth of cotton. Topping is a commonly used method that can effectively control the growth height of cotton and increase yield (He *et al*., 2021). To solve the problem of how to quickly and efficiently identify cotton top buds in field topping machinery, the recognition of cotton apical buds based on a neural network algorithm is deeply studied.

The cotton topping period is from late July to early August and the operation time is short. Simple manual work is labor-intensive, low-efficiency, and high-cost (Han *et al*., 2022). Chemical topping has a strong inhibitory effect and can easily lead to cotton peach deformities and mechanical topping is an important direction for future development (Biradar *et al*., 2011). However, there are some problems with mechanical topping, such as high miss rate and poor accuracy in cotton apical bud recognition (Tang *et al*., 2008). To solve this problem, optimizingthe target detection algorithm of cotton apical bud has become an important way to improve the identification accuracy, making it better serve the automation and intelligence of cotton topping machinery and equipment.

Target detection methods can be divided into two categories: Two-stage algorithms and single-stage algorithms. The two-stage algorithms include R-CNN, SPP-Net (He *et al*., 2015), Faster R-CNN (Ren *et al*., 2015) and the single-stage algorithms include YOLO (Redmon and Farhadi, 2018; Cardellicchio *et al*., 2023; Wang and Liu, 2022), SSD (Feng *et al*., 2019; Yao *et al*., 2022) and RetinaNet (Wu *et al*., 2023). These algorithms are widely used in the identification of agricultural products, fruits and vegetables, plants, and cotyledons. Based on the YOLOV5s algorithm, Wang and He (2021) detected the apples in the fruit thinning period of channel pruning. Tian *et al.* (2019) used the YOLOV3 model to identify and detect apples from different periods. To distinguish the maturity status of tomatoes, Egi *et al.* (2022) used YOLOV5 and Deep SORT algorithms to detect their maturities. To identify plants and cotyledons, Pan *et al.* (2022) used a two-stage algorithm, faster R-CNN,

to automatically identify and detect the sugarcane seedlings to estimate the yield of this round of sugarcane. Islam *et al*. (2021) used the KNN algorithm to detect the weeds in pepper seedlings. The features of apple, tomato, sugarcane, and weeds are relatively distinct in the captured image environment. The image environment is controllable and the recognition accuracy and speed can meet practical application requirements.

For the identification and detection of cotton plants and cotyledons, recognition and detection methods based on deep learning have been widely applied to the identification of cotton plants and the detection of the tops of cotton plants. Liu (2022) used SSD and YOLO algorithms to detect targets on the top of cotton in natural environments, achieving excellent performance. They also introduced the decoupled head module and Anchor Free method to improve the Head of YOLOV4, achieving an accuracy of 90.15%. Xiaochen and Shen (2018) input the preprocessed images into the improved VGG network to improve the accuracy of cotton top target detection in complex field environments. Its recognition accuracy is 83.4%, which can effectively detect and locate the top of cotton. Siqi (2021) replaced the original Residual Unit with a dense connection block and replaced the original dense connection block with separable convolution, integrating multi-scale receptive fields and improving the YOLOv3 network, with a recognition accuracy of 90.93%. Although the accuracy of cotton top bud recognition based on YOLO algorithms has met the requirements of practical applications, they have the common problems of slow recognition speed and long inference time.

Due to the complex natural environment, high planting density, and relatively small apical bud area, as well as the presence of cotyledon occlusion of cotton plants, this study aims to reduce computational complexity and obtain more accurate prediction boxes. Based on the traditional RetinaNet deep learning algorithm, optimizations and improvements are made to accurately determine the position of the cotton apical buds.

Materials and Methods

Image Processing of Cotton Plants

Preparation of Test Materials

The hardware system mainly contains the computer and camera (IPHONE12). Ubuntu 20.04.1 is adopted as the operating system, PyTorch 1.7 and CUDA 11.0 as the deep learning environment, and python3.8 as the development language, in Table 1.

To obtain the optimal optical effect of cotton plants, the shooting time was arranged at 12 noon and the shooting distance of the camera in the recognition system of the cotton plant top was simulated. When shooting, cotton plants with different growth levels were selected. The distance between the camera and the top of the cotton was controlled as 10-20cm and the shooting was at a vertical angle of 90°C directly above the cotton apical buds. A total of 800 top photos of cotton plants with a resolution of 2532×1170 were collected in the experiment.

To obtain continuous image data of cotton apical buds, the images were collected four times between July 5 and July 8, each with 200 images of different occlusion conditions and lighting angles. The dataset sample is shown in Fig. 1.

Image Preprocessing of Cotton Plants

Data Set Expansion

The training of deep learning networks typically requires a large amount of data and a small amount of data can easily lead to overfitting in network training (Hu *et al*., 2018; Woo *et al*., 2018; Dosovitskiy *et al*., 2020). Here we use data augmentation to expand the dataset samples of cotton apical buds to enhance the generalization ability and robustness of the trained model ((Pinto *et al*., 2019; Rohaziat *et al*., 2020). The main methods such as horizontal flipping, mirror transformation, image brightness adjustment, and image black-and-white processing are used to expand the data set of cotton top buds. The rendering is shown in Fig. 2 and a total of 4000 cotton top bud images were obtained. 90% (3600) images were randomly selected from the total sample as the training set and the remaining 400 images were used as the set to verify the performance of the experimental evaluation model.

Target Detection Algorithm for Cotton Apical Buds Based on Retinanet

Settings of the Retinanet Network Structure

RetinaNet is specially designed to solve class imbalance and multi-scale problems in target detection, with optimal detection performance and efficient computing speed (Wu *et al*., 2023; Ju *et al*., 2019). RetinaNet introduces a new Feature Pyramid Network (FPN) and focal loss function (Li *et al*., 2023). The RetinaNet network architecture is shown in Fig. 3, which mainly consists of a backbone network and a feature pyramid network. First, the backbone feature extraction network of RetinaNet is ResNet18, which is stacked by a residual network structure, in Fig. 4. It is used for preliminary feature extraction of the target.

Table 1: Experimental environment

Software and hardware configuration Parameters	
Operating system	Ubuntu 20.04.1
CPU	Intel® Xeon® Gold 5218 CPU@2.3Hz
GPU	GTX3090 (24GB)
The programming language	Python 3.8
Deep learning framework	PyTorch 1.7, CUDA 11.0

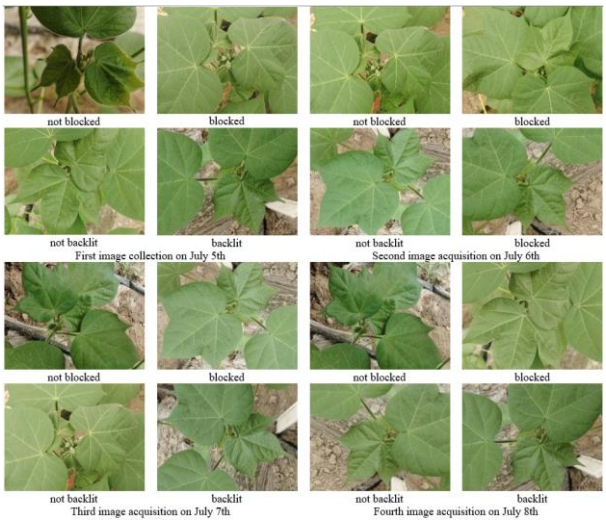

Fig. 1: Sample data graph

Fig. 2: Example of data set expansion part

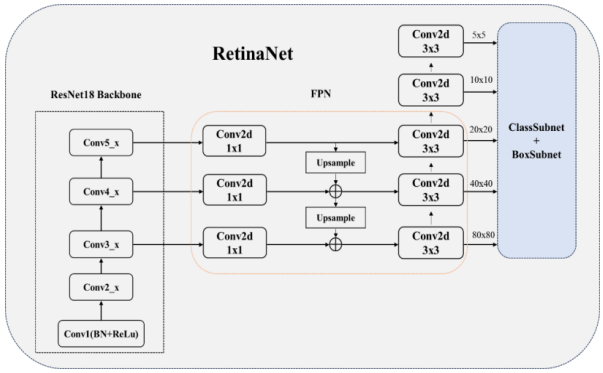

Fig. 3: RetinaNet algorithm network architecture

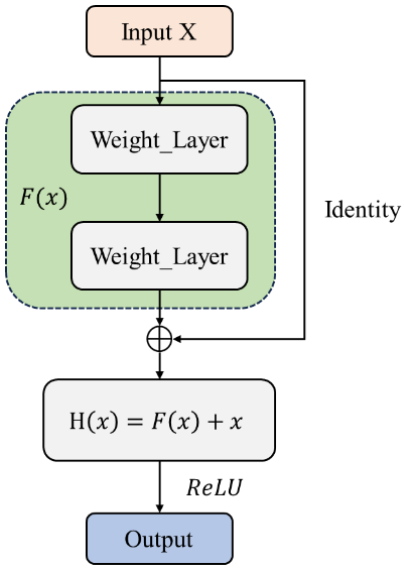

Fig. 4: Residual structure

By introducing an adjustable balance parameter, setting the loss function of RetinaNet, and reducing the weight of the loss function, the model pays more attention to the cotton plant samples that are difficult to classify andimprovesthe performance ofdifficult samples. The total loss of RetinaNet detection model training can be expressed as follows:

$$L = \frac{1}{N_{pos}}\sum_i l_{cls}^i + \frac{1}{N_{pos}}\sum_j l_{reg}^j \quad (1)$$

where, l_{cls} is the classification loss Focal Loss; l_{reg} is the border regression loss smooth L_1 Loss; N_{pos} is the number of positive and negative samples; i is the positive sample and j is the negative sample.

Focal loss is defined as follows:

$$FL(p_t) = -\alpha_t * (1-p_t)^{\gamma} * log(p_t) \quad (2)$$

where, p_t is the target category probability predicted by the model; α_t is an adjustable balance parameter used to adjust the loss weight of different categories; γ is an adjustment factor that is used to control the attention of difficult samples.

RetinaNet Algorithm Validation and Comparison of Typical Algorithms

RetinaNet Algorithm Validation

In the experiment, the RetinaNet algorithm is used for target detection of cotton apical buds. The experimental parameter settings are shown in Table 2.

Table 2: Verification experiment and parameter settings of the RetinaNet algorithm

Parameters	Value settings
Learning rate	0.001
Batch size number	64.000
Iterations	100.000
Input feature size	640*640

Table 3: Experimental result data of the RetinaNet algorithm

Model	Accuracy/ %	Recall/ %	Harmonic mean/%	Average precision (mAP@0.5)/%
RetinaNet	80.11	83.61	76.11	77.64

Table 4: Experimental results of the YOLOv5s algorithm

Model	Accuracy/%	Recall/%	Average precision (mAP@0.5)/%
YOLOv5s	67.20	82.01	73.68

Table 5: The Experimental result of the SSD algorithm

Model	Accuracy/%	Recall/%	Average precision (mAP@0.5)/%
SSD	70.11	81.90	75.57

Fig. 5: Partial detection effect of the RetinaNet algorithm on the dataset of cotton apical buds

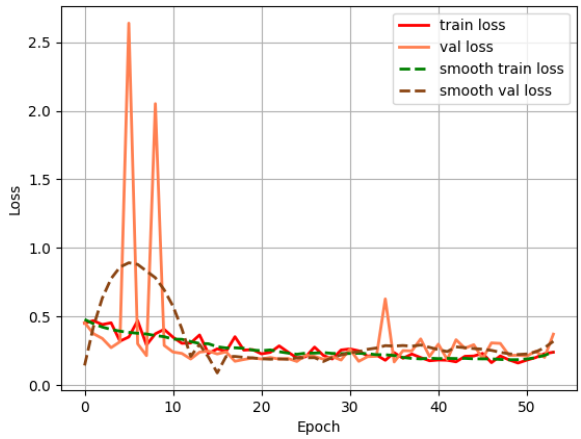

Fig. 6: RetinaNet Algorithm loss function curve

The detection results are shown in Fig. 5, the loss function curve is shown in Fig. 6 and the test results are shown in Table 3.

According to the experimental result data shown in Table 3, the RetinaNet algorithm has a certain performance in the detection of cotton apical buds. The accuracy rate is 80.11%, indicating the ratio between the number of apical buds correctly predicted by the algorithm and the total predicted number. The recall rate is 83.61%, indicating the ratio between the number of apical buds successfully detected by the algorithm and the actual number of apical buds. The harmonic average is 76.11%, indicating the balance between accuracy and recall. The average accuracy is 77.64%, which is the detection results calculated under different confidence thresholds. In general, the experimental results show that RetinaNet has achieved relatively good performance in the recognition and detection of cotton apical buds. Figure 6, the training loss and testing loss show significant differences within the first 10 iterations. However, after 10 iterations, the training loss is generally stable, while the validation loss significantly decreased. This indicates there is an overfitting in the RetinaNet model in the preliminary experiment. In the case of sufficient sample dimensions, this indicates that excessive noise in the dataset causes the model to over-identify the noise features, thus ignoring the true relationship between input and output. In other words, the model has weak feature selection ability in the case of a large number of samples.

Comparison of Typical Algorithms

Based on the dataset in Chapter 1, YOLOv5s and SSD algorithms are used to conduct experiments on the target detection of cotton apical buds. The hardware configuration and experimental environment of these algorithms are consistent with those used in the RetinaNet experiment. After 100 training epochs, the performance metrics of the YOLOv5s algorithm and the detection results are shown in Table 4.

The initial learning rate of the SSD algorithm in the detection experiment is set to 0.01, the BatchSizeto64, the momentum size to 0.937, the weight delay size is 0.0005 and the epoch size to 100. The detection results are shown in Table 5.

Comparing the RetinaNet, YOLOv5s, and SSD algorithms after 100 training epochs, it is evident that YOLOv5s can achieve an accuracy of 67.20%, a recall rate of 82.01% and an average accuracy (mAP@0.5) of 73.68%. On the other hand, the corresponding SSD algorithm has an accuracy of 70.11%, a recall rate of 81.90%, and an average precision (mAP@0.5) of 75.57%. Comparing the pre-training, it is observed that the RetinaNet algorithm demonstrates superior performance in the recognition of cotton apical buds, with an accuracy of 80.11% and a recall rate of 83.61%. These results indicate that the RetinaNet has the best performance in target detection, with superior recognition capability for cotton apical buds.

Improvement of the Target Detection Algorithm for Cotton Apical Buds Based on RetinaNet

Multi-Head Attention Mechanism Backbone Network Based on Channel Rearrangement

To solve the problem of the weak feature selection ability of the RetinaNet algorithm, a new global

information extraction module, ShuffleViT Block is designed. The ShuffleViT Block utilizes the "shuffling" mechanism in ShuffleNet and the slicing mechanism in MobileViT. First, the block features in the ViT are sliced and split to scramble the information within the block. In this way, the interaction of local feature information in the block is achieved while keeping the block sequence information unchanged, which can reduce the calculation and overcome the problems of high training cost and unsatisfactory inference speed in MobileViT. At the same time, through the slicing mechanism, the network can retain the positional information of the slices, eliminating the noise and retaining the key local features.

Analysis of the ShuffleNet Network Module Model

ShuffleNet is a lightweight convolutional neural network structure that minimizes computational and storage complexity (Yu *et al*., 2023; Bi *et al*., 2019). It uses the hierarchical convolution to divide the input and convolution kernel into several groups. Each group operates the convolution independently and eventually connects the outputs of all groups. The convolution structure is shown in Fig. 7. The input features and convolution kernels are divided into *G* groups respectively. The number of channels in each group is C/G and that of convolution kernels in each group is K/G. For the input feature *Xg*, each set of convolutions can be expressed as follows:

$$y_g = W_g \otimes x_g + b_g \tag{3}$$

where, X_g represents the *G* group of input features; W_g represents the *G* group of convolution kernel; b_g represents the *G* group of offset term and Y_g represents the output result of the *G* group. The final output result is the connected set of all group results, namely:

$$y = \left[y_1, y_2, \ldots, y_G\right] \tag{4}$$

Assuming that its feature input and output are $W \times H \times C_1$ and $W' \times H' \times C_2$, respectively and the size of a single convolutional kernel is $k \times k$ and the input features are divided into g groups. Then the feature input data of each group is $W \times H \times \frac{C_1}{g}$; the single convolution kernel in each group $k \times k \times \frac{C_2}{g}$ and each set of the output feature data is $W' \times H' \times g$. Then the parameters of the grouped convolution are calculated as follows:

$$Params = k^2 \times \frac{C_1}{g} \times \frac{C_2}{g} \times g = k^2 \frac{C_1}{g} C_2 \tag{5}$$

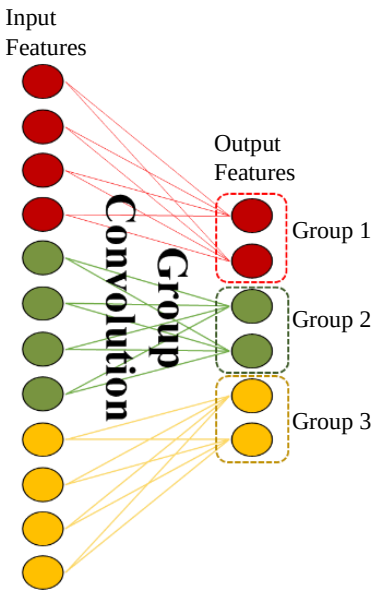

Fig. 7: Diagram of the group convolutional structure

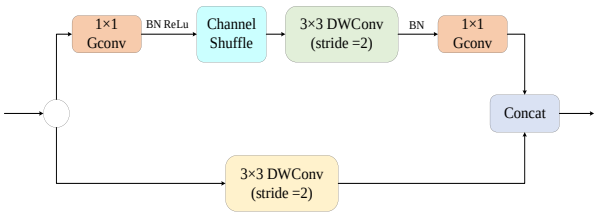

Fig. 8: Structure of the ShuffleBlock module network

$$FLOPs = k^2 \times \frac{C_1}{g} C_2 \times W' \times H' \tag{6}$$

The structure of the ShuffleNet network module is shown in Fig. 8. Based on the simple residual structure, an average pooling layer with a pooling core of 3 is added; 1*1 group convolutions are used and the feature information of packet convolution is interacted by channel rearrangement. Then a 3*3 depth separable convolution and a 1*1 grouped convolution are used to extract features and restore the channel dimension of the aforementioned "shuffled" (channel rearrangement) data. Through the Concat function, the dimension of the input feature channel is expandedwithout additional calculation.

Analysis of Vit Network Module Model

ViT is an image classification network model based on the Transformer model (Yu *et al*., 2023). The features of the image are extracted and classified by segmenting the image into small pieces and then feeding them into the Transformer model as a sequence. The ViT network structure model is shown in Fig. 9:

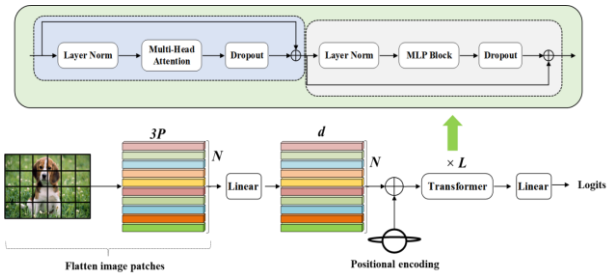

Fig. 9: ViT network structure

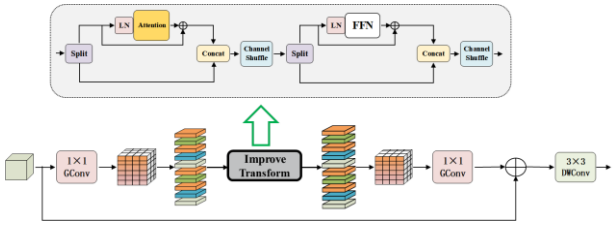

Fig. 10: Diagram of the ShuffleViT block module structure

The overall structure of ViT can be divided into five components: Input layer, embedding layer, learnable embedding, position offset, and Transformer encoder layer. The image is decomposed into small pieces and serialized into vectors by VIT and the self-attention mechanism of the Transformer model is used to capture the global context information of the image.

Improved ViT Module (ShuffleViT)

To improve the accuracy and adaptability of target detection for cotton apical buds, a lightweight and high-performance backbone network is designed based on the structural principles of the visual Transformer and ShuffleNet network module models. The diagram of the ShuffleViT Block module structure is shown in Fig. 10.

Figure 10, the ShuffleViT Block model is different from ViT. Because it uses point-wise convolution and depth-separable convolution as feature extractors, which can better capture important information and multi-scale information in the image and improve the representation ability of features. In terms of data processing, the input feature map $X \in \mathbb{R}^{B\times C\times W\times H}$ (*B*, *C*, *W*, and *H* respectively represent the batch, channel number, width, and height of the convolution feature) is directly divided into multiple sub-tensors along the channel dimension. That is, a 1*1 group convolution is projected to a fixed *D* dimension $X \in \mathbb{R}^{B\times D\times W\times H}$ and then the features are divided into blocks of $P^h \times P^w$ ($P^h, P^w \le H, W$; h and w respectively represent the height and width of the block features). The processed features are $X_p \in \mathbb{R}^{B\times P^h\times P^w\times \frac{HW}{P^hP^w}}$, through which the calculation and parameters of the model are reduced, maintaining the effectiveness and expressive ability of the model.

Next, feature a is segmented into two parts based on the second dimension: $X_{ps1}, X_{ps2} \in \mathbb{R}^{B\times \frac{P^hP^w}{2}\times \frac{HW}{P^hP^w}\times D}$ Image slicing is performed to reduce the computational complexity of the network and improve the processing capability of large-size features as follows:

$$X_{ps1}, X_{ps2} = Split(X) \tag{7}$$

where, X_{ps2} serves as the input of the self-attention module; X_{ps1} serves as the residual connection and X_{ps2} is concatenated according to the second dimension. Then, the "shuffling" mechanism is used to "shuffle" the concatenated features in the second dimension and the formula can be obtained as follows:

$$F_{attn}\left(X_{ps1}, X_{ps2}\right) = Cat\left(Attn\left(LN\left(X_{ps2}\right)\right) + X_{ps2}, X_{ps1}\right) \tag{8}$$

The "*shuffle*" process is to scramble pixels inside the slice to achieve interaction within the block without changing the original block position. In the subsequent nonlinear transformation stage of the feedforward neural network, a similar processing method as described above is also used as follows:

$$X_{shuffle} = Shuffle\left(F_{attn}\right) \tag{9}$$

$$X'_{ps1}, X'_{ps2} = Split\left(X_{shuffle}\right) \tag{10}$$

$$F_{ffn}\left(X'_{ps1}, X'_{ps2}\right) = Cat\left(FFN\left(LN\left(X'_{ps2}\right)\right) + X'_{ps2}, X'_{ps1}\right) \tag{11}$$

$$X_{ShuffleViT} = Shuffle\left(F_{ffn}\right) \tag{12}$$

where, F_{attn} represents the output feature of the multi-head attention network part; $X_{shuffle}$ represents the feature data after the secondary channel rearrangement; X'_{ps1} and X'_{ps2} represents the block feature vectors after the secondary slicing process; *LN* represents the normalization operation; *FFN* represents the feedforward Neural network; F_{ffn} represents the output of the feedforward network after channel rearrangement and $X_{ShuffleViT}$ represents the output characteristics of the entire improved module.

According to the improved method, the calculation of the ShuffleViT Block is significantly reduced. Through sharing and channel rearranging, the ability to extract global features is improved without losing the position information of slices, thereby addressing the issue of algorithm overfitting (Appendix 1 for detailed code).

Fig. 11: Structure diagram of multi-head attention mechanism backbone network based on channel rearrangement

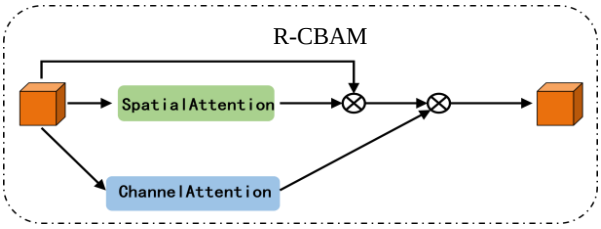

Fig. 12: R-CBAM module structure

Overall Structure of the Backbone Network with Multi-Head Attention Mechanism Based on Channel Rearrangement

The Shuffle Block module and ShuffleViT Block module are combined to form the overall network structure, in Fig. 11. A convolution module with a 3×3 size convolution kernel, a pooling module with a 3×3 size pooling kernel, and three down sampling Shuffle Block modules are selected to down sample the input features. The non-down sampling Shuffle Block modules and ShuffleViT modules in different structural regions are used to extract features of the input image data.

Introduction of R-CBAM Hybrid Attention Mechanism

When detecting cotton apical buds in the field, it is difficult to accurately detect small targets due to the small size and complex background of the target in real-scene image data. To this end, attention mechanisms need to be added. CBAM (Convolutional Block Attention Module) is an attention mechanism module that is used to enhance the modeling ability of convolutional neural networks for spatial and channel attention (Hu and Mingyu, 2022; Ran *et al*., 2023).

CBAM has the disadvantages of increasing the complexity and calculation of the network, as well as increasing the time and resource consumption of training and inference (Wang *et al*., 2023). CBAM needs to be improved. The improved structure is shown in Fig. 12. The fully connected layer in the channel attention is replaced with a 1×1 convolution and the attention in the channel domain and spatial domain is adjusted to a parallel state, with the original input feature $X \in \mathbb{R}^{B\times C\times W\times H}$ serving as the input of the two-dimensional attention module. Then, the feature information extracted from the spatial domain is multiplied with the original input feature $X \in \mathbb{R}^{B\times C\times W\times H}$ to obtain the mixed information of the spatial domain. The obtained mixed information of the spatial domain is multiplied with the output feature of the channel domain to obtain the complete spatial channel mixed attention feature information. Finally, the feature information of the two dimensions is achieved without mutual interference, thereby effectively avoiding overfitting caused by the weighted overlap. The improved model is called R-CBAM (detailed code can be found in Appendix 2).

The R-CBAM module is define $F_{sa}(X)$d as representing the spatial domain output features; $F_{ca}(X)$ representing the channel domain output features; $F^1_{MLP}(X)$ and $F^2_{MLP}(X)$ representing the output features of two fully connected neural network layers, respectively, and $F_{R\text{-}CABM}(X)$ representing the output features of the entire attention module, then $X \in \mathbb{R}^{B\times C\times W\times H}$. When inputting features, the inference process of R-CBAM can be expressed as follows:

$$F_{sa}(X) = s\left(Conv\left(Cat\left(Mean(X), Max(X)\right)\right)\right) \tag{13}$$

$$F^1_{MLP}(X) = Conv\left(ReLu\left(Conv\left(AvgPool(X)\right)\right)\right) \tag{14}$$

$$F^2_{MLP}(X) = Conv\left(ReLu\left(Conv\left(MaxPool(X)\right)\right)\right) \tag{15}$$

$$F_{ca}(X) = \sigma\left(F^1_{MLP}(X) + F^2_{MLP}(X)\right) \tag{16}$$

$$F_{R\text{-}CABM}(X) = \left(F_{ca}(X) \ddot{A} X\right) \ddot{A} F_{sa}(X) \tag{17}$$

where, σ represents the sigmoid function; Conv represents the convolution module with a core of 7×7; Mean represents the mean function; Relu represents the Rectified Linear Unit activation function; Max represents the maximum value function; Avg pool and Max pool represent the mean pooling and maximum pooling functions, respectively.

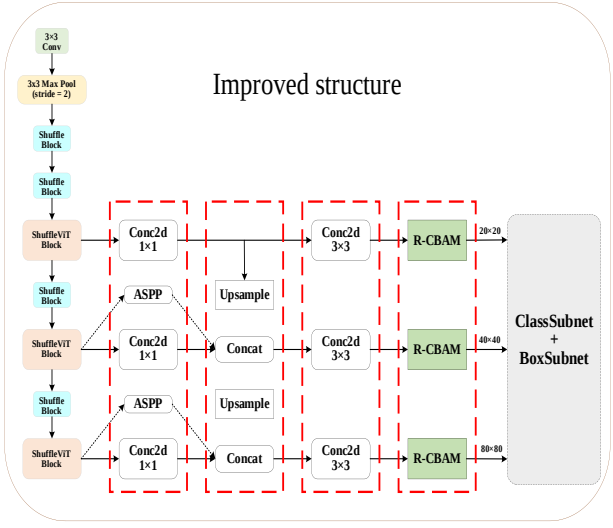

Fig. 13: Improved RetinaNet network structure

Improved RetinaNet Network Architecture

The overall network structure of the target detection algorithm for cotton top buds based on improved RetinaNet is shown in Fig. 13. The network introduces a multi-head attention mechanism to capture feature information of different scales and semantics and rearranges the features of different channels. The multi-head attention mechanism is shown in Fig. 13. Figure 13, there are two intervals in the three 1×1 convolution kernels. The ASPP atrous convolution module is added to the intervals to connect the convolution kernel (Conv2d). Then the receptive field of the convolution kernel is expanded through two up-samples, so as to combine the two up-sampling results with Concat and output it as a 3×3 convolution kernel. Finally, before extracting the feature values, the CBAM module is introduced to combine the channel attention mechanism and the spatial attention mechanism, which can pay attention to the feature information of the channel dimension and the spatial dimension at the same time. By applying a parallel spatial attention mechanism in the connection part of the detection head, the sensitivity of the network to the target location and scale can be improved. In this way, the accuracy and robustness of target detection can be improved, especially for small targets and occluded targets.

Results

Verification of the Target Detection Algorithm for Cotton Apical Buds Based on RetinaNet

To further illustrate the improved target detection algorithm for cotton apical buds based on the RetinaNet algorithm, the effectiveness of the improved algorithm will be verified through multiple sets of experiments.

Experimental Parameters and Environment Settings

In the experiment, 4,000 images were used as a data set, which was divided into a training set and a test set at a ratio of three to one. The process of data reading in the experiment includes the following steps:

(1) Use the voc_annotation.py file to divide the data set into four Txt files, including the original image of the training set, the label image of the training set, the original image of the test set, and the label of the test set image
(2) In the data loader, read these four txt files and obtain the path information of the image
(3) Read the corresponding picture according to the path information and return it as part of the data
(4) At the same time, label information is provided for each picture and the position of cotton apical buds is m. Through this data reading process, the images and labels in the training set and the test set are paired for model training and performance evaluation on the test set, marked as 1, indicating that this is the position of the cotton core

Table 6: Experimental parameters and environmental configuration

Types	Setting content
Data enhanced	Rotation, mirror image, black and white, brightness
Input feature size	640*640
Category	1
Batch size	32
Training times	100 epoch
Learning rate adjustment strategy	Simulated cosine annealing algorithm
Graphics card	RTX3090 (24GB)
System	Ubuntu 20.0

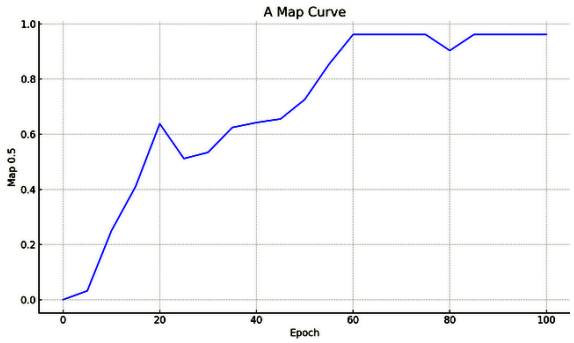

Fig. 14: Improved RetinaNet algorithm confidence curve

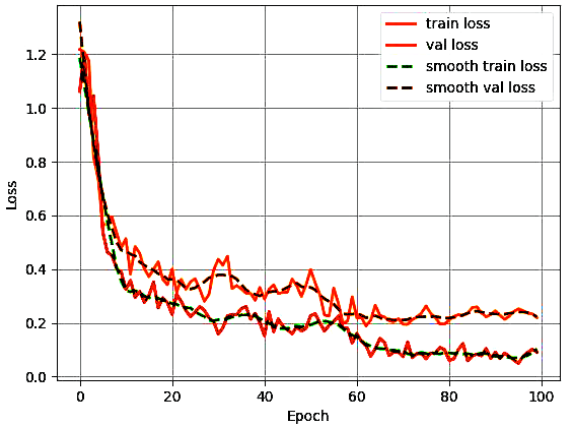

Fig. 15: Loss function of the improved RetinaNet algorithm

The specific experimental parameters and environmental settings are in Table 6.

Verification Results of the Target Detection Algorithm for Cotton Apical Buds based on RetinaNet

The detection and recognition results in Fig. 14. It can be seen from Fig. 14 that the improved algorithm can accurately identify cotton apical buds.

The confidence curve is shown in Fig. 15. It can be seen from Fig. 15 that the confidence performance is optimal, which tends to be stable after training for about 60 times and the confidence of the prediction results is relatively high.

The image of the loss function of the improved RetinaNet algorithm is shown in Fig. 15. It can be seen that both the training set loss and the test set loss tend to

decrease, indicating that the model detection is normal and the model parameters are well modulated.

From the above experimental results, it can be concluded that the model has a high accuracy in determining the cotton apical buds. The improved RetinaNet algorithm has a recall rate of 91.10%, mAP@0.5 and the average accuracy is 96.25%.

Discussion

In order to improve the recognition accuracy of cotton top buds, an improved method based on RetinaNet network for accurate identification of cotton top buds under natural light is proposed. By introducing improved modules such as ShuffleViT Block, ASPP and R-CBAM, compared with the original Retina Net model, the improved recall rate can reach 91.10%, mAP@0.5 can reach 0.96, the recall rate is 83.61%, and the average precision of mAP@0.5 is 77.64%, which are increased by 7.49% and 18.61% respectively. The effectiveness of the improved RetinaNet algorithm has been verified, and its advantages in cotton top bud recognition have been demonstrated.

However, this algorithm lacks a critical discussion of the limitations and potential sources of errors in the results, and it has not been tested in the field, nor has it taken into account the impact of wind on cotton cotyledons that cover the top buds of cotton. In the future, we will focus on analyzing the limitations and potential sources of errors of this algorithm, further optimizing it, and verifying it in cotton fields.

Conclusion

Through introducing improved modules such as ShuffleViT Block, ASPP, and R-CBAM, the improved RetinaNet algorithm has achieved satisfactory results in the recognition of cotton apical buds compared with the original RetinaNet model, the recognition and positioning accuracy of the improved model is significantly improved. The improved recall rate can reach 91.10%, and mAP@0.5 can reach 0.96. Compared with the traditional RetinaNet algorithm, the recall rate is 83.61% and mAP@0.5 (average precision mean) is 77.64%, increased by 7.49% and 18.61% respectively. The effectiveness of the improved RetinaNet algorithm is demonstrated and its advantages in the recognitionof cotton apical buds are demonstrated. However, the algorithm has not experimented in the field, nor has it taken into account the effect of wind on the cotton cotyledons that shade the cotton apical buds. Our future work would focus on the verification of the effectiveness of the algorithm in the field.

Acknowledgment

Thank you to the team members for their contributions in the research, to the journal platform for recognizing this paper, and to the School of Mechanical and Electrical Engineering at Tarim University, the Key Laboratory of Modern Agricultural Engineering at Xinjiang Uygur Autonomous Region College, and the Key Laboratory of Oasis Agriculture in Tarim for their support.

Funding Information

This study was financially Supported by the Bingtuan Science and Technology Program (Grant No. 2021CB018) and the Master Talent Project of the Tarim University Presidents Fund (Grant No. TDZKSS202228).

Author's Contributions

Jikui Zhu and Shijie Lin: Organize the data and written the full text.

Fengkui Zhang: Prepared the experimental materials and collected the experimental data.

Ting Zhang: Participated in the design of all experiments.

Shijie Zhao: Checked the article.

Ping Li: Participated in the structural design of the article.

Ethics

Authors should address any ethical issues that may arise after the publication of this manuscript.

References

Bi, P., Luo, J., & Chen, W. (2019). Research on lightweight convolutional neural network technology. *Computer. Eng. Appl*, *55*, 25-35. http://doi.org/10.3778/j.issn.1002-8331.1903-0340

Biradar, D. P., Basavanneppa, M. A., Yadahalli, G. S., Udikeri, S. S., Alagawadi, A. R., & Patil, V. C. (2011). Agronomic performance and economics of BT-cotton as influenced by intercrops and plant protection schedules. *International Journal of Agricultural and Statistical Sciences*, *7*(1), 117-123. http://doi.org/10.3098/ah.2011.85.3.398

Cardellicchio, A., Solimani, F., Dimauro, G., Petrozza, A., Summerer, S., Cellini, F., & Renò, V. (2023). Detection of tomato plant phenotyping traits using YOLOv5-based single-stage detectors. *Computers and Electronics in Agriculture*, *207*, 107757. https://doi.org/10.1016/j.compag.2023.107757

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16×16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. http://doi.org/10.48550/arXiv.2010.11929

Egi, Y., Hajyzadeh, M., & Eyceyurt, E. (2022). Drone-computer communication-based tomato generative organ counting model using YOLO V5 and deep-sort. *Agriculture*, *12*(9), 1290. https://doi.org/10.48550/arXiv.2010.11929

Feng, Y., Chen, J., Xiong, T., Chen, R., Li, X., Tang, Y., & Lei, X. (2019). Application of the ssd algorithm in a people flow monitoring system. *In 2019 15th International Conference on Computational Intelligence and Security (CIS*), (pp. 341-344). IEEE. http://doi.org/10.1109/CIS.2019.00079

Han, X., Lan Y. B., Wang J., Liu Z.Q., Zhao Z. Y., Xi. C. S., Chen J. Y., & Sha. L. M. (2022). Cotton topping unmanned aerial vehicle having cutter discs and front grain lifting baffle plates. U.S. Patent Application No. 17/435001. US2022135220A1[2023-11-13] https://www.freepatentsonline.com/y2022/0135220.html

He, C., Wu, C., Li, N., & Miao, Z. (2021, July). Research on Auto-follow Row Assist Technology of Cotton Picker with Adaptive Speed. *In 2021 40th Chinese Control Conference* (*CCC*), (pp. 3840-3844). IEEE. http://doi.org/10.23919/CCC52363.2021.9550464

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(9), 1904-1916. http://doi.org/10.1109/TPAMI.2015.2389824

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (pp. 7132-7141). http://doi.org/10.1109/CVPR.2018.00745

Hu, & Mingyu. (2022). Research on apple tree bud image recognition method based on deep learning. *Chinese Academy of Agricultural Sciences*. http://doi.org/10.27630/d.cnki.gznky.2021.000951

Islam, N., Rashid, M. M., Wibowo, S., Xu, C. Y., Morshed, A., Wasimi, S. A., ... & Rahman, S. M. (2021). Early weed detection using image processing and machine learning techniques in an Australian chilli farm. *Agriculture*, *11*(5), 387. http://doi.org/10.3390/agriculture11050387

Ju, M., Luo, H., Wang, Z., Hui, B., & Chang, Z. (2019). The application of improved YOLO V3 in multi-scale target detection. *Applied Sciences*, *9*(18), 3775. http://doi.org/10.3390/app9183775

Li, Z., Zhou, Y., Lyu, S., Huang, Y., Yi, Y., & Zhao, C. (2023). Design of Fruit-Carrying Monitoring System for Monorail Transporter in Mountain Orchard. *Journal of Circuits, Systems and Computers*, *32*(15), 2350264. http://doi.org/10.1142/s021812662350264x

Liu, H. (2022). Research on cotton top bud identification system based on deep learning. *Shandong University of Science and Technology*, 2022. http://doi.org/10.27276/d.cnki.gsdgc.2022.000424

Pan, Y., Zhu, N., Ding, L., Li, X., Goh, H. H., Han, C., & Zhang, M. (2022). Identification and counting of sugarcane seedlings in the field using improved faster R-CNN. *Remote Sensing*, *14*(22), 5846. http://doi.org/10.3390/rs14225846

Pinto, P. F. A., Busson, A. J. G., De Melo, J. P. F., Colcher, S., & Milidiú, R. L. (2019, October). PVBR-Recog: A YOLOv3-based Brazilian Automatic License Plate Recognition Tool. In *Anais Estendidos do XXV Simpósio Brasileiro de Sistemas Multimídia E-Web*, (pp. 121-124). SBC. http://doi.org/10.5753/webmedia_estendido.2019.8149

Ran, D., Xiong, X., & Gao, L. (2023, March). An improved YOLOv5 method for small object detection in high resolution images. In *International Conference on Mechatronics Engineering and Artificial Intelligence* (*MEAI 2022*), (Vol. *12596*, pp. 284-289). SPIE. http://doi.org/10.1117/12.2673151

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv: 1804.02767*. http://doi.org/10.48550/arXiv.1804.02767

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, *28*. http://doi.org/10.1109/TPAMI.2016.2577031

Rohaziat, N., Tomari, M. R. M., Zakaria, W. N. W., & Othman, N. (2020). White blood cells detection using yolov3 with CNN feature extraction models. *International Journal of Advanced Computer Science and Applications*, *11*(10). http://doi.org/10.14569/IJACSA.2020.0111058

Siqi, H. (2021). Research on cotton main stem growth point identification based on machine vision. *Hebei Agricultural University*. In Chinese. http://doi.org/10.27109/d.cnki.ghbnu.2021.000229

Xiaochen, & Shen. (2018). Research on cotton plant height identification technology of cotton topping machine. *Shihezi University, Xinjiang*. In Chinese.

Tang, J., Luo X., Hu B., Zhang Q., & Li X. (2008). The Research on Structure Design and Performance Test of 3MDZK-12 Cotton Uniline Profile Modeling Topping Machine.*Journal of Shihezi University (Natural Science Edition*), 511-514. In Chinese. https://link.cnki.net/doi/10.13880/j.cnki.65-1174/n.2008.04.014

Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., & Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and Electronics in Agriculture*, *157*, 417-426. http://doi.org/10.1016/j.compag.2019.01.012

Wang, D., & He, D. (2021). Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. *Biosystems Engineering*, *210*, 271-281. http://doi.org/10.1016/j.biosystemseng.2021.08.015

Wang, H., Yu, T., Yi, G., Lin, D., & Luo, M. (2023). Detection of Citrus Psyllid Based on Improved YOLOX Model. *Plant Diseases and Pests*, *14*(1), 17-21. http://doi.org/10.19579/j.cnki.plant-d.p.2023.01.005

Wang, K., & Liu, M. (2022). Toward structural learning and enhanced YOLOv4 network for object detection in optical remote sensing images. *Advanced Theory and Simulations*, *5*(6), 2200002. http://doi.org/10.1002/adts.202200002

Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. *In Proceedings of the European Conference on Computer Vision (ECCV)*, (pp. 3-19). http://doi.org/ 10.1007/978-3-030-01234-2_1

Wu, J., Fan, P., Sun, Y., & Gui, W. (2023). Ghost-RetinaNet: Fast Shadow DetectionMethod for Photovoltaic Panels Based on Improved RetinaNet. *CMES-Computer Modeling in Engineering and Sciences*, *134*(2). http://doi.org/10.32604/cmes.2022.020919

Yao, J., Wang, Z., Liu, C., Huang, G., Yuan, Q., Xu, K., & Zhang, W. (2022). Detection method of crushing mouth loose material blockage based on SSD algorithm. *Sustainability*, *14*(21), 14386. http://doi.org/10.3390/su142114386.

Yu, J., Zhang, Y. & Zhang, W. (2023). Simulation of Autonomous Grab of Manipulator Based on YOLOv4-Tiny and RRT-Connect Algorithm. *Modeling and Simulation*, Corpus ID: 258976198. http://doi.org/10.12677/mos.2023.123254\a