Article

Bud-YOLO: A Real-Time Accurate Detection Method of Cotton Top Buds in Cotton Fields

Xuening Zhang [1] and Liping Chen [1,2,3,*]

1   College of Information Engineering, Tarim University, Alaer 843300, China; 10757223138@stumail.taru.edu.cn
2   Key Laboratory of Tarim Oasis Agriculture, Tarim University, Ministry of Education, Alaer 843300, China
3   Key Laboratory of Modern Agricultural Engineering, Tarim University, Alaer 843300, China
*   Correspondence: 119970011@mail.taru.edu.cn

Abstract: Cotton topping plays a crucial and indispensable role in controlling excessive growth and enhancing cotton production. This study aims to improve the operational efficiency and accuracy of cotton topping robots through a real-time and accurate cotton top bud detection algorithm tailored for field operation scenarios. We propose a lightweight structure based on YOLOv8n, replacing the C2f module with the Cross-Stage Partial Networks and Partial Convolution (CSPPC) module to minimize redundant computations and memory access. The network's neck employs an Efficient Reparameterized Generalized-FPN (Efficient RepGFPN) to achieve high-precision detection without substantially increasing computational cost. Additionally, the loss calculation of the optimized prediction frame was addressed with the Inner CIoU loss function, thereby enhancing the precision of the model's prediction box. Comparison experiments indicate that the Bud-YOLO model is highly effective for detecting cotton top buds, with an AP50 of 99.2%. This performance surpasses that of other YOLO variants, such as YOLOv5s and YOLOv10n, as well as the conventional Faster R-CNN model. Moreover, the Bud-YOLO model exhibits robust performance across various angles, occlusion conditions, and bud morphologies. This study offers technical insights to support the migration and deployment of the model on cotton topping machinery.

Keywords: cotton; cotton top bud; cotton topping; deep learning network; real-time detection; Bud-YOLO

1. Introduction

Cotton is a vital crop in China, which is closely related to the daily life of the nation, and China's seed cotton production reached 18.122 million tons in 2022, ranking first in the world's production [1]. Within the cotton plant ecosystem, cotton top buds play a crucial role in shaping the growth trajectory and yield [2]. Cotton topping is critical for reducing the growth of ineffective branches, regulating nutrient distribution, and promoting early and increased boll setting while minimizing shedding.

Cotton topping is typically performed through three methods: manual topping [3], chemical topping [4], and mechanical topping [5]. Manual topping is labor-intensive and inefficient; chemical topping is prone to environmental pollution; and mechanical topping is prone to omission and misdirection. Domestic research institutes have investigated mechanical equipment for cotton topping [6,7], focusing largely on automating control of the operations. In the complex environment of a cotton field, factors such as variations in lighting conditions, differences in cotton plant growth, weed growth, and changes in the size and color of cotton top buds can significantly impact the accuracy of detection and spatial positioning.

The ongoing advancements in artificial intelligence and deep learning have led to substantial progress in object detection through convolutional neural networks (CNNs). At present, mainstream object detection techniques are classified into two categories: two-stage

and one-stage methods. Two-stage methods mainly include an RCNN [8], Faster R-CNN [9], and Cascade R-CNN [10]; these detection algorithms offer higher detection accuracy but slower detection speeds. One-stage methods mainly include SSDs [11], Retinanet [12], and the YOLO [13] series; these detection algorithms have average accuracy but fast detection speeds and are widely used in several fields.

The YOLO algorithm has seen numerous and diverse applications [14,15]. However, maintaining model lightweightness while ensuring high detection accuracy during object detection remains a research challenge. In order to design a lightweight target detector for vehicle-mounted applications, Chen Xue et al. [16] proposed a new method called a Sparsely Connected Asymptotic Feature Pyramid Network (SCAFPN). Jin Gao et al. [17] used a cross-layer feature fusion network to retain model lightweightness in a model designed to detect cherry tomatoes in unstructured environments. While these methods can effectively decrease the number of model parameters, significant modifications to the model structure might lead to performance degradation or require extensive tuning.

In recent years, researchers have proposed various crop detection methods [18–20]. Traditional approaches rely on color, shape, texture, and other background aspects of the crops, extracting features through algorithms for crop recognition. For example, Longsheng Fu et al. [21] employed RGB and depth features in an R-CNN-based approach to detect apples in densely foliated fruit wall trees, facilitating robotic harvesting. Guichao Lin et al. [22] developed a reliable algorithm based on Red-Green-Blue-Depth (RGB-D) images for detecting and localizing citrus in real outdoor orchard environments for robotic picking. However, the small size of cotton top buds, along with leaf shading, light conditions, and uneven growth, poses challenges for cotton top bud detection.

Numerous researchers have conducted extensive studies on the detection and identification of small targets in complex environments [23–25], making significant progress. Yifan Bai et al. [26] proposed a real-time recognition algorithm (Improved YOLO) to accurately identify strawberry seedlings, addressing issues of small flower and fruit size, similar color, and overlapping occlusion. Yanxu Wu et al. [27] developed an enhanced end-to-end RGB-D multimodal object detection network for tea bud detection based on YOLOv7, which has an AP50 of 91.12% in the face of complex outdoor tea photography. In cotton topping object detection, it is essential to maintain the accuracy of small object detection, improve detection speed, and ensure the model's lightweight nature to support the migration and deployment of the model on cotton topping machinery.

Over the past few years, numerous improved object detection models have been introduced to tackle the difficulties associated with detecting cotton top buds [28,29]. To address the problem of cotton top bud detection, Peng Song et al. [30] proposed an improved Cascade R-CNN to detect cotton top bud regions on RGB images, and three-dimensional (3D) coordinates of objects were obtained by combining color images and depth images from RGB-D cameras. C. Wang et al. [31] drastically reduced the parameter count in the YOLOv3 model by applying deep separable convolution and enhanced the model's ability to learn multi-scale features through a hierarchical multi-scale approach. Xuan PENG et al. [32] added an object detection layer to the YOLOv5s structure, incorporating the CPP-CBAM attention mechanism with the SIoU bounding box regression loss function to improve cotton top bud detection accuracy. While these methods enhance detection accuracy to some extent, they suffer from slow detection speeds and do not account for the shooting angles of the cotton top buds.

To address the limitations and shortcomings of existing research, this paper focuses on cotton top bud recognition in complex environments and proposes an accurate, real-time recognition algorithm for natural environments based on the YOLOv8n object detection algorithm. This approach effectively handles variations in angles, shapes, and occlusion scenarios. The research improves the detection accuracy of small objects by modifying the YOLOv8n network structure, including changes to the loss function, lightweight convolution, and fusing of the pyramid feature network. Ultimately, the proposed object detection

model not only enhances detection accuracy but also lays a robust foundation for future research endeavors.

The key contributions of this study are as follows:

(1) We propose replacing the C2f module of the YOLOv8n backbone network with the Cross-Stage Partial Networks and Partial Convolution (CSPPC) lightweight module to reduce redundant computations and optimize memory access.

(2) The neck network employs the Efficient Reparameterized Generalized-FPN (Efficient RepGFPN) to achieve high-precision detection without significantly increasing computational cost.

(3) The study introduces the Inner CIoU loss function to compute regression loss, regulating the generation of auxiliary bounding boxes with a scale factor ratio to expedite convergence. The enhanced model's effectiveness in detecting cotton top buds under natural conditions has been validated through experiments, offering technical support for the advancement of intelligent cotton topping machinery.

The method was assessed and benchmarked against existing techniques using a cotton top bud dataset. The results indicate that the proposed method attains higher precision, recall, AP50, and F1 scores while maintaining real-time processing speed, significantly enhancing detection performance compared to existing methods.

The structure of the remaining sections of the paper is as follows: Section 2 elaborates on each module of the proposed Bud-YOLO model. Section 3 presents the experimental setup, results, and discussion. Section 4 summarizes the paper and highlights its main contributions.

2. Materials and Methods

2.1. Dataset Sample Collection

The cotton top bud dataset used in this paper was obtained from a cotton field in the 10th Regiment of Alaer City, Xinjiang, China. The images were collected using a smartphone HUAWEI P50E (manufactured by Huawei in Shenzhen, Guangdong Province, China) from mid-June to mid-July 2022 and had a resolution of 4069 × 3072 pixels.

During the image data acquisition process, we used two angles, namely a top shot and a side shot, and top buds with less than 30% occlusion were selected for photography. In addition to the influence of the objective factors of acquisition time on the detection of cotton top buds, their shape is also complex. The morphology of the cotton top bud varies at different developmental stages, as shown in Figure 1. The cotton top bud is smaller and lighter in color in the early stage of development, and more plump and darker in color in the later stage of development.

The collected images of cotton top bud encompass different angles, occlusion situations, and morphologies to ensure data diversity and enhance the model's robustness. Together, these samples constitute the dataset, with a total of 800 raw images collected. To verify the effectiveness of the model training, cotton top bud images were collected from mid-June to late June 2023, However, these images were not included in the training, validation, and testing datasets and were only used for prediction.

2.2. Annotation Alteration of the Dataset

We focused on cotton top bud detection under natural conditions. Manual labeling is necessary before training on the cotton top bud image data, for which we utilized the LabelImg tool [33]. Each cotton top bud was boxed and labeled as "bud", and the labeling focused on the location and category information of the cotton top bud. The labeled files were saved in PASCAL VOC format as XML files. Following annotation, the cotton top bud image data were converted to the YOLO dataset format.
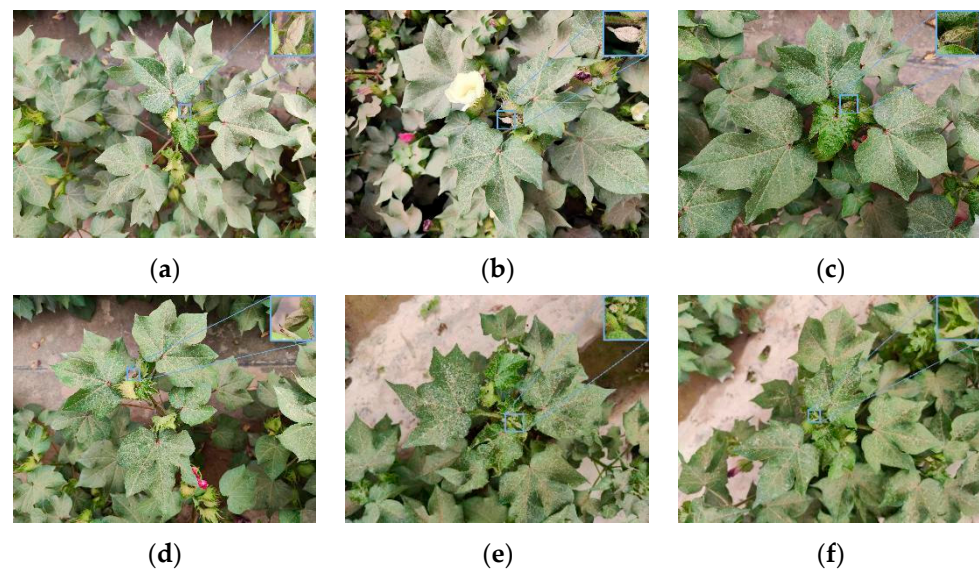
Figure 1. Images of cotton top bud data collected from mid-June to mid-July 2022. (**a**) Cotton top bud photographed from top view; (**b**) Cotton top bud photographed from side view; (**c**) Unocclusion cotton top bud; (**d**) Occlusion cotton top bud; (**e**) Early shape of cotton top bud; (**f**) Late shape of cotton top bud.

2.3. Dataset Augmentation

The sample size of the data in the cotton top bud dataset was insufficient for the model to converge during training. To improve the generalization of the model and prevent overfitting due to a lack of training data, we used data augmentation techniques, including brightness and contrast adjustments, as shown in Figure 2, resulting in a total of 4000 object samples.

Figure 2. Impact of data augmentation. (**a**) Brightness enhancement; (**b**) Brightness reduction; (**c**) Contrast enhancement; (**d**) Contrast reduction.

2.4. Dividing the Dataset

The labeled dataset was partitioned into training, validation, and test subsets with a ratio of 8:1:1. This distribution yielded 3200 images for training, 400 images for validation, and 400 images for testing.

2.5. Bud-YOLO Model for the Detection of Cotton Top Buds

2.5.1. Selection of the YOLOv8 Model

Ultralytics released the YOLOv8 algorithm in January 2023 [34], representing a significant technological improvement over the YOLOv5 object detection model. There are five versions of YOLOv8: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x, which progressively increase in width and depth. Considering the model size and its deployability on mobile platforms, the YOLOv8n model was chosen as the base for the experiments.

The YOLOv8 model consists of three main components: a backbone, neck, and head. The backbone utilizes the C2f and SPPF modules for feature extraction, adjusting the number of channels to improve the efficiency of this process. The neck network retains the PAN-FPN architecture from the YOLOv5 model to achieve bidirectional fusion of low- and high-level features, thereby improving target detection across multiple scales. The head network comprises three detection layers designed to detect features of varying scales generated by the neck network, employing an anchor-free approach to improve detection accuracy and flexibility across multiple scales. The network structure of YOLOv8 is illustrated in Figure 3:
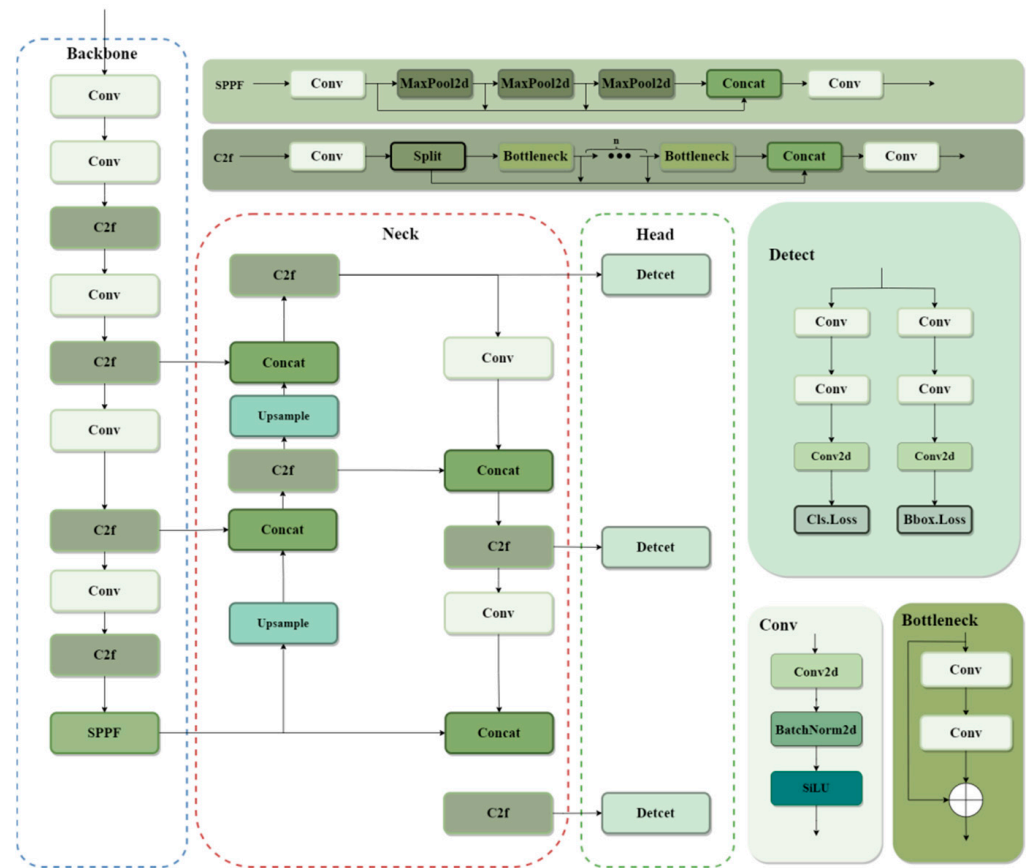

Figure 3. Network structure of YOLOV8.

2.5.2. Bud-YOLO Network Structure

For the cotton top bud dataset, we propose a lightweight detection algorithm: Bud-YOLO. The algorithm reduces the model size and improves the computational efficiency while increasing AP50 in object detection and reducing false detections and omissions, thereby enhancing model robustness.

The backbone network uses the CSPPC module, reducing redundant computations and memory access with minimal impact on detection accuracy. The neck network employs an Efficient Reparameterized Generalized-FPN (Efficient RepGFPN) to ensure high-accuracy detection without significantly increasing computational cost. Finally, the Inner CIoU function is introduced to compute the regression loss, with auxiliary bounding boxes generated based on the scale factor ratio to compute the loss and accelerate convergence. The network structure of Bud-YOLO is illustrated in Figure 4:
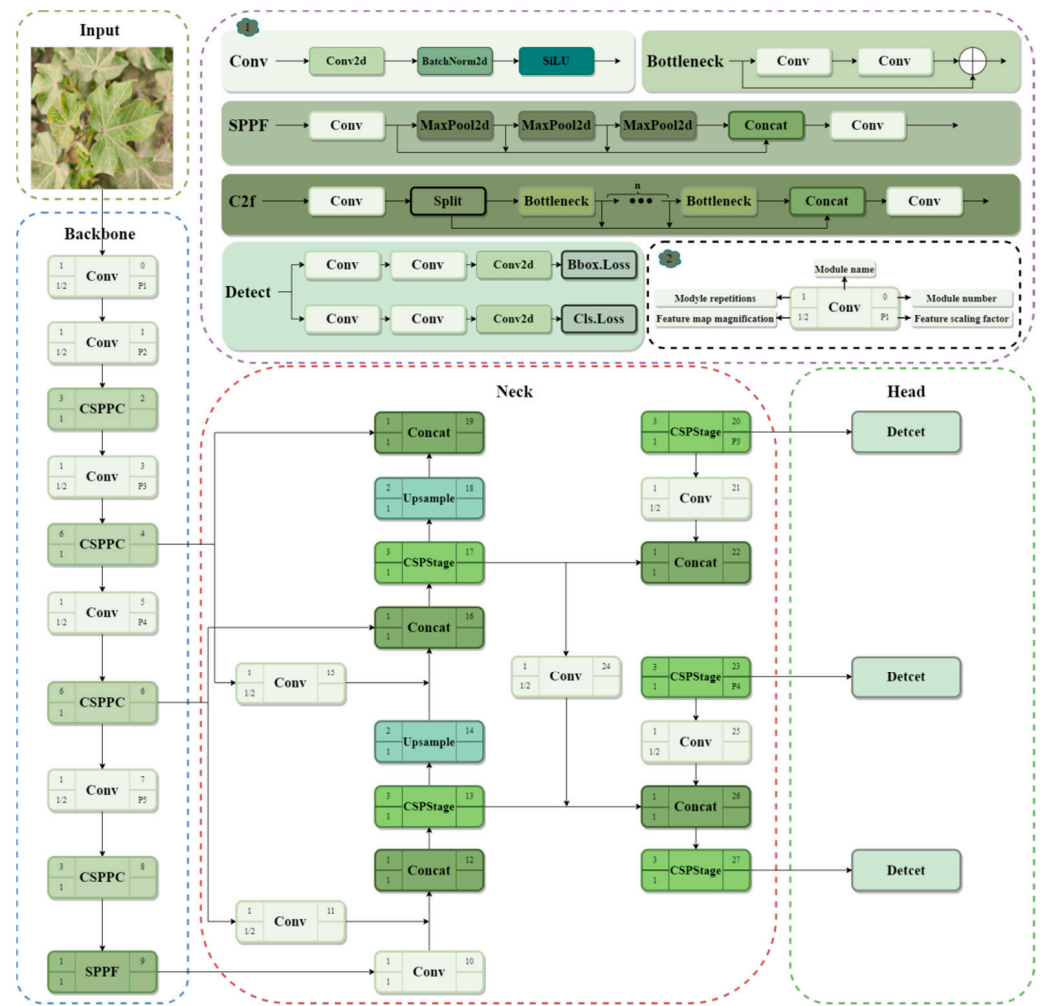
Figure 4. Network structure of Bud-YOLO. Label 1 and Label 2 show the network structure of the modules and explain the symbols in the boxes, respectively. Other modules (CSPPC, CSPStage) not specified here are described in detail below.

2.5.3. CSPPC Module

In this study, the CSPPC module proposed by Liu [35], which has two PConvs [36] in series in the output process, was used to replace the conventional convolution and reduce the number of parameters. This module replaces the conventional C2f and is incorporated into the algorithm's backbone network. This integration removes redundant channel characteristics, minimizes computational redundancy and memory accesses, accelerates detection speed, and enables the more efficient extraction of spatial features. Suppose the input size is (c, h, ω), the convolution kernel is $\kappa \times \kappa$, and the output size is (c, h, ω). Then, the FLOPs for regular convolution are shown in Equation (1) and the memory accesses are shown in Equation (2):

$$h \times \omega \times \kappa^2 \times c^2 \tag{1}$$

$$h \times \omega \times 2c + k^2 \times c^2 \approx h \times \omega \times 2c \tag{2}$$

To maintain memory continuity, consecutive channels in the front or back segments are chosen to represent the entire feature map. Then, the FLOPs of PConv are shown in Equation (3), and the memory accesses are shown in Equation (4):

$$h \times \omega \times \kappa^2 \times c_p^2 \tag{3}$$

$$h \times \omega \times 2c_p + k^2 \times c_p^2 \approx h \times \omega \times 2c_p \tag{4}$$

The architecture of the CSPPC module is depicted in Figure 5. The CSPPC module significantly reduces the model size, facilitating seamless deployment on mobile devices and lowering the costs associated with hierarchical device development.
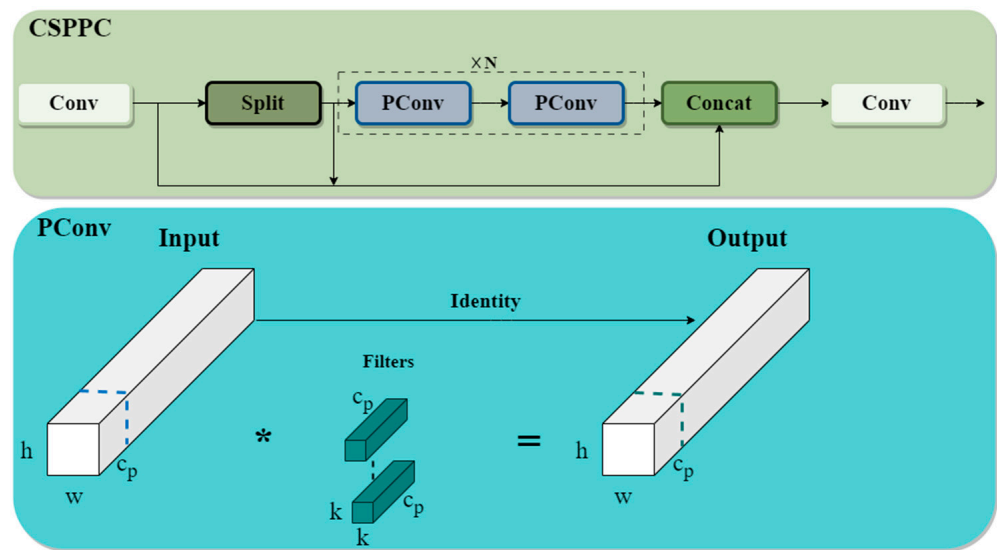

Figure 5. Structure of CSPPC network.

2.5.4. Efficient RepGFPN Module

In a Feature Pyramid Network (FPN), the purpose of multi-scale feature fusion is to combine features from various layers of the backbone network, enhancing the expressiveness of the output features and improving model performance. Traditional FPNs introduce a top-down path to fuse multi-scale features.

In this paper, we utilize the Efficient RepGFPN proposed by Xianzhe Xu et al. [37]. This Efficient RepGFPN enhances the FPN concept for object detection by fusing multi-scale features more efficiently, capturing both high-level semantics and low-level spatial details. The main improvements of the Efficient RepGFPN include the following:

(1) Adopting different channel dimensions for feature maps at different scales; optimizing performance within computational resource constraints.
(2) Reducing latency by eliminating the additional up-sampling operation in Queen-Fusion.
(3) Combining CSPNet with an Efficient Layer Aggregation Network (ELAN) and reparameterization to improve feature fusion without significantly increasing computational requirements.

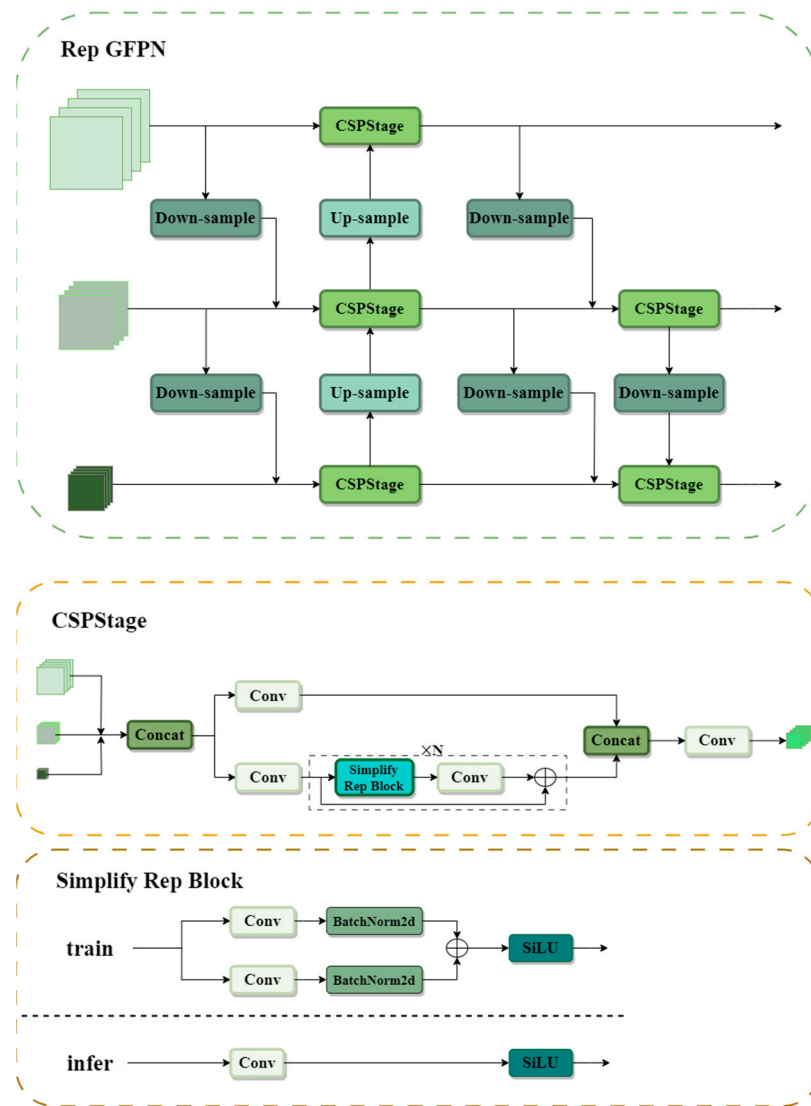The architecture of the Efficient RepGFPN network is shown in Figure 6.

Figure 6. Efficient RepGFPN network architecture diagram.

2.5.5. Improved Loss Function

The CIoU loss function in the YOLOv8n model effectively captures geometric differences in bounding boxes, thereby enhancing model positioning accuracy. However, it exhibits slower convergence and higher loss values when applied to the cotton terminal bud dataset, primarily due to the considerable variation in bud shapes. The CIoU loss function is defined as shown in Equation (8):

$$\text{IoU} = \left| \frac{b \cap b^{gt}}{b \cup b^{gt}} \right| \tag{5}$$

where the actual bounding box and the anchor box are denoted as b^{gt} and b.

$$\nu = \frac{4}{\pi^2} \left[\arctan(\frac{\omega^{gt}}{h^{gt}}) - \arctan(\frac{\omega}{h}) \right]^2 \tag{6}$$

$$\alpha = \frac{\nu}{1 - \text{IoU} + \nu} \tag{7}$$

$$\text{L}_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha\nu \tag{8}$$

In Equations (6)–(8), the width and height of the actual bounding box are denoted as ω^{gt} and h^{gt}, respectively, and those of the anchor box are denoted as ω and h; ν measures the consistency of the aspect ratio; α denotes the equilibrium parameter; b denotes the prediction frame; b^{gt} denotes the labeling frame; c is the diagonal distance encompassing both the predicted and true boxes; ρ denotes the Euclidean distance between the centroid of the prediction frame and the labeling frame; and L_{CIoU} denotes the CIoU loss function. From Equation (6), it is evident that when the aspect ratios of the predicted and labeled boxes are identical, ν equals 0. At this point, the effectiveness of the CIoU loss function is affected, leading to varying sensitivity to objects of different scales, which is particularly unfavorable for small target localization. However, due to the large number of small targets in the cotton top bud image, it is easy to miss detections using this loss function.

To address the aforementioned issues, this paper incorporates the Inner IoU loss function, as proposed by Hao Zhang et al. [38]. This method accelerates convergence by utilizing an auxiliary bounding box without introducing any additional loss terms. By distinguishing different regression samples and using various ratios of auxiliary bounding boxes to calculate the loss, the process of border regression can be effectively accelerated.

As shown in Figure 7, $\left(x_c^{gt}, y_c^{gt}\right)$ denotes the centroid of the actual bounding box and Inner actual bounding box, and $\left(x_c, y_c\right)$ denotes the centroid of the anchor box and Inner anchor box. The variable ratio corresponds to scale factors, which usually range from 0.5 to 1.5.
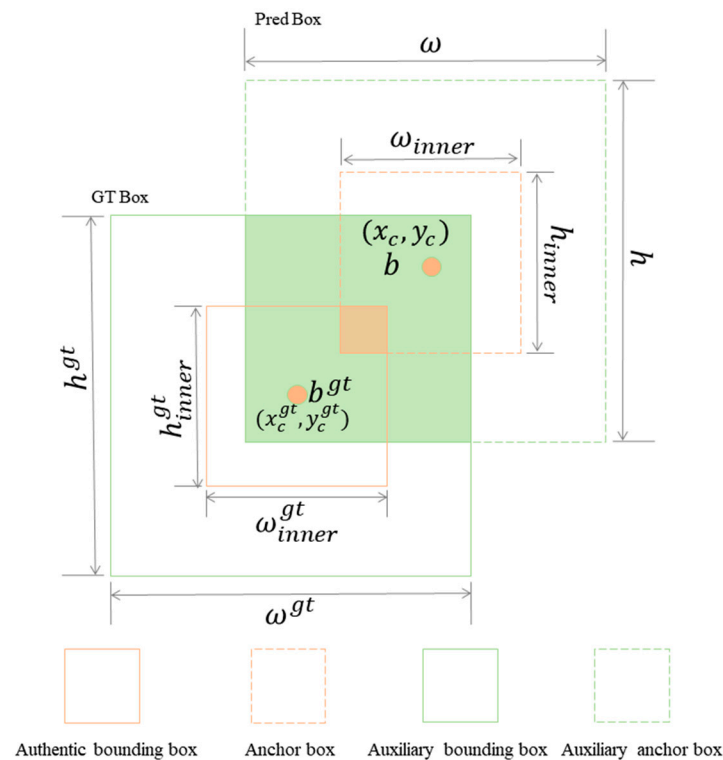

Figure 7. Schematic diagram of Inner IoU.

The Inner CIoU loss replaces the standard CIoU loss in the original loss function, and is defined as follows:

$$
\begin{cases}
b_l^{gt} = x_c^{gt} - \dfrac{\omega^{gt} * \text{ratio}}{2}, b_r^{gt} = x_c^{gt} + \dfrac{\omega^{gt} * \text{ratio}}{2} \\[2mm]
b_t^{gt} = y_c^{gt} - \dfrac{h^{gt} * \text{ratio}}{2}, b_b^{gt} = y_c^{gt} + \dfrac{h^{gt} * \text{ratio}}{2} \\[2mm]
b_l = x_c - \dfrac{\omega * \text{ratio}}{2}, b_r = x_c + \dfrac{\omega * \text{ratio}}{2} \\[2mm]
b_t = y_c - \dfrac{h * \text{ratio}}{2}, b_b = y_c + \dfrac{h * \text{ratio}}{2}
\end{cases}
\tag{9}
$$

$$
\text{inter} = \left(\min\left(b_r^{gt}, b_r\right) - \max\left(b_l^{gt}, b_l\right) \right) * \left(\min\left(b_b^{gt}, b_b\right) - \max\left(b_t^{gt}, b_t\right) \right)
\tag{10}
$$

$$
\text{union} = \left(\omega^{gt} * h^{gt}\right) * (\text{ratio})^2 + (\omega * h) * (\text{ratio})^2 - \text{inter}
\tag{11}
$$

$$
\text{IoU}^{\text{inner}} = \frac{\text{inter}}{\text{union}}
\tag{12}
$$

$$
L_{\text{Inner-CIoU}} = L_{\text{CIoU}} + \text{IoU} - \text{IoU}^{\text{inner}}
\tag{13}
$$

In Equations (9)–(13), b_l^{gt} represents the transverse coordinate of the auxiliary bounding box's left boundary, while b_r^{gt} denotes its right boundary. The scaling factor ratio controls the size of the auxiliary bounding box. The longitudinal coordinates of the auxiliary bounding box's lower and upper boundaries are represented by b_t^{gt} and b_b^{gt}, respectively. Similarly, b_l and b_r represent the transverse coordinates of the left boundary of the auxiliary anchor frame, while b_t and b_b correspond to the longitudinal coordinates of the lower and upper boundaries of the auxiliary anchor frame. The term "inter" refers to the area where the auxiliary anchor frame intersects the auxiliary bounding box, and the term "union" describes the merged area of these two regions. The IoU of Inner IoU is denoted by $\text{IoU}^{\text{inner}}$, and $L_{\text{Inner-CIoU}}$ represents the Inner CIoU loss function.

2.6. Performance Evaluation Indicators

Model performance is typically evaluated based on three key factors: accuracy, real-time processing capability, and complexity. The commonly used accuracy metrics for target detection models include precision (P), recall (R), F1 score (F1), average precision (AP), and mean average precision (mAP), which are defined as follows:

$$
P = \frac{TP}{TP + FP}
\tag{14}
$$

$$
R = \frac{TP}{TP + FN}
\tag{15}
$$

$$
F1 = \frac{2 * P * R}{P + R}
\tag{16}
$$

$$
AP = \int_0^1 P(R)\,dR
\tag{17}
$$

$$
mAP = \frac{\sum_{i=1}^{n} AP_i}{n}
\tag{18}
$$

TP represents the number of actual positive samples correctly predicted as positive, while FP denotes the number of actual negative samples predicted as positive. FN indicates the number of actual positive samples predicted as negative. P is the proportion of predicted positive samples that are actually positive. R represents the proportion of actual positive samples that are correctly predicted by the model. To balance the trade-off between precision and recall, the F1 score was introduced. AP represents the average precision for a specific class of targets across various recall points, corresponding to the area under the Precision–Recall (PR) curve. mAP is the average of the AP values across n target classes. In this study, *n* was set to 1. The average precision is expressed in terms of AP, specifically AP50 when the IoU threshold is set to 0.5.

Real-time performance was evaluated using Frames Per Second (FPS). Higher FPS values indicate better real-time detection performance. These metrics were used to evaluate the model's performance in detecting cotton top buds. Complexity metrics include FLOPs and model size, the latter referring to the size of the best model after training.

3. Results and Discussion

3.1. Experimental Platform

The hardware environment of the server platform was Intel(R) Xeon(R) Gold 6152 CPU (manufactured by Intel, Santa Clara, CA, USA) and NVIDIA GeForce RTX 3090 (24GB) GPU (manufactured by Nvidia, Santa Clara, CA, USA). The software environment was the Linux operating system, CUDA version 12.1, Python 3.10 programming language, and Pytorch 2.3.1 deep learning framework.

3.2. Experimental Parameters

The model received images with dimensions of 640×640 pixels as inputs. To optimize performance while considering the parameters, computational requirements, and memory usage associated with networks of varying depths and widths, the hyperparameters were set as follows: the epoch was 150, the batchsize was 32, the workers was 8, the initial learning rate was 0.01, the weight decay coefficient was 0.0005, the momentum parameter was 0.937, and the optimizer was Adam. To mitigate overfitting, we implemented an early stopping mechanism that terminated training if AP50 failed to exhibit significant improvement over 30 consecutive iterations.

3.3. Comparative Performance Analysis against Alternative Models

Based on the cotton top bud dataset, the Bud-YOLO model was compared with six target detection algorithms: YOLOv5s, YOLOv7-tiny, YOLOv8n, YOLOv9T, YOLOv10n, and Faster R-CNN. All models underwent training and validation in a controlled experimental environment. The detection results are presented in Table 1.

Table 1. Results of comparative experiments.

Model	P	R	AP50	F1	FPS
YOLOv5s	0.931	0.925	0.969	0.927	73.1
YOLOv7-tiny	0.841	0.778	0.853	0.808	45.1
YOLOv8n	0.990	0.960	0.991	0.974	70.9
YOLOv9T	0.962	0.897	0.973	0.928	39.1
YOLOv10n	0.964	0.941	0.987	0.952	86.9
Faster R-CNN	0.967	0.615	0.967	0.751	50.0
Bud-YOLO	0.977	0.990	0.992	0.983	69.3

The comparison experiments indicate that the AP50 of the Bud-YOLO model was 2.3%, 13.9%, 0.1%, 1.9%, 0.5%, and 2.5% higher than those of YOLOv5s, YOLOv7-tiny, YOLOv8n, YOLOv9T, YOLOv10n, and Faster R-CNN, respectively. The recall of the Bud-YOLO model exceeded these models by 6.5%, 21.2%, 0.1%, 9.3%, 4.9%, and 37.5%, respectively. Similarly, the F1 score of the Bud-YOLO model was 5.6%, 17.5%, 0.9%, 5.5%, 3.1%, and 23.2% higher than those of YOLOv5s, YOLOv7-tiny, YOLOv8n, YOLOv9T, YOLOv10n, and Faster R-CNN. The precision of the Bud-YOLO model was 0.977, with a recall of 0.99, an AP50 of 0.992, an F1 score of 0.983, and an FPS of 69.3. Although its FPS was lower than that of YOLOv8n and YOLOv10n, the Bud-YOLO model outperformed them in all other evaluation metrics, making it more suitable for detecting cotton top buds.

Figure 8 illustrates an example of the detection results obtained using the Bud-YOLO model. It demonstrates that the model is capable of effectively detecting cotton top bud under varying brightness levels, shading conditions, and shapes.

(**a**) Brightness-enhanced images of cotton top bud

(**b**) Brightness-reduced images of cotton top bud

(**c**) Unoccluded cotton top bud

(**d**) Occluded cotton top bud

(**e**) Early stage of cotton top bud

(**f**) Late stage of cotton top bud

Figure 8. Example plot of Bud-YOLO model detection results.

3.4. Effect of Inner CIoU Loss on the Model

To evaluate the impact of the Inner CIoU loss function on cotton top bud detection in Bud-YOLO, we conducted training experiments with the Bud-YOLO model using both the CIoU and Inner CIoU loss functions. The optimal weight files were then obtained for comparative analysis, and the results are shown in Table 2. Compared with the CIoU loss function, the model using the Inner CIoU (ratio = 1.0) loss function showed improvements

of 0.1%, 2.4%, 0.3%, and 1.3% in P, R, AP50, and F1 score, respectively. This combined advantage of incorporating the Inner CIoU loss function in model training is significant.

Table 2. Comparison of detection results using different loss functions with the YOLOv8n model.

Type	P	R	AP50	F1	FPS
CIoU	0.990	0.960	0.991	0.974	70.9
Inner CIoU (ratio = 0.5)	0.985	0.982	0.993	0.983	44.6
Inner CIoU (ratio = 1.0)	0.987	0.991	0.989	0.988	75.1
Inner CIoU (ratio = 1.5)	0.991	0.984	0.994	0.987	70.4

Additionally, this study verifies the convergence of the Bud-YOLO loss function. Figure 9 illustrates the curves of the two loss functions across the number of iterations. The four curves represent the edge loss when different scale factors for Inner CIoU and CIoU are used, respectively.
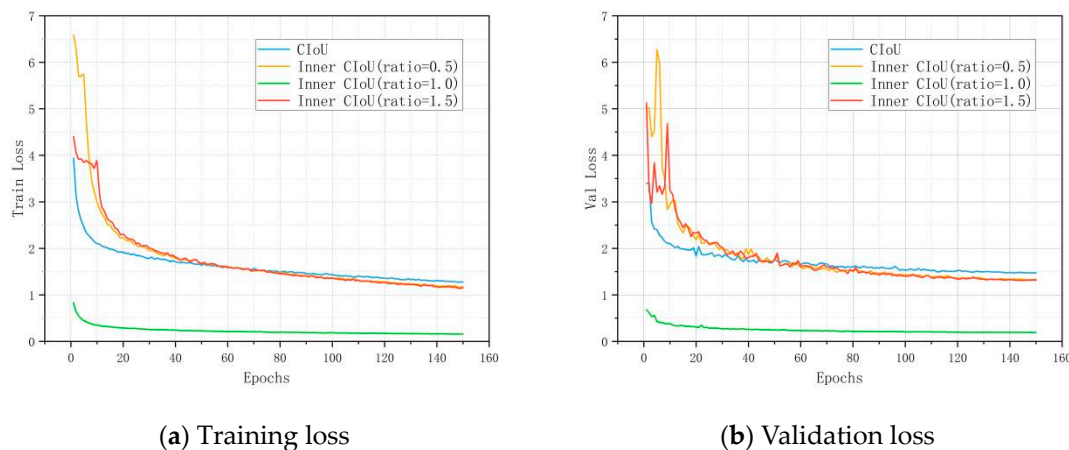

(**a**) Training loss   (**b**) Validation loss

Figure 9. Loss curves for different models.

As observed in Figure 9, all four edge losses eventually converge as the number of iterations increases. However, compared to CIoU, Inner CIoU (ratio = 1.0) has a smaller loss value and exhibits greater stability. Therefore, selecting Inner CIoU (ratio = 1.0) as the border loss function in this study could improve the model's detection performance for cotton top buds.

3.5. Ablation Experiments

The Bud-YOLO model proposed in this study is based on the YOLOv8n framework and is divided into three parts for improvement. To verify the validity of each improvement stage, ablation experiments were conducted using the experimental dataset.

As shown in Table 3, all performance metrics of the model changed following the modifications. The "√" in the table indicates that the method was used in the improvement based on the YOLOv8n model. The FLOPs and model size of the CSPPC-only model were significantly reduced. Although the model size increased by 0.6 MB with the addition of the Efficient RepGFPN compared to the CSPPC-only model, the values of P, R, and F1 score increased by 0.2%, 1.5%, and 0.9%, respectively. Compared to YOLOv8n, the Bud-YOLO model experienced a 1.3% decrease in P, a 2.25% decrease in FPS, a 3% increase in R, a 0.1% increase in AP50, and a 0.9% increase in F1 score. Despite a reduction in inference speed, the Bud-YOLO model maintained a high frame rate of 69.3 FPS.

Table 3. Comparison results before and after adding different modules.

Model	CSPPC	Efficient RepGFPN	Inner CIoU	P	R	AP50	F1	FPS	FLOPs	Model Size /MB
YOLOv8n				0.990	0.960	0.991	0.974	70.9	8.1	6.0
YOLOv8n-C	√			0.979	0.964	0.986	0.971	72.4	6.8	5.1
YOLOv8n-CG	√	√		0.981	0.979	0.983	0.980	64.6	7.1	5.7
Bud-YOLO	√	√	√	0.977	0.990	0.992	0.983	69.3	7.1	5.7

3.6. Detection Performance in Complex Scenarios in Cotton Fields

Detection performance evaluation experiments in complex cotton field scenes were conducted to assess the models' effectiveness for target detection by considering three factors: shooting angle, occlusion, and varying morphologies. Images taken from mid-June to late June 2023 were randomly selected for comparison and analysis. The detection effectiveness of the YOLOv8n and Bud-YOLO models on cotton top bud images under varying conditions was compared. Figure 10 displays some detection results, where red boxes indicate correct recognition and blue boxes indicate non-recognition. Compared to the YOLOv8n model, the Bud-YOLO model can detect cotton top buds of various morphologies, including those with different viewing angles, occlusion conditions, and different shapes. These results demonstrate the robust performance of the Bud-YOLO model.

Figure 10. Detection results of 2023 cotton top bud image. (**a**) YOLOv8n detects cotton top bud photographed from top view; (**b**) Bud-YOLO detects cotton top bud photographed from top view; (**c**) YOLOv8n detects cotton top bud photographed from side view; (**d**) Bud-YOLO detects cotton top bud photographed from side view; (**e**) YOLOv8n detects unoccluded cotton top bud; (**f**) Bud-YOLO

detects unoccluded cotton top bud; (**g**) YOLOv8n detects occluded cotton top bud; (**h**) Bud-YOLO detects occluded cotton top bud; (**i**) YOLOv8n detects early cotton top bud; (**j**) Bud-YOLO detects early cotton top bud; (**k**) YOLOv8n detects late cotton top bud; (**l**) Bud-YOLO detects late cotton top bud.

3.7. Discussion

Comparison, ablation, and detection experiments were conducted to verify the performance of the improved Bud-YOLO model for detecting cotton top buds in natural scenes. The comparison experiments demonstrated that the Bud-YOLO model achieved the highest mAP value. Although the FPS of the Bud-YOLO model is lower than that of YOLOv8n and YOLOv10n, the speed loss is acceptable, and its inference speed meets real-time requirements. The ablation experiments indicate that the CSPPC module improves the inference speed of the model and reduces its size without significantly affecting the AP50 of the algorithm. Additionally, the inclusion of the Efficient RepGFPN module enhances the recall of the model without adding more parameters, mitigating missed detections of cotton top buds with varying shapes and under occlusion conditions. The Inner CIoU loss function enhances the P, R, AP50, and F1 score, and simultaneously accelerates the convergence of the model, stabilizing the loss value at 0.15. The model accounts for leaf shading, different angles, and various shapes of cotton top buds, exhibiting a high detection rate and strong resistance to external environmental conditions. Therefore, the model is more robust and effective in detecting cotton top buds under complex natural scenarios.

4. Conclusions

This study proposes a Bud-YOLO detection algorithm capable of accurately identifying cotton top buds in real-time. Initially, a dataset of cotton top bud images in complex natural scenes was constructed, labeled using LabelImg (version 1.8.1). A total of 800 labeled images were selected, and through data expansion, a dataset containing 4000 images was generated. A network architecture for the accurate real-time detection of cotton top buds was proposed, utilizing the CSPPC lightweight convolution module to replace the C2f module in the backbone network, thereby reducing redundant computations and memory access with minimal impact on detection accuracy. Incorporating an Efficient RepGFPN in the neck network maintains high accuracy in cotton top bud detection without significantly increasing computational costs. Finally, the Inner CIoU loss function was introduced to compute the regression loss, with the generation of auxiliary bounding boxes controlled by a proportional factor ratio to compute the loss and accelerate convergence. Comparative experimental results indicate that the Bud-YOLO model achieved a precision of 0.977, a recall of 0.99, an AP50 of 0.992, an F1 score of 0.983, and an FPS of 69.3, meeting the real-time detection requirements. The performance evaluation experiments demonstrate that the Bud-YOLO model achieves a high detection rate in complex natural scenes, including varying shooting angles, occlusion, and different morphologies.

In future work, we plan to extend the cotton thimble image dataset by incorporating images taken under various weather conditions (e.g., sunny, cloudy, and rainy) and different lighting scenarios. This will enhance the model's generalization across different environmental conditions, improving the reliability of cotton toppers in field operations. Although the Bud-YOLO model shows promising results, further optimization such as model pruning should be explored to reduce the model size and complexity, aiming for more efficient deployment in real-world applications.

Author Contributions: Conceptualization, X.Z. and L.C.; methodology, X.Z.; software, X.Z.; validation, X.Z.; formal analysis, X.Z.; investigation, X.Z.; resources, X.Z.; data curation, X.Z. and L.C.; writing—original draft preparation, X.Z.; writing—review and editing, L.C.; visualization, X.Z.; supervision, L.C.; project administration, L.C.; funding acquisition, L.C. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are available on request from the corresponding author(s).

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

References

1. Kang, Y. *International Statistical Yearbook 2023*; China Statistics Press: Beijing, China, 2023.
2. Nie, J.; Li, Z.; Zhang, Y.; Zhang, D.; Xu, S.; He, N.; Zhan, Z.; Dai, J.; Li, C.; Li, W.; et al. Plant pruning affects photosynthesis and photoassimilate partitioning in relation to the yield formation of field-grown cotton. *Ind. Crop. Prod.* **2021**, *173*, 114087. [CrossRef]
3. Renou, A.; Téréta, I.; Togola, M. Manual topping decreases bollworm infestations in cotton cultivation in Mali. *Crop Prot.* **2011**, *30*, 1370–1375. [CrossRef]
4. Shi, F.; Li, N.; Khan, A.; Lin, H.; Tian, Y.; Shi, X.; Li, J.; Tian, L.; Luo, H. DPC can inhibit cotton apical dominance and increase seed yield by affecting apical part structure and hormone content. *Field Crop. Res.* **2022**, *282*, 108509. [CrossRef]
5. Aydın, İ.; Arslan, S. Mechanical properties of cotton shoots for topping. *Ind. Crop. Prod.* **2018**, *112*, 396–401. [CrossRef]
6. Wang, J.; Hu, B.; Jia, S.; Xue, X.; Li, Z. Optimization and Experiment of 3MDZ-18 Hydraulic Driven Cotton Topping Machine. *J. Agric. Mech. Res.* **2023**, *45*, 170–174. [CrossRef]
7. Zhang, J.; Zheng, C.; Zhao, J.; Zhang, R.; Zhao, X.; Zhao, F. Design and Experiment of Auto-follow Row Device for Cotton Topping Machine. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 93–101.
8. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [CrossRef]
9. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
10. Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving Into High Quality Object Detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162. [CrossRef]
11. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer: Cham, Switzerland, 2016; pp. 21–37. [CrossRef]
12. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2999–3007. [CrossRef]
13. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [CrossRef]
14. Zhao, C.; Bai, C.; Yan, L.; Xiong, H.; Suthisut, D.; Pobsuk, P.; Wang, D. AC-YOLO: Multi-category and high-precision detection model for stored grain pests based on integrated multiple attention mechanisms. *Expert Syst. Appl.* **2024**, *255*, 124659. [CrossRef]
15. Cai, Y.; Yao, Z.; Jiang, H.; Qin, W.; Xiao, J.; Huang, X.; Pan, J.; Feng, H. Rapid detection of fish with SVC symptoms based on machine vision combined with a NAM-YOLO v7 hybrid model. *Aquaculture* **2024**, *582*, 740558. [CrossRef]
16. Xue, C.; Xia, Y.; Wu, M.; Chen, Z.; Cheng, F.; Yun, L. EL-YOLO: An efficient and lightweight low-altitude aerial objects detector for onboard applications. *Expert Syst. Appl.* **2024**, *256*, 124848. [CrossRef]
17. Gao, J.; Zhang, J.; Zhang, F.; Gao, J. LACTA: A lightweight and accurate algorithm for cherry tomato detection in unstructured environments. *Expert Syst. Appl.* **2024**, *238*, 122073. [CrossRef]
18. Gao, W.; Zong, C.; Wang, M.; Zhang, H.; Fang, Y. Intelligent identification of rice leaf disease based on YOLO V5-EFFICIENT. *Crop Prot.* **2024**, *183*, 106758. [CrossRef]
19. Chen, X.; Liu, T.; Han, K.; Jin, X.; Wang, J.; Kong, X.; Yu, J. TSP-yolo-based deep learning method for monitoring cabbage seedling emergence. *Eur. J. Agron.* **2024**, *157*, 127191. [CrossRef]
20. Tang, Z.; Lu, J.; Chen, Z.; Qi, F.; Zhang, L. Improved Pest-YOLO: Real-time pest detection based on efficient channel attention mechanism and transformer encoder. *Ecol. Inform.* **2023**, *78*, 102340. [CrossRef]
21. Fu, L.; Majeed, Y.; Zhang, X.; Karkee, M.; Zhang, Q. Faster R-CNN-based apple detection in dense-foliage fruiting-wall trees using RGB and depth features for robotic harvesting. *Biosyst. Eng.* **2020**, *197*, 245–256. [CrossRef]

22. Lin, G.; Tang, Y.; Zou, X.; Li, J.; Xiong, J. In-field citrus detection and localisation based on RGB-D image analysis. *Biosyst. Eng.* **2019**, *186*, 34–44. [CrossRef]

23. Li, G.; Suo, R.; Zhao, G.; Gao, C.; Fu, L.; Shi, F.; Dhupia, J.; Li, R.; Cui, Y. Real-time detection of kiwifruit flower and bud simultaneously in orchard using YOLOv4 for robotic pollination. *Comput. Electron. Agric.* **2022**, *193*, 106641. [CrossRef]

24. Wang, T.; Zhang, K.; Zhang, W.; Wang, R.; Wan, S.; Rao, Y.; Jiang, Z.; Gu, L. Tea picking point detection and location based on Mask-RCNN. *Inf. Process. Agric.* **2023**, *10*, 267–275. [CrossRef]

25. Shuai, L.; Mu, J.; Jiang, X.; Chen, P.; Zhang, B.; Li, H.; Wang, Y.; Li, Z. An improved YOLOv5-based method for multi-species tea shoot detection and picking point location in complex backgrounds. *Biosyst. Eng.* **2023**, *231*, 117–132. [CrossRef]

26. Bai, Y.; Yu, J.; Yang, S.; Ning, J. An improved YOLO algorithm for detecting flowers and fruits on strawberry seedlings. *Biosyst. Eng.* **2024**, *237*, 1–12. [CrossRef]

27. Wu, Y.; Chen, J.; Wu, S.; Li, H.; He, L.; Zhao, R.; Wu, C. An improved YOLOv7 network using RGB-D multi-modal feature fusion for tea shoots detection. *Comput. Electron. Agric.* **2024**, *216*, 108541. [CrossRef]

28. Yin, L.; Wu, J.; Liu, Q.; Wu, W. Improved YOLOv5s recognition of cotton top buds with fusion of attention and feature weighting. In Proceedings of the International Conference on Algorithm, Imaging Processing, and Machine Vision (AIPMV 2023), Qingdao, China, 15–17 September, 2023; Zhou, H., Yang, Q., Eds.; International Society for Optics and Photonics, SPIE: Bellingham, DC, USA, 2024; Volume 12969, p. 1296928. [CrossRef]

29. Li, J.; Zhi, X.; Wang, Y.; Cao, Q. Research on Intelligent recognition system of Cotton apical Bud based on Deep Learning. *J. Phys. Conf. Ser.* **2021**, *1820*, 012134. [CrossRef]

30. Song, P.; Chen, K.; Zhu, L.; Yang, M.; Ji, C.; Xiao, A.; Jia, H.; Zhang, J.; Yang, W. An improved cascade R-CNN and RGB-D camera-based method for dynamic cotton top bud recognition and localization in the field. *Comput. Electron. Agric.* **2022**, *202*, 107442. [CrossRef]

31. Wang, C.; He, S.; Wu, H.; Teng, G.; Zhao, C.; Li, J. Identification of Growing Points of Cotton Main Stem Based on Convolutional Neural Network. *IEEE Access* **2020**, *8*, 208407–208417. [CrossRef]

32. Peng, X.; Zhou, J.; Xu, Y.; Xi, G. Cotton top bud recognition method based on YOLOv5-CPP in complex environment. *Trans. Chin. Soc. Agric. Eng.* **2023**, *39*, 191–197.

33. Tzutalin. LabelImg. 2015. Available online: https://github.com/tzutalin/labelImg (accessed on 12 September 2023).

34. Jocher, G.; Chaurasia, A.; Qiu, J. Ultralytics YOLO (Version 8.1.0). 2024. Available online: https://github.com/ultralytics/ultralytics (accessed on 25 March 2024).

35. Liu, R.M.; Su, W.H. APHS-YOLO: A Lightweight Model for Real-Time Detection and Classification of Stropharia Rugoso-Annulata. *Foods* **2024**, *13*, 1710. [CrossRef] [PubMed]

36. Chen, J.; Kao, S.H.; He, H.; Zhuo, W.; Wen, S.; Lee, C.H.; Chan, S.H.G. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 12021–12031. [CrossRef]

37. Xu, X.; Jiang, Y.; Chen, W.; Huang, Y.; Zhang, Y.; Sun, X. DAMO-YOLO: A Report on Real-Time Object Detection Design. *arXiv* **2023**, arXiv:2211.15444.

38. Zhang, H.; Xu, C.; Zhang, S. Inner-IoU: More Effective Intersection over Union Loss with Auxiliary Bounding Box. *arXiv* **2023**, arXiv:2311.02877.