# Design and Experimental Evaluation of an Intelligent Sugarcane Stem Node Recognition System based on Enhanced YOLOv5s

**Jiuxiang Dai**
South China Agricultural University
**Yangfan Luo**
South China Agricultural University
**Shenye Shi**
South China Agricultural University
**Dantong Yang**
South China Agricultural University
**Zuoxi Zhao**
fit@stu.scau.edu.cn

South China Agricultural University

---

**Research Article**

**Keywords:** Sugarcane, YOLOv5, real-time detection, edge computing device, computer vision

**Posted Date:** April 4th, 2024

**DOI:** https://doi.org/10.21203/rs.3.rs-4166889/v1



**Additional Declarations:**
No competing interests reported.

Supplementary Tables are not available with this version.

---

# Design and Experimental Evaluation of an Intelligent Sugarcane Stem Node Recognition System based on Enhanced YOLOv5s

**Jiuxiang Dai[a,b], Yangfan Luo[a,b], Shenye Shi[a,b], Dantong Yang[a,b,*], Zuoxi Zhao[a,b,*]**

[a] College of Engineering, South China Agricultural University, Guangzhou 510642, China

[b] Ministry of Education Key Technologies and Equipment Laboratory of Agricultural Machinery and Equipment in South China, South China Agricultural University, Guangzhou 510642, China

* Corresponding author. E-mail address: zhao_zuoxi@scau.edu.cn (Z.Z.); Tel.: +86-136-0004-9101; yangdt@scau.edu.cn (D.Y.)

## Abstract

The rapid and accurate identification of sugarcane internodes is of great significance for tasks such as field operations and precision management in the sugarcane industry, and it is also a fundamental task for the intelligence of the sugarcane industry. However, in complex field environments, traditional image processing techniques have low accuracy, efficiency, and are mainly limited to server-side processing. Meanwhile, the sugarcane industry requires a large amount of manual involvement, leading to high labor costs. In response to the aforementioned issues, this paper employed YOLOv5s as the original model algorithm, incorporated the K-means clustering algorithm, and added the CBAM attention module and VarifocalNet mechanism to the algorithm. The improved model is referred to as YOLOv5s-KCV. We implemented the YOLOv5s-KCV algorithm on Jetson TX2 edge computing devices with a well-configured runtime environment, completing the design and development of a real-time sugarcane internode identification system. Through ablation experiments, comparative experiments of various mainstream visual recognition network models, and performance experiments conducted in the field, the effectiveness of the proposed improvement method and the developed real-time sugarcane internode identification system were verified. The experimental results demonstrate that the improvement method of YOLOv5s-KCV is effective, with an algorithm recognition accuracy of 89.89%, a recall rate of 89.95%, and an mAP value of 92.16%, which respectively increased by 6.66%, 5.92%, and 7.44% compared to YOLOv5s. The system underwent performance testing in various weather conditions and at different times in the field, achieving a minimum recognition accuracy of sugarcane internodes of 93.5%. Therefore, the developed system in this paper can achieve real-time and accurate identification of sugarcane internodes in field environments, providing new insights for related work in sugarcane field industries.

**Keywords:** Sugarcane, YOLOv5, real-time detection, edge computing device, computer vision

## 1. Introduction

China, as one of the world's major sugarcane producers, is heavily reliant on sugarcane for its national economy, serving as a fundamental raw material for sugar and fuel production. Sugarcane also holds significant medicinal value. However, field operations in sugarcane cultivation still heavily rely on manual labor. Currently, only land cultivation and transportation in the sugarcane industry have been mechanized, while other tasks such as sugarcane yield measurement still require significant manual involvement. This labor-intensive approach results in issues such as high labor intensity, inflated labor costs, and low efficiency. Real-time and accurate identification of sugarcane internodes is crucial for the automation and efficiency enhancement of sugarcane production and serves as the foundation for sugarcane yield visual measurement. With the advancement of computer technology, an increasing number of computer applications are being integrated into the agricultural sector.

Currently, the identification of sugarcane features relies mainly on image processing techniques, which not only have low efficiency but also demand specific working environments. Moshashai et al. [4] conducted preliminary research and utilized grayscale image threshold segmentation to identify sugarcane internodes, laying the initial practical foundation for using image processing techniques in sugarcane internode identification. Shangping Lu et al. [5] explored machine vision methods for extracting and identifying sugarcane internode features. They utilized threshold segmentation on the S/H components of sugarcane images in the HSV color space to generate composite images, integrated the feature indicators of these composite images, and then employed support vector machine classification to identify internodes and internode blocks. Additionally, they applied clustering analysis to identify sugarcane internodes against a monochrome background. Yanmei Meng et al. [6] proposed a sugarcane node identification algorithm based on multi-threshold and multi-scale wavelet transforms, although this algorithm requires prior removal of sugarcane leaves before internode identification. Jiqing Chen et al. [7] initially converted collected sugarcane RGB images into HSV color images, established vertical projection functions for regions of interest, determined the minimum value points of these functions, and preliminarily located sugarcane nodes based on the obtained minimum points. Rui Yang et al. [8] proposed a method capable of identifying multiple internodes in a segment of sugarcane by utilizing image stitching and the Sobel operator to obtain gradient images and subsequently identify sugarcane nodes.

The aforementioned referenced studies are all based on traditional image processing techniques, which rely on the reflection of certain lighting and surface characteristics to identify object contours and subsequently extract target features. Moreover, these methods were conducted in relatively simple working environments. Real-time identification of sugarcane internodes in complex and dense background environments poses certain challenges. Nonetheless, these research methods provide foundational theories for feature recognition, including sugarcane internodes.

With the rapid development of the field of deep learning, visual recognition technology has been widely applied in various domains of life. In the agricultural sector, applications such as maturity detection [9], fruit selection [10], [11] and crop harvesting [12], [13] have already begun. These studies provide valuable references for the application of deep learning convolutional neural network algorithms in agriculture. TIAN et al. [14] proposed an improved version of the You Only Look Once v3 (YOLOv3) model for detecting apples of different maturities under various lighting conditions, complex backgrounds, and occlusions. Parvathi S et al. [15] used the deep learning-based neural network model Faster R-CNN to identify and select coconuts. Dihua Wu et al. [16]

discussed and compared the recognition accuracy of field apple blossoms using the YOLOv4 and YOLOv3 algorithms. Deep learning represents a novel and efficient method for intelligent sugarcane cultivation. Preeti Saini et al. [17] proposed a novel hybrid method based on CNN-Bi-LSTM_CYP deep learning for predicting sugarcane crop yield, but this method is only capable of measuring yield using current data images. Srivastava S et al. [18] proposed a novel network framework based on VGG and Inception for identifying sugarcane diseases and pests. Da Wang et al. [19] employed the YOLOv5 network model to identify sugarcane internodes and designed a pre-cutting system for sugarcane varieties based on visual recognition technology, enabling cutting of different types of sugarcane. The aforementioned studies indicate that deep learning network models can provide new insights for sugarcane-related agricultural work.

The recognition of features in complex field environments can have a negative impact, with common factors affecting the accuracy of fruit detection including crop overlap, interference from plant structures, and lighting conditions. Therefore, the key to addressing these issues is to enhance the performance of crop detection algorithms. This paper addresses the drawbacks of low efficiency, high manual involvement, and labor costs faced by current sugarcane industry-related work by designing a real-time sugarcane internode identification system based on an improved YOLOv5 network model algorithm. This system is capable of identifying sugarcane internodes in any field environment and simultaneously obtaining a large amount of sugarcane internode information, greatly improving work efficiency while laying the groundwork for intelligent operations in sugarcane fields.

# 2. Materials and Methods

*2.1. Dataset Preparation*

The sugarcane plants growing in natural environments are often affected by various factors such as partial obstruction by sugarcane leaves and different degrees of tilting, posing higher demands for the comprehensiveness and diversity of data collection. The images in the dataset should ideally encompass various natural weather conditions, including both cloudy and sunny conditions. Furthermore, data augmentation techniques such as rotation, noise addition, and cropping should be employed to expand the dataset, thereby enhancing the robustness and recognition accuracy of the trained model [20].

The data collected for this experiment were obtained from sugarcane fields in Zhanjiang and Leizhou cities, Guangdong Province, China. Image acquisition methods included handheld and tripod-mounted ZED brand cameras, with an image resolution of 512×512 pixels. A total of 1066 images were captured during two sessions (April 2022 and January 2023). Since sugarcane mostly grows in outdoor natural environments, the data collection process should comprehensively consider the impact of natural environmental conditions. Factors such as shooting time, angle, presence of sugarcane leaves, and shooting distance were considered as classification features for image acquisition. Fig. 1 illustrates three scenarios of internode obstruction: (a) where internodes are mostly exposed without obstruction by sugarcane leaves or other sugarcane plants, (b) where some internodes are partially obstructed by sugarcane leaves, and (c) where almost all internodes are obstructed by sugarcane leaves. To enable the recognition algorithm to learn from various objective scenes, images from these three scenarios were provided to the algorithm for learning. Additionally, to enhance the learning effectiveness and robustness of the model, lighting conditions were also taken into account, including front, side, and back lighting. Different lighting conditions were simulated by adjusting the camera's shooting angle to obtain sugarcane photos in different light directions, including with, perpendicular to, and against the direction of light propagation, mimicking front, side, and back lighting, respectively. Furthermore, images were collected depicting individual sugarcane internodes, single sugarcane stalks, tilted dense clusters of sugarcane stalks, and multiple sugarcane stalks within the camera's field of view, covering various recognition target states as shown in Fig. 2.

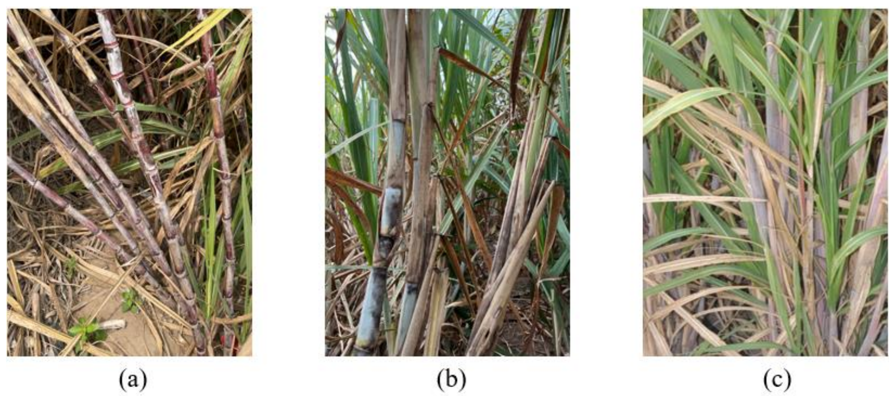



**Fig. 1** Schematic representation of sugarcane node obstruction at different levels

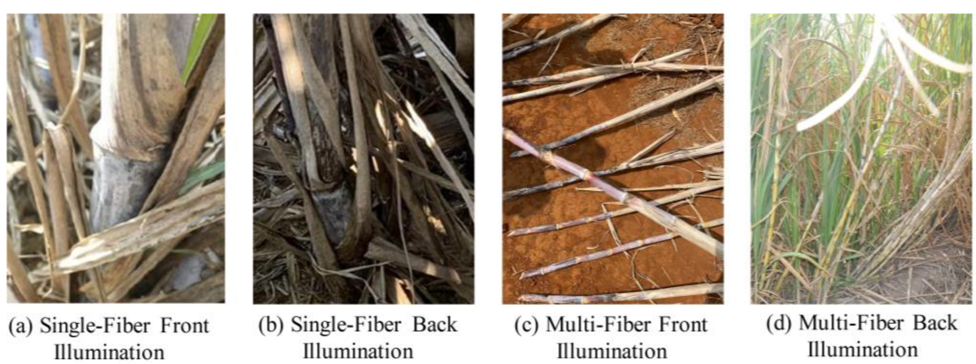



**Fig. 2** Schematic representation of sugarcane photography at various angles

During dataset acquisition, it was not possible to guarantee that all images could be used. Therefore, to enhance the robustness and detection accuracy of the model, we selected clear and usable images before annotating image features. Images that were blurry due to lack of focus or unusable for various other reasons were excluded, deliberately retaining those with a significant number of sugarcane plants under various lighting and background conditions. This approach facilitated better learning and understanding of the detailed features of sugarcane internodes during the model training process, thus improving its recognition accuracy. After selection, we ultimately obtained 976 high-quality images. Additionally, to enhance the robustness of model training and network generalization, we used web crawling technology to collect an additional 180 sugarcane images, resulting in a total of 1156 original experimental images.

### *2.2. Annotation of Dataset Features*

The sugarcane image dataset was manually labeled using the LabelImg software, with each image being annotated with bounding boxes. The label name assigned was "jingjie," and the labeling information was stored in text files in a format recognizable by YOLO. Fig. 3 illustrates the interface used for annotating sugarcane internodes, with the labeled bounding boxes aimed at facilitating the rapid identification of internode features during training and learning processes. Following the organization of experimental data, the entire dataset was fully annotated to ensure an adequate learning volume while preventing overfitting of the neural network during training. The LabelImg image annotation tool was used to annotate sugarcane internode features in all experimental data, resulting in a total of 11,106 annotated features.

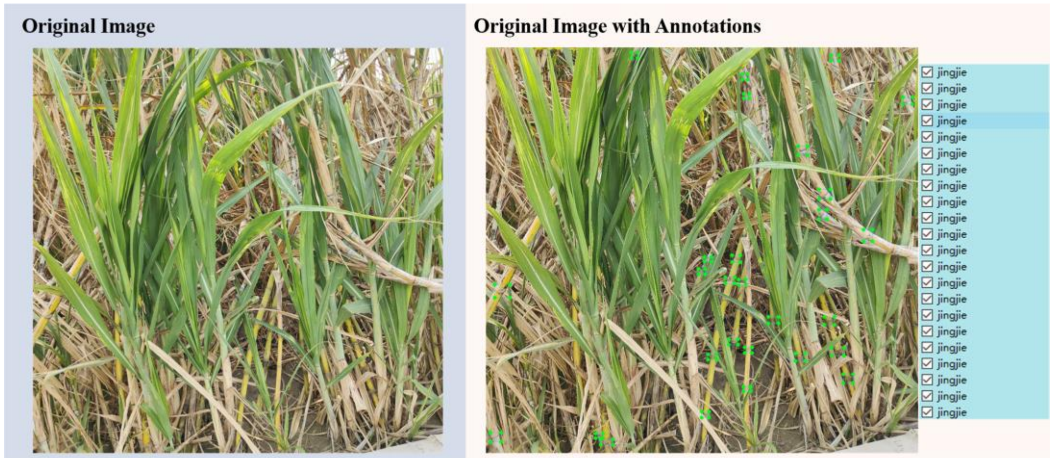



**Fig. 3** Schematic representation of dataset annotation

*2.3. Image Augmentation in the Dataset*

To avoid overfitting, increase the volume of training data, enhance model generalization, and address sample imbalances, we performed small-scale and effective data augmentation [21], [22] on the existing dataset. The main augmentation techniques included random rotation, noise addition, and random cropping, primarily to address the uncertainty of sugarcane plant growth under natural conditions. Fig. 4 demonstrates the effect of image augmentation, with the expanded dataset containing a greater variety of sugarcane states and morphologies, aiming to consider the operational environment in actual sugarcane fields as much as possible and to include a greater number of sugarcane internodes in the images. In this study, Imgaug [23] was employed as the code tool for data augmentation. Imgaug is an open-source Python package suitable for various augmentation techniques, providing simplicity, convenience, and efficiency.

The final experimental dataset comprised 2056 images, containing 26,768 internode features. All experimental data were divided into training, validation, and test sets in a ratio of 7:2:1, with the data volumes for each dataset shown in Table 1.


**Fig. 4** Schematic Representation of Data Augmentation Effects

Table 1 Dataset and Corresponding Quantities

| Dataset | Training | Validation | Test | Total |
|---|---|---|---|---|
| Image | 1440 | 411 | 205 | 2056 |
| Label | 18738 | 6692 | 1338 | 26768 |

## 2.4. YOLOv5

YOLOv5, proposed by Glen Jocher et al. in 2020, is a single-stage object detection algorithm that has been widely applied across various industries [24]. Building upon the YOLOv4 network model [25], YOLOv5 introduces several new improvements, offering smaller storage footprint, higher precision, and faster execution speed. It presents stronger advantages in the rapid deployment of network models, greatly facilitating the deployment of algorithms onto edge computing devices. YOLOv5 consists of four versions: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, each catering to different computational demands, ranging from simple to complex calculations. As the computational complexity increases, so does the model's volume and precision.

The YOLOv5 network model algorithm primarily comprises four parts: the input end, backbone, neck, and head. The backbone is responsible for extracting target features and passing them to the neck for further processing. The neck structure utilizes Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) to enhance target features through forward and backward fusion [26]. The head generates anchor boxes to identify features enhanced by the neck and determine feature categories [27]. Considering the need for real-time detection of sugarcane internodes on edge computing devices, this study adopts YOLOv5s as the network baseline model, whose structure is illustrated in Fig. 5.

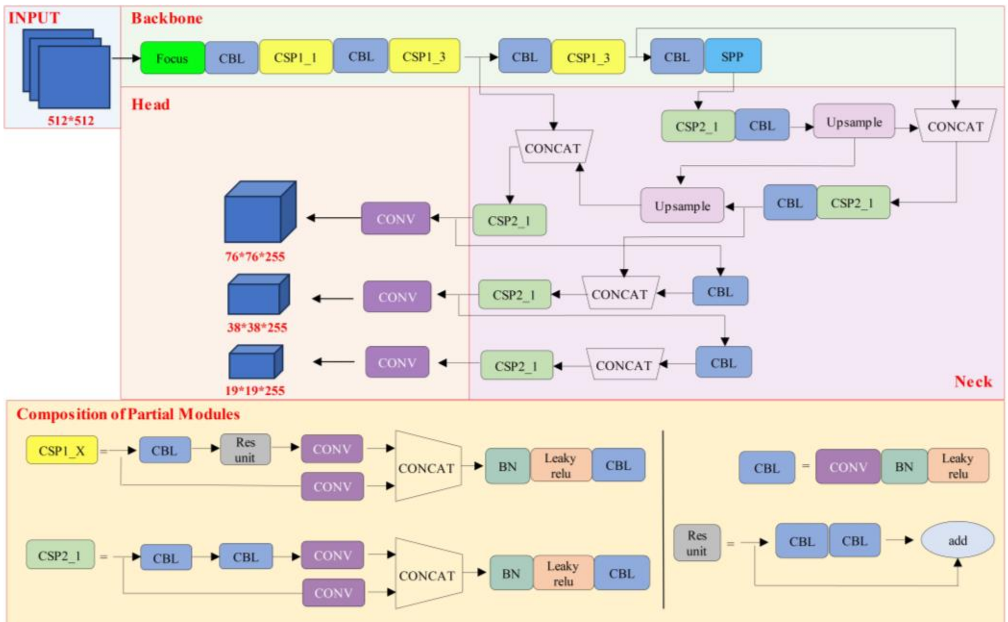



**Fig. 5** Schematic Representation of YOLOv5 Network Architecture

*2.5 Improvements to YOLOv5s*

Although this study focused solely on the identification of sugarcane internodes, the objective was to transplant the model onto edge computing devices and perform real-time detection of sugarcane internodes in complex field environments. In order to enhance the operational performance of the model, the following improvements were made to the YOLOv5s network model:

*2.5.1 K-means Clustering Algorithm*

The K-means clustering algorithm [28] divided all data samples into K clusters based on the distances between samples, aiming to ensure that the points within each cluster were as close as possible while maximizing the distances between clusters. After assigning all samples to their respective clusters, a new round of distance calculation was conducted. This iterative process continued until certain specified conditions were met. To expedite convergence and stability within clusters, efforts were made to maintain the centers unchanged during the sample selection process. The objective function formula of the K-means clustering algorithm is shown in Equation (1), where *i* represents the number of cluster sets, *K* denotes the cluster centers, and *N* indicates the number of samples contained in the cluster.

$$J=\sum_{i=1}^{N}\sum_{k=1}^{K}||\mathrm{x}_{\mathrm{i}}-\mu_k||^2 \tag{1}$$

After annotating the sugarcane stem features in the images using the Labelimg tool, corresponding txt files were generated. These files recorded the positional information of all features in the original images, including the coordinates of the top-left corner of the annotation box ($x$, $y$), as well as the width ($w$) and height ($h$) of the annotation box, following the format shown in Equation (2). The K-means clustering algorithm was then employed to calculate the heights and widths of as many anchor boxes as possible. Subsequently, the intersection over union (IOU) between each anchor box and each labeled box was computed, and the maximum IOU was selected through comprehensive evaluation. This iterative calculation process continued until the dimensions of the anchor boxes best suited for the sugarcane stem features were determined. The calculation formulas are depicted in Equations (3) and (4), where $w$ represents the width of the anchor box during each calculation, $h$ represents the height of the anchor box during each calculation, $W$ denotes the computed width of the anchor box, and $H$ denotes the computed height of the anchor box.

*<object-class><x><y><width><height>* (2)

$$W_i' = \frac{1}{N_i}\sum w_i \tag{3}$$

$$H_i' = \frac{1}{N_i}\sum \mathrm{h}_i \tag{4}$$

In sugarcane fields, issues such as complex environmental conditions, significant mutual plant occlusion, and small sugarcane stem features presented certain difficulties in identifying sugarcane stems. The default sizes of anchor boxes in the YOLOv5 network model are determined based on features from the COCO dataset, which are not suitable for sugarcane stems. Moreover, the sizes of stem anchor boxes are mostly very similar and follow a certain pattern of data values. Therefore, recalculating the sizes of anchor boxes suitable for sugarcane stems using the K-means clustering algorithm is deemed meaningful.

### *2.5.2. CBAM Attention Module*

To enhance the accuracy of small object detection, the CBAM module was introduced into the backbone network [29], [30]. Comprising Channel Attention Module (CAM) and Spatial Attention Module (SAM), the CBAM module processes feature maps transmitted to the neural network. Successively passing through the CAM and SAM, the newly acquired attention feature maps optimize the initial input feature maps. The fusion of these two attention modules not only simplifies the model structure but also reduces computational pressure. In this study, the CBAM attention module was incorporated into the YOLOv5s network architecture, effectively increasing the weight of small objects across the entire feature map for easier network learning. The flowchart of the CBAM algorithm is illustrated in Fig. 6.

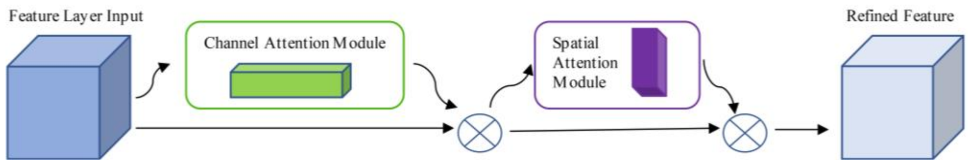



**Fig. 6** Structural Diagram of the CBAM Attention Mechanism Module

The Channel Attention Mechanism: The input feature map undergoes successive processing through max-pooling and average-pooling layers, resulting in the compression of feature map dimensions. The compressed feature map, after passing through the Share MPL module, reduces the channel count to only 1% of the original, which is then expanded back to the original number of channels. After applying the ReLU activation function, two activated results are obtained. These two output results are combined element-wise, and after being processed through the sigmoid activation function, a weight value is obtained. This weight value is then merged with the original feature map, serving as the input to SAM. The input content is defined as *Mc(F)* value, as detailed in Equation (5).

Spatial Attention Mechanism: For the input feature layer, the maximum and average values are taken on each channel at every feature point. These values are then combined, and after adjusting the convolutional channel count and processing the fusion result with the sigmoid activation function, the weight value of each feature point in the input feature layer (ranging from 0 to 1) is obtained. Multiplying this weight value by the original input feature layer yields the *Ms(F)* value, as expressed in Equation (6). Here, $F$ represents the input feature map, $c$ denotes the number of neurons, $W_0$ and $W_1$ represent the feature dimensions of the MLP, and $\sigma$ represents the sigmoid function.

$$M_c(F) = \sigma(W_1(W_0(F^c_{avg})) + W_1(W_0(F^c_{\max}))) \tag{5}$$

$$M_s(F) = \sigma(\int^{7\times 7}(F^s_{avg}; F^s_{\max})) \tag{6}$$

The dataset utilized in this study primarily consisted of images directly captured from the field, resulting in an intertwining layout of sugarcane plants. Moreover, due to the small size of sugarcane stem features, they were challenging to identify. Therefore, the CBAM attention module enhanced the network model's focus on identifying sugarcane stem joints, further ensuring the reliability and accuracy of this system for field application.

### *2.5.3. VarifocalNet Mechanism*

In determining the accuracy of target feature recognition and the precision of their positioning, the common practice involves using Intersection over Union-based Classification Score (IACS) [31]. Depending on the value of IACS, adjustments can be effectively made to refine the object detection results. The Varifocal Loss focuses the detector on a sparse set of challenging instances and predicts IACS on the dense feature detector obtained after training. It combines star bounding box feature representations to estimate IACS and improve rough bounding boxes. Additionally, a novel efficient star-shaped bounding box feature representation was devised for estimating IACS and enhancing rough bounding boxes. By integrating these two new modules and the bounding box optimization branch, while eliminating redundant network data, a new object detector named VarifocalNet [32] was constructed, with its loss function VFL formula presented in (7). Here, $\alpha$ represents sample evaluation weights, $\gamma$ denotes sample adjustment parameters, $p$ signifies the perceptual classification score, and $q$ denotes the detected target IOU score.

$$VFL(p,q)=\begin{cases}-q(q\log(p)+(1-q)\log(1-p) & q>0\\ -\alpha p^{\gamma}\log(1-p) & q=0\end{cases} \tag{7}$$

In the dataset of sugarcane nodes, dense overlapping of sugarcane nodes made it difficult to be recognized. Therefore, the VarifocalNet mechanism was introduced to enhance the recognition of fine and dense sugarcane node features by the network model.

In summary, we named the improved model as YOLOv5s-KCV, which clearly indicates the locations of the CBAM attention module and VarifocalNet mechanism within the network model, marked with red boxes. Its structure is shown in Fig. 7.

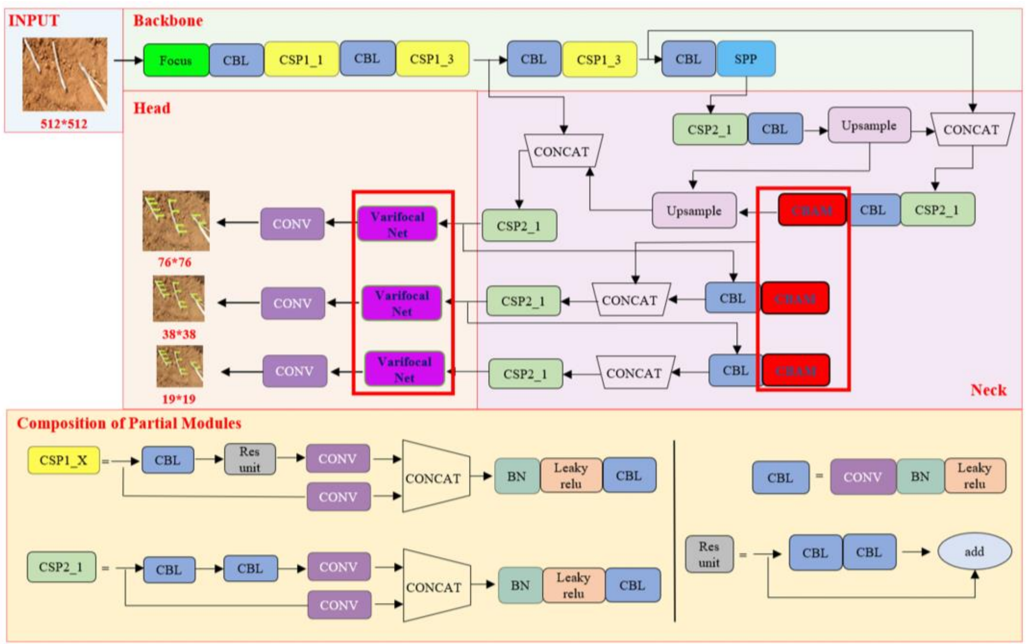



**Fig. 7** Schematic Representation of YOLOv5s-KCV Structure

*2.6. Model Training*

*2.6.1. Training Platform and Parameter Settings*

The operating system of the training platform is Windows 10, with an Intel(R) Core(TM) i5-7500CPU processor and 16GB of RAM. It is equipped with an NVIDIA GeForce GTX 1060 3GB graphics card. CUDA 10.0 and Cudnn7.4.1 are used, along with convenient and efficient libraries such as PyTorch and OpenCV, running on the PyCharm compiler. PyCharm serves as the program editor, while PyTorch serves as the deep learning framework.

To achieve rapid detection of sugarcane nodes, we first pretrained the model weights of YOLOv5s on the COCO dataset and made appropriate adjustments to the network model size and depth. The key parameters set during training are as follows: batch size of 4, epoch value of 300, learning rate of 0.001, and categories value of 1.

### *2.6.2. Evaluation Metrics*

The dataset contained approximately 26,768 targets of uneven sizes distributed across various locations in all images. The dataset essentially covered all scenarios and positions of nodes in sugarcane fields, which positively contributed to improving model training effectiveness and recognition accuracy. Precision, recall, and mean Average Precision (mAP) were utilized in this study to measure the detection performance of the model [33]. The inference time per single image (ms/pic) was employed to gauge the detection speed of the model [34]. Precision serves as a performance metric for classification models, representing the proportion of samples correctly predicted by the model to the total number of samples. Recall, another performance metric for classification models, signifies the ratio of successfully detected positive class samples to the total number of positive class samples. mAP is used to evaluate the performance of object detection models, particularly in multi-class object detection, measuring the model's accuracy across different categories while considering class imbalances. The formulas defining these metrics are presented in equations (9)-(11), where $P$ denotes precision (%), $R$ represents recall (%), mAP signifies the average accuracy (%), C denotes the number of categories needing identification in the sample, $TP$ indicates the number of positive class samples successfully identified by the model, *FP* represents the number of negative class samples incorrectly identified as positive by the model, and *FN* represents the number of positive class samples not correctly identified by the model.

$$P = \frac{TP}{TP + FP} \times 100\% \tag{8}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{9}$$

$$mAP = \sum_{i=1}^{C} AP_i / C \tag{10}$$

*2.7. Real time detection system for sugarcane stem nodes*

To enable the improved network model to operate in the field, we designed and developed a real-time sugarcane node detection system [35]. The sugarcane node real-time detection system designed in this paper is based on the Jetson TX2 edge computing device. The platform of this system can be powered by a portable mobile power supply, meeting the requirements for simplicity and portability of field operation equipment. This system can be deployed and operated in sugarcane fields, allowing for the rapid acquisition of a large number of sugarcane plant images and the quick identification of sugarcane nodes. This reduces manual labor and significantly improves work efficiency. Table 2 shows the basic configuration of the Jetson TX2 platform, which supports multiple processors to accelerate the computation speed of neural network models. The YOLOv5s-KCV model is deployed on the Jetson TX2 as the core algorithm support. Fig. 8 illustrates a working example of the system, while the overall workflow of this paper is clearly presented in the form of a flowchart, as shown in Fig. 9.

Table 2 Basic Configuration of Jetson TX2 Platform.

| Hardware | Configure | Environment | Version |
|---|---|---|---|
| System | Ubuntu18.04 | CUDA | 10.2 |
| GPU | NVIDIA Pascal Architecture 256×CUDA Core | Cudnn | 7.6.2 |
| CPU | 64×Denver 2 CPU 64×ARM Cortex-A57 CPU | Pytorch | 1.6.0 |
| RAM | GB LPDDR4 RAM | JetPack | 4.5 |
| Hard disk | 32GB eMMC | Cudatoolkit | 10.8.2 |

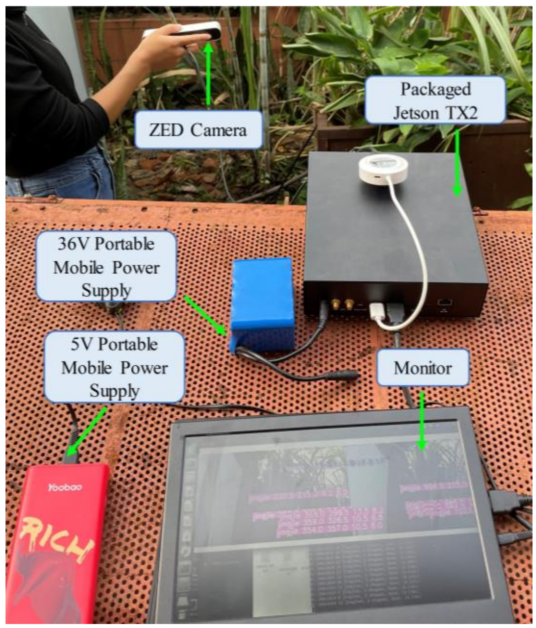



**Fig. 8** Real-Time Sugarcane Nodes Detection System Example.

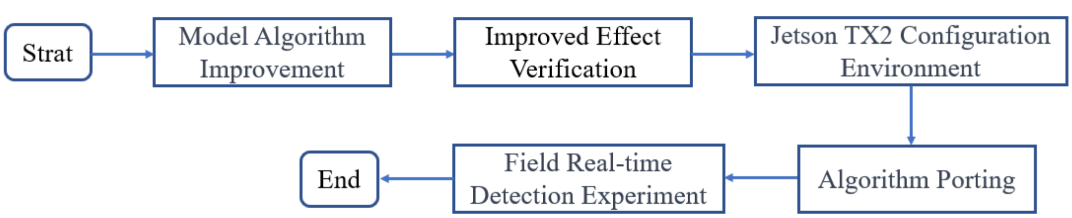



**Fig. 9** Workflow diagram in this paper

# 3. Results and Analysis

## 3.1. Ablation Experiments

The ablation experiments conducted on the experimental platform mentioned in section 2.6.1 demonstrated the effectiveness of the proposed method for sugarcane node recognition. Detailed experimental results are presented in Table 3. YOLOv5s-KCV achieved a precision of 89.89%, recall of 89.95%, and mAP of 92.16% for sugarcane node recognition, outperforming all other experimental methods. By replacing SE and ECA with CBAM on the basis of YOLOv5s-KCV, the experimental results showed that the CBAM attention module had a better effect on sugarcane node recognition. It can be observed from the experimental results that the addition of the CBAM attention module led to improvements in all metrics, demonstrating its effectiveness in recognizing dense features. Although the detection speed of YOLOv5s-KCV per image was 23ms, slightly slower than YOLOv5s+k-means, there was a significant improvement in sugarcane node detection.

Our experimental results indicate that the addition of K-means, CBAM, and VarifocalNet can indeed improve the performance of the YOLOv5 model. However, we need comparative experiments on network models to further analyze the impact of these improvements on the model and ensure that they can generate practical benefits in real-world applications.

Table 3 Comparison of Experimental Results among Different Models.

| Method | Precision(%) | Recall(%) | mAP(%) | Detection Speed (ms/pic) |
|---|---|---|---|---|
| YOLOv5s | 83.23 | 84.03 | 84.72 | 25 |
| YOLOv5s+k-means | 85.66 | 86.06 | 85.68 | 22 |
| YOLOv5s+CBAM | 84.69 | 85.27 | 85.46 | 23 |
| YOLOv5s+VarifocalNet | 84.07 | 84.79 | 84.94 | 23 |
| YOLOv5s+k-means+CBAM | 87.16 | 87.64 | 89.67 | 23 |
| YOLOv5s+k-means+VarifocalNet | 86.71 | 86.52 | 89.06 | 22 |
| YOLOv5s+CBAM+VarifocalNet | 87.24 | 87.44 | 90.22 | 24 |
| YOLOv5s+SE+VarifocalNet+k-means | 83.55. | 83.23 | 83.46 | 24 |
| YOLOv5s+ECA+VarifocalNet+k-means | 84.16 | 84.20 | 84.46 | 23 |
| YOLOv5s-KCV | 89.89 | 89.95 | 92.16 | 23 |

### *3.2. Comparative Analysis of Various Network Models*

To further validate the effectiveness of the improvements proposed in this study, we conducted comparative experiments between YOLOv5s-KCV and mainstream visual recognition network models including VGG19, YOLOv4, YOLOv5s, and the newer version YOLOv7. The experimental results are presented in Table 4 YOLOv5s-KCV showed performance improvements of 5.22%, 5.96%, and 7.2% in various metrics compared to YOLOv7, which exhibited the best performance among the other models. The inference time per image increased by 28ms/pic. In terms of detection speed performance, the ranking from fastest to slowest is YOLOv5s-KCV, YOLOv4, YOLOv5s, YOLOv7, VGG19. Although YOLOv4 had the same single-image detection time as YOLOv5s-KCV, its detection performance was relatively poor, with an mAP of only 84.64%. This makes it unsuitable for real-time detection of sugarcane nodes in the field, making YOLOv5s-KCV the preferred model in this experiment. The experimental results further demonstrate that the improvement methods proposed in this study have a positive effect on the accuracy of sugarcane node feature recognition and detection speed in YOLOv5s. Fig. 10 provides a more intuitive comparison of performance among different models.

Fig. 11 presents a comparison of the detection results of each model algorithm on the same image. In the image, we marked the nodes that were not detected with red, and those features that were incorrectly detected as nodes with green, and labeled each detection with a number. VGG19 failed to detect nodes marked 1 and 2 in the image; YOLOv4 incorrectly identified other features as nodes at position 3 and missed detecting the node at position 4; YOLOv5s yielded consistent results with YOLOv4. YOLOv7 did not misidentify other features as nodes but missed detecting the node at position 7. YOLOv5s-KCV successfully detected all nodes in the image and did not misidentify other features as nodes. Fig. 12 illustrates the detection results before and after model improvements. The experimental results confirm the effectiveness of the improvement approach proposed in this study.

Table 4 Comparative Experimental Results of Different Models.

| Model | Precision(%) | Recall(%) | mAP(%) | Detection Speed (ms/pic) |
|---|---|---|---|---|
| VGG19 | 82.84 | 83.66 | 84.27 | 98 |
| YOLOv4 | 82.89 | 84.89 | 84.64 | 22 |
| YOLOv5s | 83.23 | 84.03 | 84.72 | 25 |
| YOLOv7 | 84.67 | 83.99 | 84.96 | 50 |
| YOLOv5s-KCV | 89.89 | 89.95 | 92.16 | 22 |

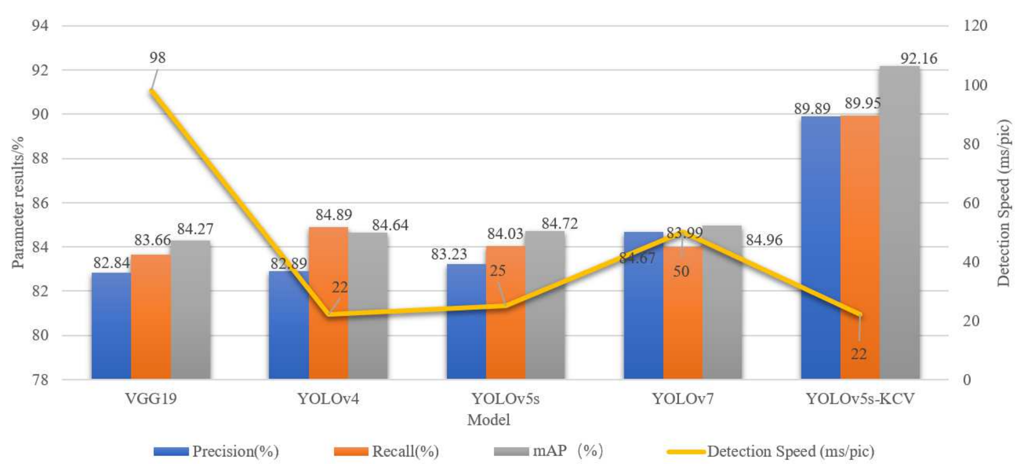



**Fig. 10** Performance Comparison of Different Models

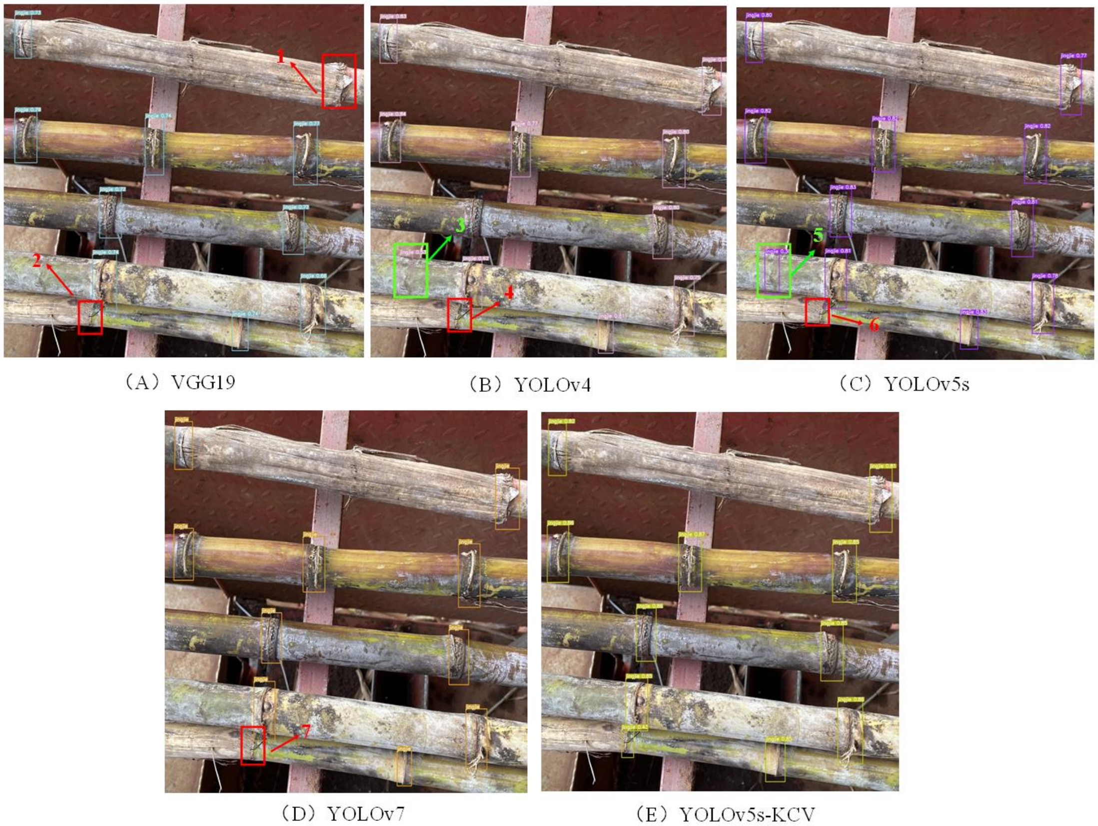
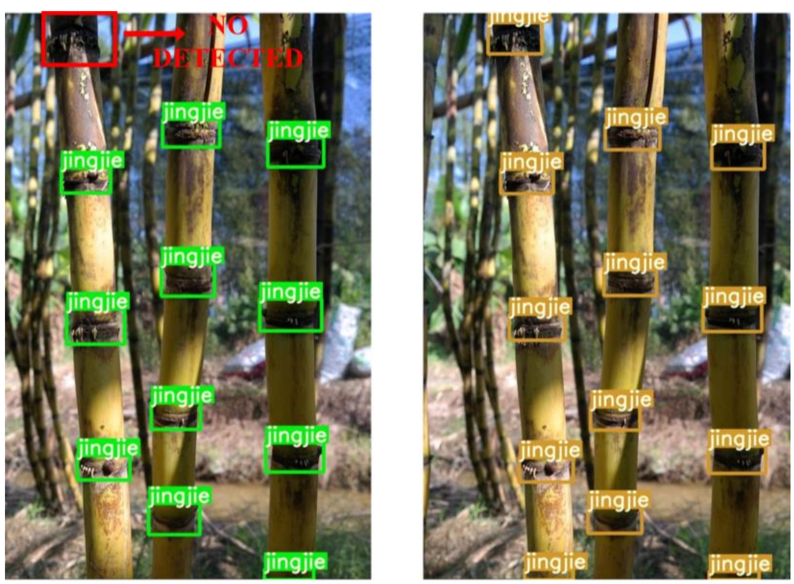



**Fig. 11** Comparative Detection Results Among Different Models




YOLOv5s YOLOv5s-KCV

**Fig. 12** Comparative Detection Results Before and After Model Improvement

### *3.3. Sugarcane Nodes Real-Time Detection System*

#### *3.3.1. Field Experiment of the Real-Time Sugarcane Nodes Detection System*

To ensure that the system can perform effectively in complex field environments, field experiments were conducted in sugarcane fields in Qujie Town, Xuwenzhen County, Zhanjiang City, Guangdong Province, China, as depicted in Fig. 13.

To compare the accuracy of the system's operation, we randomly selected plants from different areas of the experimental field and manually counted the number of nodes in each selected area. We then compared these manually counted node numbers with the system's recognized node numbers to determine the reliability of the system's operation.

We manually counted the number of sugarcane nodes in the selected samples and conducted recognition experiments on the selected samples using the system. To minimize the impact of errors on the test results, samples were taken from different plants in different areas, and sugarcane nodes were consecutively detected. Subsequently, both manual and system recognition accuracy were evaluated and statistically assessed. Details of the sampling statistics, manual confirmation, and system recognition accuracy are presented in Table 5.

**Fig. 13** Example of Sugarcane Field

Table 5 Sampling Statistics and Comparative Experiments between Manual and System Recognition.

| Area | Plants | Stem Nodes by Artificial | Target by System | Stem Node by System |
|---|---|---|---|---|
| 1 | 20 | 360 | 349 | 347 |
| 2 | 25 | 475 | 452 | 449 |
| 3 | 19 | 342 | 335 | 334 |
| 4 | 21 | 399 | 389 | 386 |
| 5 | 28 | 448 | 426 | 423 |

Based on the analysis of the experimental results, the system demonstrates a recognition accuracy and recall rate for sugarcane nodes of over 94%, indicating high reliability and meeting the demands of field agricultural work.

Furthermore, to further demonstrate the stability of the system in field operations, stability tests were conducted under different environmental conditions to identify sugarcane nodes. A range of target samples comprising 20 plants with a total of 355 nodes was selected. The experiments were conducted in varying weather conditions, including overcast and sunny weather during the morning, noon, and evening. The experimental interface of the system is shown in Fig. 14, and detailed experimental results are presented in Fig. 15. Analysis of the experimental results reveals that the system's overall performance in sunny conditions surpasses that in overcast conditions, with the best performance observed during sunny noon hours. Relatively poorer performance is noted in overcast conditions during the evening; however, the detection accuracy still reaches 93.5%, fully meeting the system's field operational requirements.

The experiments confirm the effectiveness and feasibility of the system's algorithms, demonstrating its ability to operate in various environments and conditions in sugarcane fields.

Moreover, the system's accuracy and performance in field operations meet the basic requirements for such operations.

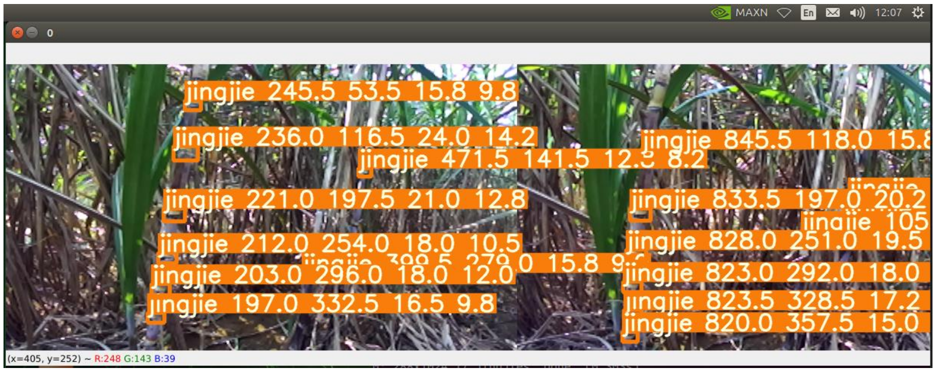



**Fig. 14** System Interface Example

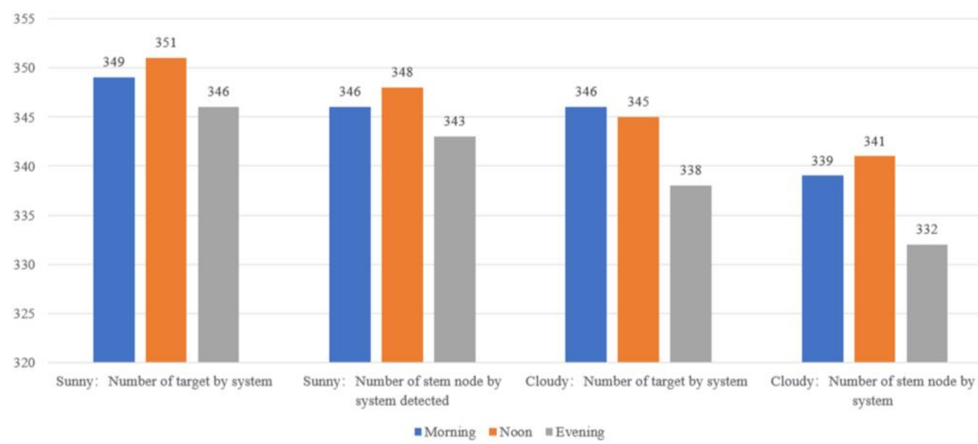



**Fig. 15** Experimental Results of System Stability Testing

*3.3.2. Summary*

The recommended methodology for this system includes the following steps:

1. Randomly select the sugarcane field area to be identified.
2. Assemble and set up the system according to the configuration shown in Fig. 8.
3. Capture photos or videos using the system, and leverage the improvements made to YOLOv5s as described in Section 2.5 of this article to enhance the identification of sugarcane nodes in complex field environments.

4. Visualize and display the identification results of sugarcane nodes on a portable display screen.

5. Based on the identification results of sugarcane nodes, perform tasks such as yield estimation or harvesting.

# 4. Advantages and limitations

The proposed YOLOv5s-KCV significantly improved the detection performance of the neural network model. YOLOv5s-KCV's advantages lie in: (1) Sugarcane nodes, characterized by small and dense features, underwent targeted improvements for accurate identification, resulting in a 6.66% increase in recognition accuracy. Ablation experiments (Table S2) and comparative experiments with other models (Table S3) demonstrated that YOLOv5s-KCV performed exceptionally well in ters of recognition accuracy and speed. (2) YOLOv5s-KCV had a small memory footprint and ran fast, enabling successful deployment on Jetson TX2. Feasibility experiments conducted in the field validated the practical usability of the system. Although many studies have made excellent algorithmic improvements, no actual field experiments were conducted. For example, Xu et al. [36] made significant improvements to the YOLO algorithm but did not test the improved model's field performance. Although YOLOv5s-KCV could be deployed on Jetson TX2 and demonstrated advantages in terms of small size, fast operation, and high precision, there is still room for improvement in future research. For instance, while the accuracy increased by 4.39% compared to YOLOv4, the processing speed remained at 22ms/pics, indicating some room for improvement in terms of operational speed [37]. Additionally, our feasibility experiments were conducted in various environments, but the model's performance was still influenced by environmental factors. For instance, in overcast nighttime conditions, the recognition accuracy of the YOLOv5s-KCV model dropped to its lowest [38].

## 5. Discussion and Conclusions

This study aimed to enhance the level of intelligence in the sugarcane farming industry, particularly in operations such as sugarcane yield measurement and harvesting. We developed a real-time sugarcane node recognition system based on an improved YOLOv5 algorithm, which combines image acquisition hardware with deep learning algorithms to achieve efficient and accurate sugarcane node recognition. To address the challenges posed by the complex environment of sugarcane fields, we made improvements to the YOLOv5s algorithm. By using the K-means clustering algorithm to obtain anchor box sizes suitable for sugarcane node features, we improved training efficiency and recognition accuracy. Additionally, we introduced the CBAM attention module and Varifocalnet mechanism, enhancing the algorithm's recognition accuracy and speed for dense and small features.

The results of ablation experiments demonstrate the effectiveness of the improvements made to the YOLOv5s algorithm in sugarcane node recognition, showing excellent performance. Comparative experiments with various models reveal that YOLOv5s-KCV achieved a precision rate of 89.89%, recall rate of 88.95%, and mAP of 90.16% for sugarcane node recognition. Compared to VGG19, YOLOv3, and YOLOv4, our algorithm exhibited significant advantages in all metrics. Furthermore, the inference time for a single image was as low as 22ms/pic, meeting the real-time detection requirements in the field. In tests conducted under different weather and time conditions, the system achieved a minimum sugarcane node detection accuracy of 93.5%.

In conclusion, this study enables real-time detection of sugarcane nodes in the field, but further improvements are needed to enhance the operational performance of the system in field conditions, providing more reliable equipment for the sugarcane industry. Continuous refinement and optimization of the system are also key areas for future research.

## Declarations

- **Funding**

The authors would like to acknowledge the support of this study by, the State Key Research Program of China (Grant No. 2022YDF2001901-01), the Guangdong Provincial Department of Agriculture's Modern Agricultural Innovation Team Program for Animal Husbandry Robotics (Grant No. 2019KJ129), the post-Expert Project of National Sugarcane Industry Technical System (Grant No. CARS-170404), the Major Science and Technology Project of Yunnan Province (Grant No. 202102AE090028).

- **Conflict of interest/Competing interests**

All authors have no conflicts of interest and have agreed to the publication of the article.

- **Data availability, Materials availability, Code availability**

We deeply appreciate the requirements of SCI journals regarding data sharing. However, due to policies within our laboratory or confidentiality agreements, we are unable to provide the raw data. We have meticulously described the experimental design, analysis, results, as well as the process of data analysis and handling. Should the editors and reviewers have any queries regarding specific data, we will make every effort to provide detailed explanations and clarifications.

- **Author contributionEthical Approval**

Conceptualization, Jiuxiang Dai, Dantong Yang and Zuoxi Zhao; methodology, Jiuxiang Dai and Dantong Yang; software, Jiuxiang Dai; validation, Yangfan Luo and Shenye Shi; formal analysis, Yangfan Luo; investigation, Shenye Shi; resources, Dantong Yang; data curation, Yangfan Luo; writing—original draft preparation, Jiuxiang Dai; writing—review and editing, Jiuxiang Dai and Yangfan Luo; Visualization, Jiuxiang Dai; project administration, Zuoxi Zhao and Dantong Yang; funding acquisition, Zuoxi Zhao All authors have read and agreed to the published version of the manuscript.

- **Ethical Approval**

**Not Applicable.**

## References


[1] Nare, B.; Tewari, V.K.; Kumar Chandel, A.; Prakash Kumar, S.; Chethan, C.R. A mechatronically integrated autonomous seed material generation system for sugarcane: A crop of industrial significance. Ind. Crops Prod. 2019, 128, 1–12. [CrossRef]

[2] Sara E. Ali, Qin Yuan, Shengpeng Wang, Mohamed A. Farag,More than sweet: A phytochemical and pharmacological review of sugarcane (Saccharum officinarum L.),Food Bioscience,Volume 44, Part B,2021,101431,ISSN 2212-4292,https://doi.org/10.1016/j.fbio.2021.101431.

[3] Angelo Aliano Filho, Daniela Renata Cantane, Paulo Roberto Isler, Helenice de Oliveira Florentino,An integrated multi-objective mathematical model for sugarcane harvesting considering cumulative degree-days,Expert Systems with Applications,Volume 232,2023,120881,ISSN 0957-4174,https://doi.org/10.1016/j.eswa.2023.120881.

[4] Moshashai, K.; Almasi, M.; Minaei, S.; Borghei, A.M. Identification of sugarcane nodes using image processing and machine vision technology. Int. J. Agric. Res. 2008, 3, 357–364.

[5] Lu, S.; Wen, Y.; Ge, W. Recognition and features extraction of sugarcane nodes based on machine vision. Trans. Chin. Soc. Agric.Mach. 2010, 41, 190–194.

[6] Meng, Y.; Ye, C.; Yu, S.; Qin, J.; Zhang, J.; Shen, D. Sugarcane node recognition technology based on wavelet analysis. Comput.Electron. Agric. 2019, 158, 68–78.

[7] Chen, J.; Wu, J.; Qiang, H.; Zhou, B.; Xu, G.; Wang, Z. Sugarcane nodes identification algorithm based on sum of local pixel of minimum points of vertical projection function. Comput. Electron. Agric. 2021, 182, 105994.

[8] Rui Yang., Jie Li., Qi Liu., Wenshan Huang., Kai Yin., Xi Qiao., Gradient-based method for the identification of multi-nodes in sugarcane. Information Processing in Agriculture. 2020, 12, 491-499.

[9] Li, S.; Zhang, S.; Xue, J.; Sun, H. Lightweight Target Detection for the Field Flat Jujube Based on Improved YOLOv5. Comput.Electron. Agric. 2022, 202, 107391.

[10] Guo, X.; Li, J.; Zheng, L.; Zhang, M.; Wang, M. Acquiring soybean phenotypic parameters using Re-YOLOv5 and area search algorithm. Trans. Chin. Soc. Agric. Eng. 2022, 38, 186–194.

[11] Fu, X.; Li, A.; Meng, Z.; Yin, X.; Zhang, C.; Zhang, W.; Qi, L. A Dynamic Detection Method for Phenotyping Pods in a Soybean Population Based on an Improved YOLO-v5 Network. Agronomy 2022, 12, 3209.

[12] Yan, B.; Fan, P.; Lei, X.; Liu, Z.; Yang, F. A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. Remote Sens. 2021, 13, 1619.

[13] Lawal, O.M. YOLOMuskmelon: Quest for Fruit Detection Speed and Accuracy Using Deep Learning. IEEE Access 2021, 9,15221–15227.

[14] TIAN Y, YANG G, WANG Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J].Computers and Electronics in Agriculture, 2019, 157: 417-426.

[15] 15. Wu, D.; Lv, S.; Jiang, M.; Song, H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. Comput. Electron. Agric. 2020, 178, 105742.

[16] Parvathi, S.; Selvi, S.T. Detection of maturity stages of coconuts in complex background using Faster R-CNN model. Biosyst. Eng.2021, 202, 119–132.

[17] Preeti Saini a, Bharti Nagpal b, Puneet Garg c, Sachin Kumar d. CNN-BI-LSTM-CYP: A deep learning approach for sugarcane yield prediction. Sustainable Energy Technologies and Assessments.2023, 57, 103263.

[18] Srivastava, S.; Kumar, P.; Mohd, N.; Singh, A.; Gill, F.S. A Novel Deep Learning Framework Approach for Sugarcane Disease Detection. SN Comput. Sci. 2020, 1, 1–7.

[19] Da Wang, Rui Su, Yanjie Xiong, Yuwei Wang, and Weiwei Wang. Sugarcane-Seed-Cutting System Based on Machine Vision in Pre-Seed Mode. Sensors 2022, 22, 8430.

[20] Stien Heremans, Bert Bossyns, Herman Eerens, Jos Van Orshoven,Efficient collection of training data for sub-pixel land cover classification using neural networks,International Journal of Applied Earth Observation and Geoinformation,Volume 13, Issue 4,2011,Pages 657-667,ISSN 1569-8432,https://doi.org/10.1016/j.jag.2011.03.008.

[21] Fabio Garcea, Alessio Serra, Fabrizio Lamberti, Lia Morra,Data augmentation for medical imaging: A systematic literature review,Computers in Biology and Medicine,Volume 152,2023,106391,ISSN 0010-4825,https://doi.org/10.1016/j.compbiomed.2022.106391.

[22] Alhassan Mumuni, Fuseini Mumuni,Data augmentation: A comprehensive survey of modern approaches,Array,Volume 16,2022,100258,ISSN 2590-0056,https://doi.org/10.1016/j.array.2022.100258.

[23] Shengzheng Wang, Yunfan Zhang, Tsung-Hsuan Hsieh, Wei Liu, Fei Yin, Bin Liu,Fire situation detection method for unmanned fire-fighting vessel based on coordinate attention structure-based deep learning network,Ocean Engineering,Volume 266, Part 5,2022,113208,ISSN 0029-8018,https://doi.org/10.1016/j.oceaneng.2022.113208.

[24] Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 779–788.

[25] Bochkovskiy A, Wang C Y, Liao H Ymark. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2021-06-04]. https://arxiv.org/abs/2004.10934.

[26] Wang, Z., Jin, L., Wang, S., Xu, H., 2022. Apple stem/calyx real-time recognition using YOLO-v5 algorithm for fruit automatic loading system. Postharvest Biol. Technol. 185, 111808.

[27] Zhao, J., Zhang, X., Yan, J., Qiu, X., Yao, X., Tian, Y., Zhu, Y., Cao, W., 2021. A wheat spike detection method in UAV images based on improved YOLOv5. Remote Sens. 13, 3095.

[28] Likas A, Vlassis N, Verbeek J J. The global k -means clustering algorithm. Pattern Recognition. 2003, 36(2):451-461.

[29] WOO S，PARK J，LEE J Y，et al. CBAM：convolutional block attention module［C/OL］//FERRARI V. ，HEBERT M，SMIN-CHISESCU C，et al.ECCV 2018：Proceedings of the 2018 European Conference on

Computer Vision： Lecture Notes in Computer Science. Cham： Springer， 2018： 3-19. https： //doi. org/10.1007/978-3-030-01234-2_1.

[30] Yangyang Guo, Samuel E. Aggrey, Xiao Yang, Adelumola Oladeinde, Yongliang Qiao, Lilong Chai,Detecting broiler chickens on litter floor with the YOLOv5-CBAM deep learning model,Artificial Intelligence in Agriculture,Volume 9,2023,Pages 36-45,ISSN 2589-7217,https://doi.org/10.1016/j.aiia.2023.08.002.

[31] Kai Yang, Haijun Zhang, Dongliang Zhou, Li Dong, Jianghong Ma,IASA: An IoU-aware tracker with adaptive sample assignment,Neural Networks,Volume 161,2023,Pages 267-280,ISSN 0893-6080,https://doi.org/10.1016/j.neunet.2023.01.038.

[32] ZHANG H Y，WANG Y，DAYOUB F，et al. VarifocalNet：an IoUaware dense object detector ［C］// Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Piscataway：IEEE，2021：8514-8523.

[33] Sabzi, S.; Pourdarbani, R.; Rohban, M.H.; Garcia-Mateos, G.; Paliwal, J.; Molina-Martinez, J.M. Early Detection of Excess Nitrogen Consumption in Cucumber Plants Using Hyperspectral Imaging Based on Hybrid Neural Networks and the Imperialist Competitive Algorithm. Agronomy 2021, 11, 575.

[34] Chen, C.; Wang, F.; Cai, Y.;Yi, S.; Zhang, B. An Improved YOLOv5s-Based Agaricus bisporus Detection Algorithm. Agronomy 2023,13, 1871. https://doi.org/10.3390/agronomy13071871

[35] Wang, Q.; Yang, H.; He, Q.;Yue, D.; Zhang, C.; Geng, D.Real-Time Detection System of Broken Corn Kernels Based on BCK-YOLOv7. Agronomy 2023, 13,1750.

[36] Weiyue Xu, Tao Xu, J. Alex Thomasson, Wei Chen, Raghupathy Karthikeyan, Guangzhao Tian, Yeyin Shi, Changying Ji, Qiong Su,A lightweight SSV2-YOLO based model for detection of sugarcane aphids in unstructured natural environments,Computers and Electronics in Agriculture,Volume 211,2023,107961,ISSN 0168-1699.

[37] Lijun Zhang, Jiawen Fan, Yi Qiu, Zhe Jiang, Qingsong Hu, Bowen Xing, Jingxiang Xu,Marine zoobenthos recognition algorithm based on improved lightweight YOLOv5,Ecological Informatics,Volume 80,2024,102467,ISSN 1574-9541.

[38] L. Wang, H. Qin, X. Zhou, X. Lu and F. Zhang, R-YOLO: A Robust Object Detector in Adverse Weather, Transactions on Instrumentation and Measurement, vol. 72, pp. 1-11, 2023, Art no. 5000511.