Sugarcane Bud Detection Method Based onYOLOv3-CSE Network

Hongzhen Xu (xu_hz086@163.com)
Guilin University of Technology
Jiaodi Liu
Guilin University of Technology
Jie He
Guilin University of Technology
Manlin Shen
Guilin University of Technology
Yulong Duan
Guilin University of Technology

Research Article

Keywords:

Posted Date: February 10th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1298165/v1

Sugarcane Bud Detection Method Based on YOLOv3-CSE Network

Jiaodi Liu[1], Jie He[1], Hongzhen Xu[1,2,*], Manlin Shen[1], and Yulong Duan[1]

[1]College of Mechanical and Control Engineering, Guilin University of Technology, Guilin, Guangxi, 541000, China
[2]Guangxi Mining Metallurgy and Environmental Science Experimental Center, Guangxi, 541000, China
[*]corresponding.xu_hz086@163.com

ABSTRACT

It is specified in agronomic requirements of sugarcane sowing that sugarcane buds should be placed toward the walls on both sides of the sowing ditch, while the traditional detection model of small sugarcane bud targets cannot meet the requirements of intelligent directional seeding machine for sugarcane bud identification due to such shortcomings as low accuracy, low recognition speed, and low training speed. To this end, a network model targeting sugarcane buds, called YOLOv3-CSE was proposed in this paper. On the basis of analyzing the advantages and disadvantages of the YOLOv3 network, the original YOLOv3 network was improved to achieve accurate and rapid identification of small and medium-sized targets in sugarcane buds. Besides, to further enhance the detection ability of the model for small object regions such as sugarcane buds, the original DarkNet-53 network structure and the complete intersection over union (CIoU) bounding box regression loss function were improved to make the real box regression more stable, thus avoiding IoU divergence in the training process and ameliorating the regression effect on sugarcane bud identification. Mosaic data augmentation method was applied to enrich the data diversity, so as to solve the inadequate generalization ability during small dataset training. Finally, SE-ResNet module was embedded to increase the ability of network model to identify sugarcane bud features. The test results of the YOLOv3-CSE network and the original YOLOv3 network indicated that the precision and mean average precision (mAP) of the YOLOv3-CSE network were 96.93% and 95.87%, which were 5.66% and 4.95%, respectively, higher than those of the original YOLOv3 network. Compared with other object detection models with the same dataset, the YOLO v3-CSE network proposed in this paper boasts stronger robustness, better instantaneity, higher precision and higher detection velocity in identifying small objects of sugarcane buds. In addition, it can rapidly identify the sugarcane buds, providing a technical guarantee for the application of the intelligent directional seeding machine for sugarcane seeds.

Introduction

Sugarcane, one of the main economic crops in the world, has a planting area of about 300 million mu worldwide and up to 22 million mu in China, ranking third in the globe. Sugarcane planting is one of the most important links of production,and its quality directly affects the yield of the year. It is specified in agronomic requirements that during planting, sugarcane seeds should be laid flat, with the buds facing bilateral walls of the seed ditch,and the sugarcane buds shall not be placed towards the bottom of the seed ditch, so as to promote early germination and emergence of sugarcane buds and improve the germination rate of sugarcane seeds[1]. The existing sugarcane planting machines in China and foreign countries utilize the blind planting method, lacking sugarcane bud identification function. Besides,they cannot meet the agronomic requirements for planting that the sugarcane seeds should be laid flat and the sugarcane buds should be placed toward the walls of the seed ditch[2], so the problems including late germination, slow rooting and low germination rate of the sugarcane seeds may arise,directly influencing the sugarcane yield. Currently, the directional planting of sugarcane seeds is mainly realized by identifying the sugarcane buds via human eyes, manually determining the direction of sugarcane buds and placing the sugarcane seeds, and such factors as high labor intensity and low work efficiency seriously hinder the development of the sugarcane industry. However, the detection and recognition of the sugarcane seeds at home and abroad primarily focus on the stem node identification at present[3,4], aiming to realize automatic sugarcane seed cutting, but there is no related research report on the mechanized directional planting of sugarcane seeds.

Significant breakthroughs have been made in deep learning technology for object detection in recent years[5]. There are three major categories of object detection algorithms, namely, multi-stage algorithms including typical algorithm Cascade-regions with convolutional neural network features RCNN[6], two-stage algorithms main typical algorithms: Faster RCNN[7], Mask RCNN[8], etc. and one-stage algorithms including typical algorithms YOLO Series[9–13], Retina Net[14] and SDD[15]. Among them, multi-stage and two-stage algorithms are sparse prediction models, while one-stage algorithms belong to dense prediction models. Deep learning, characterized by automatic feature extraction, can greatly improve the efficiency and precision of object detection[16], thereby promoting the wide application of object detection in agriculture. For this purpose[17]. established

a CNN classification model to distinguish good buds and bad buds on sugarcane seeds by reference to the LeNet-5 network structure[18]. simulated different light conditions through dataset expansion and identified sugarcane stem nodes using four network models. It was shown that the YOLOv4 network has the best performance in identifying sugarcane stem nodes, with a detection velocity of 69f/s and precision of 95.12%. Moreover[19]. employed a support vector machine (SVM) to detect the locations of field weeds and maize seedlings by means of K-mean clustering-based image segmentation combined with multi-feature fusion method. The results indicated that the rotation invariant LBP feature combined with GGCM can produce an average precision of up to 97.50% in identifying maize seedlings and weeds[20]. The parameters (temperature, humidity, and moisture) for the healthy growth of crop sugarcane were continuously monitored by virtue of KNN clustering and SVM classifier, and the results revealed that the accuracy of the model is as high as 96%. Furthermore, based on improved YOLOv3 network for apple fruit identification under complicated orchard environment, Zhao et al.and Li et al[21,22] optimized the model by combining the residual modules in the DarkNet-53 network with the Cross Stage Partial Network (CSPNet), introducing the Spatial Pyramid Pooling (SPP) module,substituting the Soft Non-Maximum Suppression(NMS) algorithm for traditional NMS algorithm,and finally applying the joint loss function of Focal Loss and complete intersection over union (CIoU) Loss.It was uncovered that the mean average precision (mAP),F1-score and detection velocity reach 96.3%, 91.8% and 27.8f/s, respectively. modified the YOLOv3 network to increase the real-time dynamic identification efficiency of sugarcane stem nodes, while changing the size of output feature map and reducing the number of anchors by decreasing the number of residual structures constituted by intermediate convolution layers in the YOLOv3 network. The research results demonstrated that the mAP is 90.38%, and the average time consumption for identifying the sugarcane stem nodes is 28.7ms. The team has carried out research on the intelligent directional seeding machine for sugarcane seeds that meets the agronomic requirements of directional planting of sugarcane buds. However, Li Qiang et al[23] realized the and positioning of sugarcane buds based on convolutional neural networks and an improved LeNet-5 network model. It was denoted that the recognition accuracy of the sugarcane bud position is slightly lower (only 92%),and the average detection time of a single sugarcane seed image is as long as 1.2s.Besides, this method can only identify sugarcane buds under static conditions, and cannot achieve real-time dynamic detection of the intelligent directional seeding machine for sugarcane seeds Therefore, in order to meet the agronomic requirements for directional sowing of sugarcane seeds and realize intelligent directional planting of sugarcane seeds, the rapid and dynamic detection and identification of sugarcane buds are the key technical problems to be solved first.

It is difficult for traditional identification methods to achieve precise,quick and real-time identification of sugarcane buds on sugarcane seeds because they have small sizes and different shapes, the sugarcane buds grow on the sugarcane stem nodes, there are off-white wax powder and leaf scar at the stem nodes,and dark black bulges exist at some sugarcane stem nodes. Hence, the stem node algorithms proposed in above literature cannot preferably solve the problem of detecting small objects of sugarcane buds, which has become a difficulty in rapid and accurate identification of sugarcane buds on sugarcane seeds. In this paper,a sugarcane bud detection method based on the YOLOv3-CSE network was proposed for sugarcane bud identification, and the network model was modified as follows: (1)The original DarkNet-53 network structure was improved to enhance the small object detection ability. (2)The bounding box regression loss function was improved to strengthen the regression effect on sugarcane bud identification. (3)The Mosaic data augmentation method was introduced to enrich the diversity of data. (4)The SE-ResNet module was embedded to increase the ability of network model to identify sugarcane bud features. It was verified through research that the method boasts strong robustness, good instantaneity, high precision and high detection velocity in terms of sugarcane bud identification, providing a key technology for realizing intelligent identification of sugarcane buds on sugarcane seeds as well as automatic directional planting.

Materials and methods

The project team conducted a study on the detection and identification of sugarcane buds with the intelligent directional planting machine for sugarcane seeds as a platform[24]. The sugarcane seed dataset used in this study comes from the National Agricultural Science and Technology Park in Guilin, Guangxi, China ($25°06'E, 110°31'N$), were selected as the subjects of this study, which were used for image acquisition required for model training and rapid sugarcane bud identification model training. The sugarcane seeds were collected between 2 pm and 6 pm on Sunday, October 24, 2021, the whole sugarcane was manually cut into double-bud segment cane seeds[25], and the agricultural industry standard of the People's Republic of China (Seedling of sugarcane, NY/T1796-2009)[26] was strictly implemented. The length of the double-bud segment sugarcane seeds was 30-40 cm[27] and they were placed on a conveying device of experimental table after processing and screening, followed by image acquisition (Figure 1). Which was placed at 200-250 mm above the sugarcane buds to ensure that the images obtained could reflect different photographic distance, lighting conditions, and photographic angles. The image resolution was set as 3000 pixels × 3000 pixels. A total of 2,200 images were collected, including 2,036 pieces of images labeled after being screened qualified manually. Finally, all the images were saved in JPG format. Which were divided into training set (n=1,629) and test set (n=407) at the ratio of 8:2. Part of the image collection display is shown in Figure 2. The images were manually labeled by means of the LabelImg (https ://github.com/tzutalin/labelImg). The result documents were saved in XML format, and the

stored information included the path of images, width and height, number of channels, and object box location information of sugarcane buds. Finally, the images in JPG format and the documents in XML format were sorted and renamed to guarantee the matching of images and documents, and the datasets were saved in PASCAL VOC[28] format to facilitate the training and testing of the YOLOv3-CSE network.

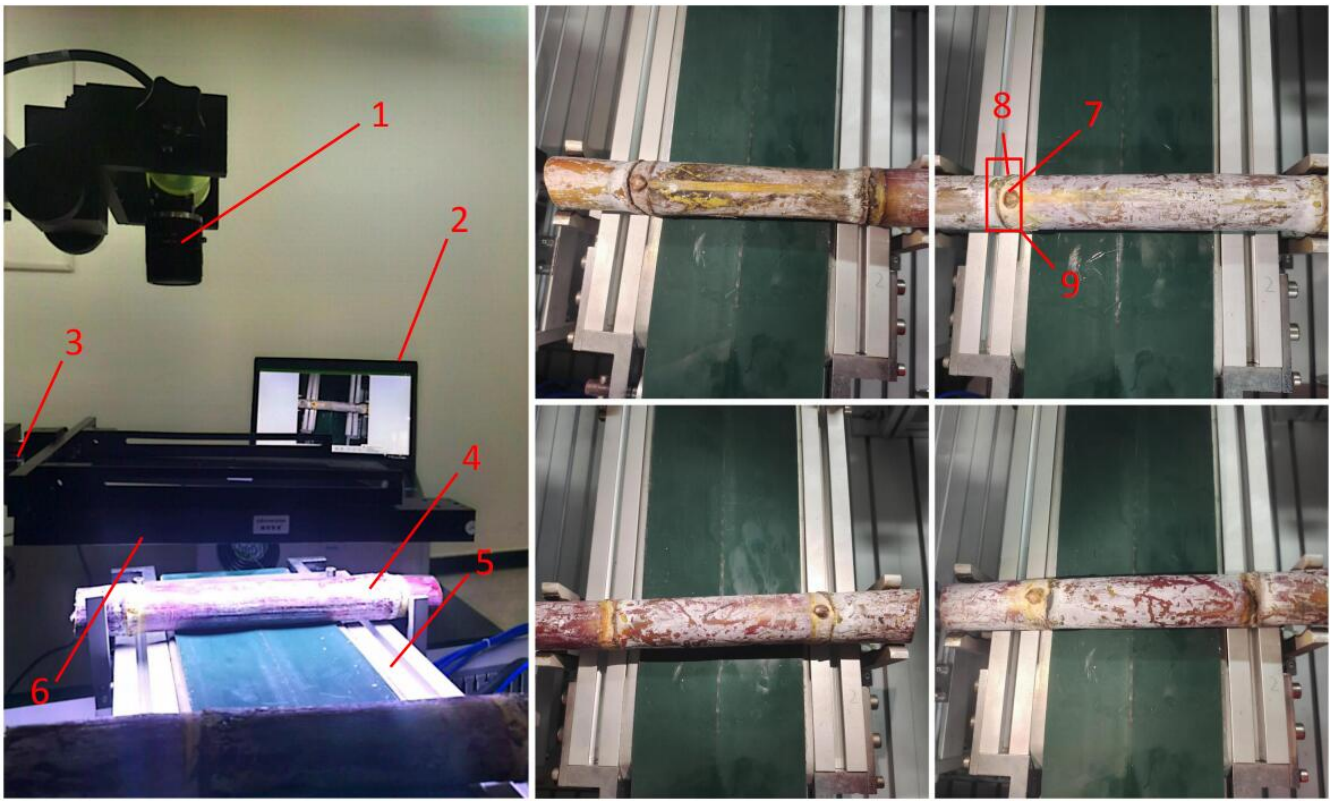

Figure 1. Image acquisition device adn display of partial image acquisition; 1.Camera, 2.Computer, 3.Bracket, 4.Cane seeds, 5.Conveying device, 6.Fill light, 7.sugarcane bud, 8.leaf scar, 9.node region.

The working principle of intelligent directional sowing of sugarcane seeds

The intelligent directional planting machine for sugarcane seeds was designed to realize the intelligent directional sowing of sugarcane seeds, and its working principle is shown in Figure 2. The machine was composed of a seed metering and conveying device, a CCD camera, a sugarcane bud direction-adjusting device, a pendulum-type retrieving and throwing mechanism and a rack. The sugarcane bud direction-adjusting device consisted of a servo motor,two-finger clamping cylinder, a lifting cylinder and other components. The working principle of the machine can be described as follows: the sugarcane seeds were transported to the designated position by the seed metering and conveying device, and then one end of the sugarcane seeds was clamped using the sugarcane bud direction-adjusting device and moved upward for a certain distance (to ensure that the sugarcane seeds were not interfered by the seed metering and conveying device during the direction adjustment process). Next, the CCD camera was triggered to capture the sugarcane seed images, the sugarcane bud position on the sugarcane seeds was identified, and the sugarcane seed steering angle was calculated. The sugarcane bud direction-adjusting device was applied to rotate the sugarcane seeds so that they were oriented at the same angle, and the sugarcane seeds were clamped by the pendulum-type retrieving and throwing mechanism for seeding and planting, so as to realize the intelligent directional sowing of the sugarcane seeds.

Automatic sugarcane planting scheme

In order to tackle the problems of intelligent identification and automatic directional planting of sugarcane buds, the technical scheme shown in Figure 3 was adopted in this study to realize the image acquisition of sugarcane seeds, rapid identification of sugarcane buds, adjustment of sugarcane bud direction and planting of sugarcane seeds. Firstly, the images of sugarcane seeds were collected using an image acquisition system, and then preprocessed and divided into datasets. Secondly, the YOLOv3 network(a typical object detection algorithm) was adopted for sugarcane bud identification,so to achieve fast and precise identification of sugarcane buds on sugarcane seeds. As for the problems of the original YOLOv3 network in sugarcane bud

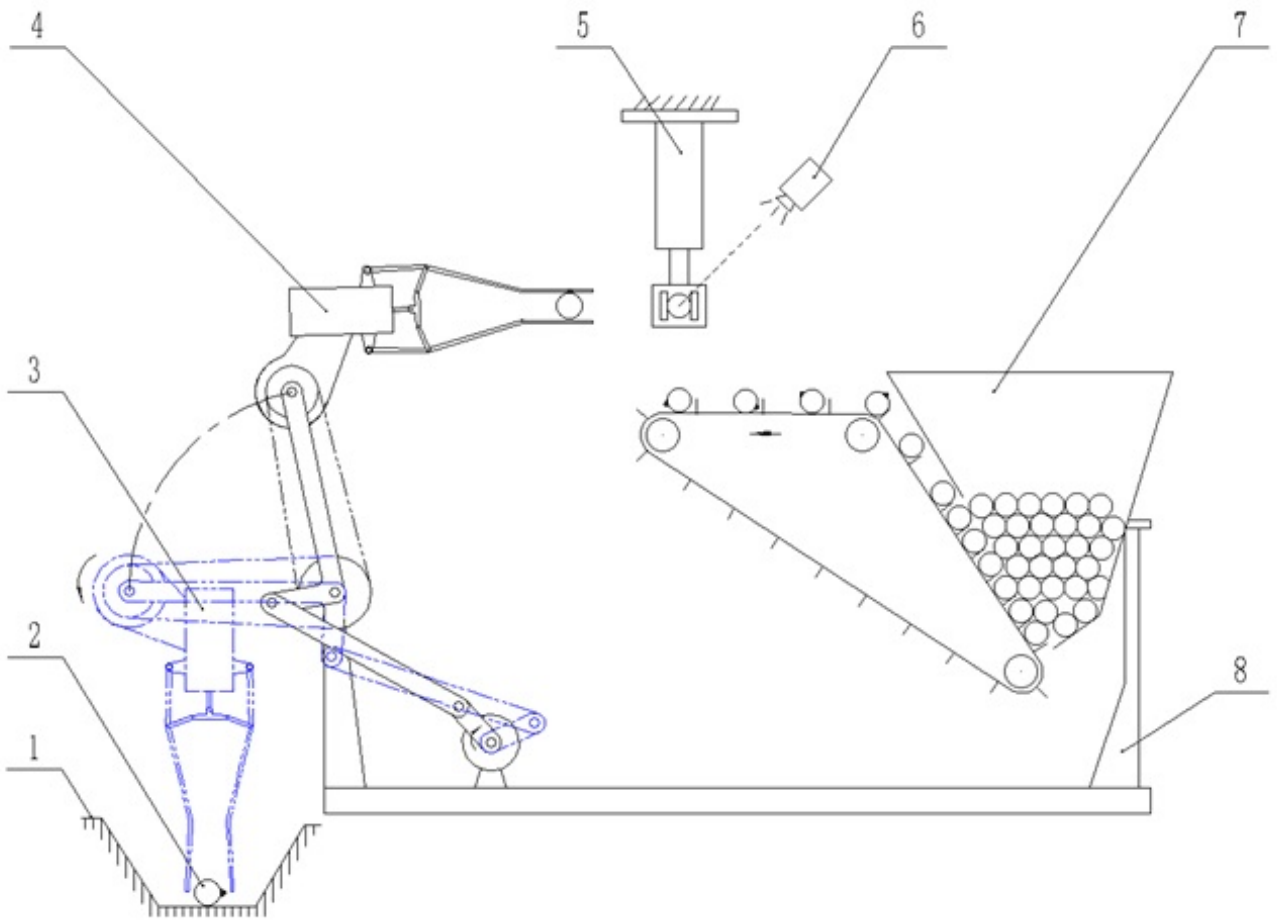

Figure 2. Schematic diagram of the working principle of intelligent directional seeding of sugarcane seeds; 1.Seed ditch, 2.Sugarcane seeds, 3.Seed placement, 4.Pendulum-type retrieving and throwing mechanism, 5.Sugarcane bud direction-adjusting device, 6.CCD camera, 7.Seed metering and conveying device, 8.Rack.

identification, such as missing detection of small objects of sugarcane buds, low identification speed and precision,and large weight files,the original model was improved,and the proposed improvement methods were compared and verified via data analysis. To transplant the network model into the embedded devices in later stage,the sugarcane seeds were captured and adjusted for the sugarcane bud direction through an end effector,thus meeting the premise of agronomic requirements for directional planting of sugarcane seeds.

Model modification

Basic YOLOv3 network

The YOLOv3 network is an improved version of YOLO network, a type of object recognition algorithm proposed by Joseph Redmon and Ali Farhadi in 2018[11], which is characterized by higher identification accuracy and speed than other networks. The framework of the YOLOv3 network consisted of two parts, namely, backbone feature extraction network and detectors (Figure 4)

(1) The backbone feature extraction network referred to a convolution layer extracted with the DarkNet-53 network as the backbone feature,which was applied to extract the features of object images. It was composed of a 3×3 convolution layer and 5 stages of residual structures.The numbers of residual structures in each stage were 1, 2, 8, 8 and 4, respectively, and each residual structure consisted of a 3×3 convolution layer and a residual block.Moreover, each residual block contained a 1×1 convolution layer, a 3×3 convolution layer and a summing operation.

(2) The detectors constituted three branches generated by the YOLOv3 network,each branch had 3 different sizes of feature maps (13×13, 26×26 and 52×52), a convolution set, a 3×3 convolution layer and a 1×1 convolution layer.Among them, the convolution set was made of continuous 1×1, 3×3, 1×1, 3×3, and 1×1 convolution layers.Through network computation,every

feature map output 3 bounding boxes, each of which predicted the center coordinates (x and y), width (w), height (h), confidence and category information.

In the YOLOv3 network, the DarkNet-53 network consisted of 53 convolution layers in total, which were mainly used to extract the object features.First of all, the size of the original images was adjusted to the input size,and the channel number of the feature model was increased using a scaled pyramid structure similar to the FPN network[29] and a 3×3 convolution kernel.Next, a 1×1 convolution kernel was utilized to reduce the channel number of the feature model. At last,3 feature maps with detection scales of 52×52,26×26 and 13×13 were obtained and then mutually fused, so that the model could detect objects of different sizes. For small objects such as sugarcane buds,however,missing detection of sugarcane buds and disability to meet the precision requirements remain problems even if the 52×52 feature map is used for output.

Basic YOLOv3 network

Improvement of feature layers of YOLOv3 network

The objects of sugarcane buds were identified based on the original YOLOv3 network. Specifically,in each prediction scale,3 bounding boxes were predicted by virtue of 3 anchors in each grid cell, so the YOLOv3 network was capable of identifying the input images with any resolution.According to Figure 5, if the center of the object fell into the grid,the grid would be responsible for predicting the object.On the basis of the original YOLOv3 network, the original network feature layers were modified in this study, and the superficial layer information could be better utilized to improve the detection ability of small sugarcane buds by increasing the scale of feature maps. As the target sugarcane buds were very small in the scene of experimental table,the output of feature maps with a scale of 13×13 or smaller could be ignored. In addition,the up-sampling of feature maps at the scale of 52×52 was combined with the output of 32 layers in this study,and the third new output scale (104×104) of feature maps was obtained through convolution operation. Meanwhile, the output of feature maps with a scale of 13×13 and 4 remaining residual units at the end of the original DarkNet-53 network were removed.

Improvement of bounding box regression loss function

IoU[30], published in 2016 and widely applied since then, is an algorithm for calculating the overlap ratio of different images, which is frequently used for object detection in the field of deep learning. The calculation method is shown in Eq. (1)

$$IoU = \frac{|M \cap N|}{|M \cup N|} \tag{1}$$

Where, M=(x,y,w,h) stands for the prediction box, and N=(x^{gt},y^{gt},w^{gt},h^{gt}) represents the ground truth box. Besides, x and y are the abscissa and ordinate, respectively, of the center of bounding box. Moreover, w and h refer to the width and height of the bounding box, respectively.

IoU is adopted in the original YOLOv3 network, which in fact is applied to measure the relative size of the overlap of two bounding boxes. In other words, a larger size of overlap between the prediction box and the ground truth box signifies a better prediction effect of the object detection network model. However, IoU has two obvious shortcomings:

(1) In the case of non-intersection between the prediction box and the ground truth box, the distance between the two boxes cannot be reflected, the loss function is not differentiable, and the non-intersection cannot be optimized. At this time, there is no gradient echo, and divergence occurs, thus disabling normal training.

(2) As for the intersection between the prediction box and the ground truth box, when the IoU value remains the same, it cannot reflect the way of intersection between the two boxes.

To solve the above two shortcomings, IoU was improved to CIoU in this study[31], and the center distance, overlap rate and scale of the prediction box and the ground truth box were considered to make the regression of the ground truth box more stable and avoid IoU divergence during training, leading to faster and more accurate training convergence. CIoU loss was defined according to Eq. (2).

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(m, n^{gt})}{C^2} + \alpha v \tag{2}$$

Where, α is the weight parameter, expressed in Eq. (3), and v is the parameter used to measure the consistency of the length-width ratio, expressed Eq. (4). Moreover, v involves the w and h to be predicted, whose partial derivatives calculated by v are shown in Eqs. (5)-(6), respectively. m and n represent the centers of the prediction box M and the ground truth box N, respectively. ρ is the Euclidean distance, and c is the diagonal length of the smallest enclosing box covering two boxes. In Figure 6, the upper left box represents the ground truth box, the lower right box stands for the prediction box, the outermost dashed box means the minimum bounding rectangle, and d refers to the Euclidean distance between the centers of the two boxes.

$$\alpha = \frac{v}{(1-IoU)+v} \tag{3}$$

$$v = \frac{4}{\pi^2}\left(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}\right)^2 \tag{4}$$

Where, w^{gt} and h^{gt} indicate the width and height of the ground truth box, respectively, and w and h are the width and height of the prediction box, respectively.

$$\frac{\partial v}{\partial w} = \frac{8}{\pi^2}\left(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}\right) \times \frac{h}{w^2+h^2} \tag{5}$$

$$\frac{\partial v}{\partial h} = \frac{8}{\pi^2}\left(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h}\right) \times \frac{w}{w^2+h^2} \tag{6}$$

Generally, the predicted values of w and h are very low, so that the value of $\frac{1}{w^2+h^2}$ is very high. As a result, gradient explosion occurs. To avoid such a situation, the value is usually set as the constant 1 in CIoU loss.

Improvement of bounding box regression loss function

In terms of image identification and deep learning, the quality of datasets also affects the robustness and accuracy of the neural network model, so data augmentation for image datasets is usually necessary for enhancing the training of neural network[32,33]. Currently, photometric transformation and geometric transformation are the commonly used data augmentation methods. The former mainly focuses on the Hue, Saturation and Value (HSV) color space of pictures, especially the random adjustment of parameters V, H and S. The latter mainly targets the random scaling, rotation, translation, image occlusion and clipping of pictures.

In this study, the Mosaic data augmentation method was introduced into the original YOLOv3 network, an improved version of the CutMix[34] data augmentation method, which are theoretically similar to some extent. The CutMix data augmentation method processed two pictures, that is, a small area of one picture was selected for masking, in which a small area of another picture was used to cover the corresponding small area of the first picture (Figure 7).

In the Mosaic data augmentation method, four pictures were processed combined with the advantages of photometric transformation and geometric transformation. The operations in the workflow were all random, which remarkably increased the environmental complexity of pictures, helped improve the diversity of datasets and enhanced the object identification ability. The fundamental principle of Mosaic data augmentation is that an intact new picture is formed by randomly selecting, clipping and splicing four pictures of the same size from the original dataset (Figure 8). To be specific, four pictures were randomly selected from the original dataset, and every picture was clipped at a random size. Then the first clipped picture was placed on the upper left corner, and the second, third and fourth pictures were placed on the lower left corner, lower right corner and upper right corner, respectively, after the same treatment. Finally, a picture was obtained after Mosaic augmentation, whose pixel size of the picture was identical to that of the original four pictures. The workflow is exhibited in Figure 9.

SE-ResNet module embedding

In images with complex distribution of environment, the traditional YOLOv3 network usually has phenomena such as identification error and missing detection of sugarcane buds due to the unbalanced confidence distribution. Hence, it is very necessary to embed the SE-ResNet module of Squeeze-and-Excitation Networks (SENet) to improve the identification accuracy and speed of the network. The SENet[35], a CNN structure of attention mechanism proposed by Jie Hu et al. from Momenta, is used to emphasize the information features and inhibit the features of non-object information by learning to use global information. There are two core operations in the SENet, including squeeze and excitation.

As shown in Figure 10[35], Ftr and Fsq represented squeeze operations, Fex indicated excitation operation, X stood for input, and U meant the result of intermediate transformation. Additionally, H, W and C were the width, height and number of layers of the network. In the squeeze operations, the feature maps of H×W×C were changed into those of 1×1×C through global pooling, so as to obtain the global feature at the channel level, meaning that the H×W pixel was compressed into 1×1 pixel, which was also realized via global average pooling. The excitation operation aimed to obtain the relationship between channels after the global features were acquired. In this operation, the bottleneck structure of two fully connected (FC) layers was employed. The

first FC layer reduced the dimension with a dimensionality reduction coefficient r (a hyperparameter) of 16, which was then activated by ReLU. The second FC layer functioned to restore the original dimension, and the activation value of each channel could be ultimately obtained by means of Sigmoid activation. In the last scale operation, the weight output by excitation operation was regarded as the importance of each feature channel following feature selection, which was then weighted to the original feature channel by channel using multiplication. It means that different channels are provided with different weights, and the darker the color is, the greater the weight will be. The whole process could be regarded as the learning of weight coefficient of each channel, so that the model was more capable of identifying the features of every channel.
Based on the idea of SENet, there are two improvement methods for the network in general: (1) directly adding SENet after the convolution layer (It is applicable to any network, but it may generate a large number of convolution layers and parameters, thus decreasing the training and learning speed, reducing the identification effect, and requiring many experiments to determine which convolution layers are added with SENet.), and (2) introducing the SE-ResNet module of SENet to replace the inception or residual layer in the original network. The SE-ResNet module is shown in Figure 11. In this method, the location of the embedded SE-ResNet module was relatively definite, and did not need to be confirmed by repeated experiments. There were multiple residual layers in the YOLOv3 network, so the network was improved by embedding the SE-ResNet module (Figure 12).

YOLOv3-CSE network

In combination with the four aforementioned improvement methods, the original YOLOv3 network was modified. Specifically, the original feature extraction network DarkNet-53 was improved to DarkNet-43 network, the number of residual modules was decreased from 5 to 4, and the numbers of residual structures in each stage were set to 1, 2, 8 and 8, respectively. In addition, the SE-ResNet module was embedded into the improved YOLOv3 network, that is, the SE-ResNet module was embedded into the last residual layer of the Residual structure at the end of each stage. In the improved YOLOv3 network, the residual structures in the last stage of the original YOLOv3 network were removed, so the original 107 layers were reduced to 94 layers. Moreover, as 3 SE-ResNet layers were embedded, the improved YOLOv3 network was composed of 97 layers instead of 107 layers in the original one. The improved YOLOv3 network was named YOLOv3-CSE network in the present study, whose structure is shown in Figure 12.

Model experiments

Environment configurations

The hardware configurations for model training and testing are as follows: Intel (R) Core (TM) i7-10700K CPU @ 3.80 GHz, 24 G DDR, with NVIDIA GeForce RTX3080 Ti Graphics Card and algorithmic program environment of Windows10 Professional, CUDA 11.0. OpenCV 4.5.3 vision library was used for morphological image processing, and PyTorch 1.90 was employed as a framework in Python 3.7 environment to implement the training, testing and application of the whole algorithm.

Scheme design

Four improvement schemes based on the YOLOv3 network were compared, and a series of indicators of the sugarcane bud identification effect in different improvement schemes were comprehensively analyzed (Table 1). The Mosaic data augmentation was introduced in Scheme 1, and the original network feature layers were improved in Scheme 2 on the basis of Solution 1. In Scheme 3, IoU was further modified into CIoU based on Scheme 2. Finally, the YOLOv3-CSE network was embedded with the SE-ResNet module on the basis of Scheme 3.

Scheme	Data augmentation	Network feature layer	Loss function	SE-ResNet module
1	✓			
2	✓	✓		
3	✓	✓	✓	
YOLOv3-CSE	✓	✓	✓	✓

Table 1. Scheme design.

Evaluation criteria

In this study, precision, recall, F1-score, and mAP were selected as the evaluation criteria to better evaluate the model designed. Precision represented the proportion of the number of positive samples of correctly identified sugarcane buds to that of all predicted positive samples [Eq. (7)]. Recall referred to the proportion of the number of positive samples of correctly identified sugarcane buds to that of all positive samples [Eq. (8)]. The mAP meant the area enclosed by the precision-recall (P-R) curve

when the IoU threshold was 0.5 [Eq. (9)]. F1-score indicated the weighted harmonic mean of precision and recall [Eq. (10)]. Based on the IoU threshold, the sample was negative when the IoU of the prediction box and the ground truth box was less than 0.5, while it was positive when the IoU exceeded 0.5.

$$P = \frac{TP}{TP+FP} \tag{7}$$

Where, TP is the number of positive samples of correctly identified sugarcane buds, and FP denotes the number of positive samples of erroneously identified sugarcane buds.

$$R = \frac{TP}{TP+FN} \tag{8}$$

Where, FN means no negative sample of sugarcane buds.

$$mAP = \frac{1}{n}\sum\int_0^1 P(R)\,dR = AP, F1 = \frac{2PR}{P+R} \tag{9}$$

Where, P(R) represents the P-R curve function, and n stands for the number of identification types. Only one identification type was involved in this study, so n=1 was adopted.

$$F1 = \frac{2PR}{P+R} \tag{10}$$

Model training

Some parameters of the YOLOv3 network could be determined through repeated testing. In order to select the optimal parameter values, the mode was tested repeatedly, and it was found that the accuracy of the model was relatively high when the learning rate was equal to 0.001. As a result, the initial learning rate was set to 0.001, which was decreased gradually with the increased number of training iterations. The final learning rate, IoU threshold, batch size, confidence and number of iterations were set as 0.0001, 0.5, 12, 0.01 and 100, respectively. The settings of model parameters are listed in Table 2. During training, small-batch training was conducted with 12 pieces of images as a batch, and the weights were updated once after the training of each batch of images. After training, 100 weights were screened to generate 10 weight files with relatively small test loss for inspection, from which the one with the highest mAP was selected as the weight file. Finally, the test set and the validation set were tested, and the test results were saved.

Parameter name	Parameter value
Batch_size	12
Training size	416×416
Number of iterations	100
IoU threshold	0.5
Initial learning rate	0.001
Final learning rate	0.0001
Confidence	0.01

Table 2. Settings of model parameters.

The change curves of loss value of five different network models for sugarcane bud training are shown in Figure 13. In the initial training stage, the sugarcane bud detection model manifested high learning efficiency and fast training convergence. As the number of iterations increased, however, the slope of the training curve was gradually decreased. When the number of training iterations reached 50, the changes in loss value tend to be stable after obvious fluctuations in convergence, and the loss value was converged slowly, uniformly and finally stably after 100 rounds of training. Additionally, the loss value rose slightly beyond 100 rounds, indicating overfitting of training set of the model. Hence, 100 rounds was determined as the termination condition of model training in comprehensive consideration of the accuracy of training the network model, so as to avoid the overfitting of the model due to excessive training times. According to the change curves of loss value of the five different network models, the YOLOv3-CSE network displayed prominently faster training convergence and milder fluctuations in convergence, as well as slightly lower loss value after training than the other four network models.

Results and analysis

Analysis of model indicators under different hyperparameters

The analysis results of the sugarcane bud identification network model with different IoU thresholds in the performance test are shown in Figure 14. Within a certain range, the IoU threshold could directly influence the precision of the model, and the recall basically remained unchanged. As the IoU threshold increased, the overlap rate between the detected prediction box and the ground truth box rose, and the number of false detections also increased. When the IoU threshold reached 0.5, the values of recall and F1 stood at 95.87% and 0.94, respectively, which means that the network model has achieved a sufficiently high identification precision, thus providing a basis for the identification and the determination of the orientation of sugarcane buds. The analysis results of the sugarcane bud identification network model with different confidence thresholds in the performance test are shown in Figure 15. As the confidence threshold increased, the value of mAP of the network model decreased. When the confidence threshold was lower than 0.6, the values of F1, recall and precision remained unchanged. On the contrary, when the confidence threshold exceeded 0.6, the values of F1 and recall began to decrease slowly, but the value of precision increased slowly. To obtain higher values of mAP and F1, the confidence threshold in this study was set at 0.01. When the values of mAP and F1 reached 95.87% and 0.94 respectively, YOLOv3-CSE network achieved the best prediction results.

Comparison of performance improvement between network models

According to Table 3, other improvement scheme showed obvious differences in identification performance compared with the original YOLOv3 network. The values of mAP and precision of the original YOLOv3 network stood at 90.92% and 91.27%, respectively. Compared with the original network model, Scheme 1 only achieved a slight improvement in identification performance. Compared with Scheme 1, the values of mAP and precision of the network model of Scheme 2 increased by 0.36% and 2.37%, respectively. In addition, the size of weight files was reduced from 240.7 MB to 112.6 MB, and the identification time of a single image was also shortened. Compared with Scheme 2, the values of mAP and precision of the network model of Scheme 3 increased by 2.39% and 1.44%, respectively, and the identification time and frame rate of a single image were also significantly improved. Compared with the original YOLOv3 network, the values of mAP and precision of the network model of Scheme 3 increased by 2.04% and 1.61%, respectively, and Scheme 3 also showed more prominent advantages in various indicators.

Model	P/%	R/%	F1/%	mAP/%	Weight/MB	Detection time/s	Frame Rate
YOLOv3	91.27	84.75	0.87	90.92	240.7	0.25	22
Method1	91.51	84.93	0.87	91.08	240.7	0.25	22
Method2	93.88	85.24	0.88	91.44	112.6	0.15	26
Method3	95.32	87.93	0.91	93.83	112.6	0.15	26
YOLOv3-CSE	96.93	90.67	0.94	95.87	118.2	0.15	28

Table 3. Comparison of identification performance between models.

Comparison with other identification network models

To further test the effectiveness of the YOLOv3-CSE network in identifying the sugarcane bud features, the training was conducted in the same dataset, and the network model in this study was compared with other network models with the same indicators. According to Table 4, different network models showed obvious differences in performance-related indicators. Specifically, CenterNet had the lowest values of mAP and precision, despite the smallest weight and the high identification velocity of a single image. With VGG16 adopted as the backbone network, Faster RCNN had a relatively lower value of mAP, the longest identification time of a single image and the largest size of weight files. With ResNet50 adopted as the backbone network, RetinaNet had higher values of mAP and precision than those of CenterNet and Faster RCNN, but lower than those of YOLOv4[12]. YOLOv4 had better performance-related indicators but relatively larger weight files. Hence, it was still slightly inferior to YOLOv3-CSE in all performance-related indicators. The sugarcane bud identification effect of all network models is shown in Figure 16. Obviously, except for YOLOv3-CSE and YOLOv4, the other three network models showed a deviation between the target location and the real location of sugarcane buds. Moreover, CenterNet also missed identification. YOLOv3-CSE achieved better performance than YOLOv4 in terms of confidence in identifying sugarcane buds.

Conclusions

A YoloV3-CSE-based sugarcane bud identification method was proposed in this study. Before the training of the network model, data augmentation was conducted to enhance the diversity of data. The inadequate generalization ability in training small

Model	mAP/%	p/%	Weight/MB	Detection time/s
CenterNet	81.35	84.07	74.3	0.37
Faster RCNN(VGG16)	82.68	85.49	370.4	0.51
RetinaNet(ResNet50)	84.04	87.96	143.5	0.43
YOLOv4	93.76	95.13	268.5	0.2
YOLOv3-CSE	95.87	96.93	118.2	0.15

Table 4. Comparison of performance between different network models.

datasets was further strengthened. Then, the feature layer of the network and the bounding-box regression loss function were improved. Finally, the SE-ResNet module was embedded to reduce the parameters and computation, increase the identification velocity, and decrease the size of network models. Great improvements were made in identification velocity and precision, and the performance and identification effect of each network model were compared. The research results are as follows:
1) The improvements reduced the parameters and computation of the network model. For YOLOv3, the size of weight files was decreased by 50.89% from 240.7 MB to 118.2 MB. In addition, the improvement method was verified according to the evaluation criteria of network model performance. The results showed that the values of mAP and precision of the network model reached 95.87% and 96.93%, respectively, after the improvement, and the identification time was 0.15 s. Compared with the original YOLOv3 network, the values of mAP and precision rose by 4.95% and 5.66%, respectively.
2) When the IoU threshold was higher than 0.5, the values of mAP, precision, recall and F1 decreased with the rise of the IoU threshold. Given the IoU threshold, the values of mAP and F1 decreased as the confidence threshold increased. The results showed that when the IoU threshold was 0.5 and the confidence threshold was 0.01, the network model achieved the best prediction results.
3) The YOLOv3-CSE network employed solved the difficulty in dynamically identifying small targets of sugarcane buds, and boasted the advantages of strong robustness, good real-time performance, high accuracy and high detection speed, thus providing a technical guarantee for the application of the intelligent directional seeding machine for sugarcane seeds.

References

1. Xinhai., W. High-yielding techniques and implementation points of sugarcane planting. *South Cina Agric.* **12**, 29–30, DOI: https://10.19415/j.cnki.1673-890x.2018.12.015 (2018).
2. Database, C. S. Good agricultural practice for sugarcane production. *China Standards Database* **NY/T**, 2254–2012 (2012).
3. Chen, H., Xu G, J., Liu X, R. & R, H. Sugarcane stem nodes based on the maximum value points of the vertical projection function. *Ciência Rural* https://doi.org/10.1590/0103-8478cr20190797 (2020).
4. Chen M, X. J. & Cheng Q, Z. Y. C. Sugarcane stem node detection based on wavelet analysis. *IEEE Access* **9**, 147933–147946, DOI: https://doi.org/10.1109/access.2021.3124555 (2021).
5. Xiong J, D., Liu S, L. & Wang X, Z. A review of plant phenotypic image recognition technology based on deep learning. *Electronics* https://doi.org/10.3390/electronics10010081 (2020).
6. Zhao W, H., Li D, F. & W, C. Pointer defect detection based on transfer learning and improved cascade-rcnn. *Sensors (Basel* https://doi.org/10.3390/s20174939 (2020).
7. Ren S, H. R. & J, S. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intel* **39**, 1137–1149, DOI: https://doi.org/10.1109/TPAMI.2016.2577031 (2017).
8. He K, G. P. & R, G. Mask r-cnn. *IEEE Trans Pattern Anal Mach Intel* **42**, 386–397, DOI: https://doi.org/10.1109/TPAMI.2018.2844175 (2020).
9. Ahmad T, Y., Yahya M, B. & Nazir S, A. Object detection through modified yolo neural network. *Scientific Programming* https://doi.org/10.3390/s20174939 (2020).
10. Redmon J, A. Yolo9000: Better, faster, stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* https://arxiv.org/pdf/1612.08242v1.pdf (2017).
11. Redmon J, A. Yolov3: An incremental improvement. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* https://pjreddie.com/media/files/papers/YOLOv3.pdf (2018).
12. Bochkovskiy, C., A. & Liao, M., H. Yolov4: Optimal speed and accuracy of object detection. *Computer Vision and Pattern Recognition (cs.CV)* https://arxiv.org/pdf/2004.10934.pdf (2020).

13. Zhou J, P., Zou A, X. & W, H. Ship target detection algorithm based on improved yolov5. *Journal of Marine Science and Engineering* https://doi.org/10.3390/jmse9080908 (2021).

14. Y., L. T., Goyal P, G. R. K. & P, D. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell* **42**, 318–327, DOI: https://doi.org/10.1109/TPAMI.2018.2858826 (2020).

15. Wei Liu, A., Dumitru E, S., Scott R, C. & Alexander C, B. Ssd: Single shot multibox detector. *Computer Vision – ECCV 2016* https://doi.org/10.1007/978-3-319-46448-0_2 (2016).

16. Redmon J, S. & Girshick R, A. You only look once: Unified, real-time object detection. *Computer Vision and Pattern Recognition (cs.CV)* https://arxiv.org/pdf/1506.02640.pdf (2016).

17. Song H, J. N. & Xia H, Y. Study of sugarcane buds classification based on convolutional neural networks. *IEEE Trans Pattern Anal Mach Intell* **27**, 581–592, DOI: https://doi.org/10.1109/TPAMI.2018.2858826 (2021).

18. Chen W, C., Li Y, S. & X, Q. Sugarcane stem node recognition in field by deep learning combining data expansion. *Applied Sciences* https://doi.org/10.3390/app11188663 (2021).

19. Chen J, J., Qiang H, B. & Xu G, Z. Sugarcane nodes identification algorithm based on sum of local pixel of minimum points of vertical projection function. *Computers and Electronics in Agriculture* https://doi.org/10.1016/j.compag.2021.105994 (2021).

20. Kumar S, S. P. & Pragya. Precision sugarcane monitoring using svm classifier. *Procedia Comput. Sci.* **122**, 881–887, DOI: https://doi.org/10.1016/j.procs.2017.11.450 (2017).

21. Zhao H, Q. W. & Yue, Y. Apple fruit recognition in complex orchard environment based on improved yolov3. *J. Agric. Eng.* **16**, 127–135, DOI: https://doi.org/10.11975/j.issn.1002-6819 (2021).

22. Li SP, L. K. & Li KN, Y. H. Increasing the real-time dynamic identification efficiency of sugarcane nodes by improved yolov3 network. *J. Agric. Eng.* **23**, 185–191, DOI: https://doi.org/10.11975/j.issn.1002-6819.2019.23.023 (2019).

23. Li Q, L. X. & ZH, N. Recognition and location of sugarcane bud based on convolutional neural network. *Agric. Mech. Res.* **07**, 27–32, DOI: https://doi.org/10.13427/j.cnki.njyi.2022.07.005 (2022).

24. He Jie, L. Q., Shen ML, Z. W. & Duan YL, W. L. A directional sugarcane planting device based on visual recognition. *Util. model patent* **CN**, 214708719U (2021).

25. Jing, L. Key points of sugarcane planting technology. *Rural Practical Technology* (2011).

26. Database, C. S. Seedling of sugarcane. *China Standards Database* **NY/T**, 1796–2009 (2009).

27. Yongjian, L. Research on key technologies of sugarcane "healthy seeds" production. *Ph.D. dissertation. Guangxi Univ.* DOI: https://doi.org/doi:10.7666/d.Y2887450 (2015).

28. Everingham M, J., E. L. Williams CKI & A, Z. The pascal visual object classes challenge: A retrospective. *Int. J. Comput. Vis.* **111**, 98–136, DOI: https://doi.org/10.1007/s11263-014-0733-5 (2015).

29. Lin T, P., Girshick R, K. & Hariharan B, S. Feature pyramid networks for object detection. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* https://arxiv.org/pdf/1612.03144.pdf (2017).

30. Yu JH, J., Wang ZY, C. & T., H. Unitbox: An advanced object detection network. *ACM MM 2016* https://doi.org/10.1145/2964284.2967274 (2016).

31. Zheng ZH, P., Liu, R., J & D, R. Distance-iou loss: Faster and better learning for bounding box regression. *AAAI 2020* https://arxiv.org/pdf/1911.08287.pdf (2020).

32. Shorten C, M. A survey on image data augmentation for deep learning. *Journal of Big Data* https://doi.org/10.1186/s40537-019-0197-0 (2019).

33. Takahashi R, T. K. Data augmentation using random image cropping and patching for deep cnns. *Ieee Transactions on Circuits Syst. for Video Technol.* **30**, 2917–2931, DOI: https://doi.org/doi:10.1109/tcsvt.2019.2935128 (2020).

34. Yun S, D., Oh SJ, S. & Choe J, Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. *ICCV 2019 (oral talk)* https://arxiv.org/pdf/1905.04899.pdf (2019).

35. Hu J, L. S. & Sun G, E. Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell* **8**, 2011–2023, DOI: https://doi.org/10.1109/TPAMI.2019.2913372 (2020).

Acknowledgements

This research work was supported by the National Natural cience Foundation of China and the Natural Science Foundation of Guangxi province Grant No. 51565048 and No. 2021JJA160046.

Author contributions

Investigation, literature analysis, methodology , writing—original draft, validation, JD.L.; funding acquisition, project administration, J.H.; supervision, HZ.X.; revising and editing,ML.S. and YL.D. All authors have read and agreed to the published version of the manuscript.

Competing interests

The author declares no competing interests.

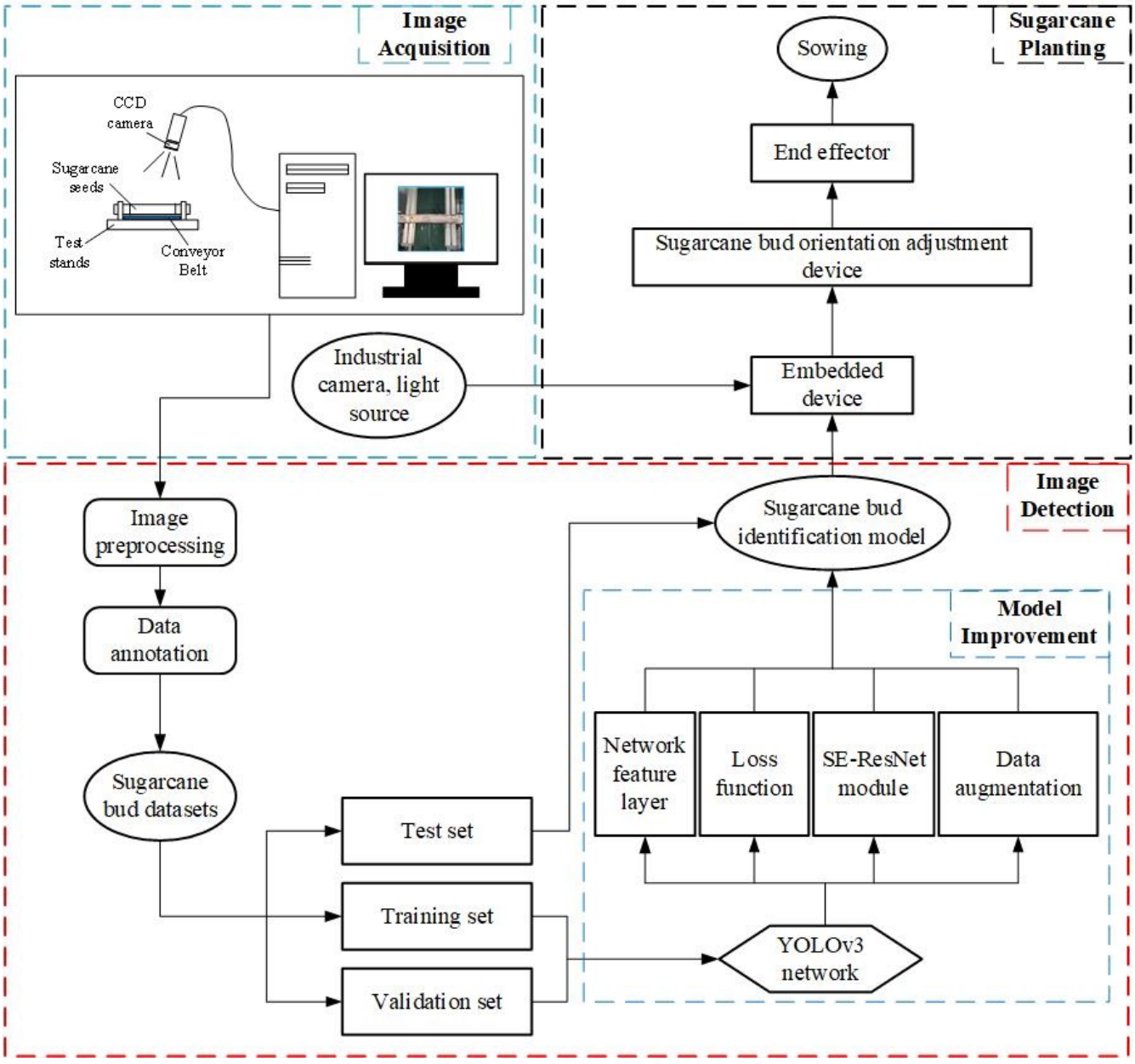

Figure 3. Automated directional planting process of sugarcane.

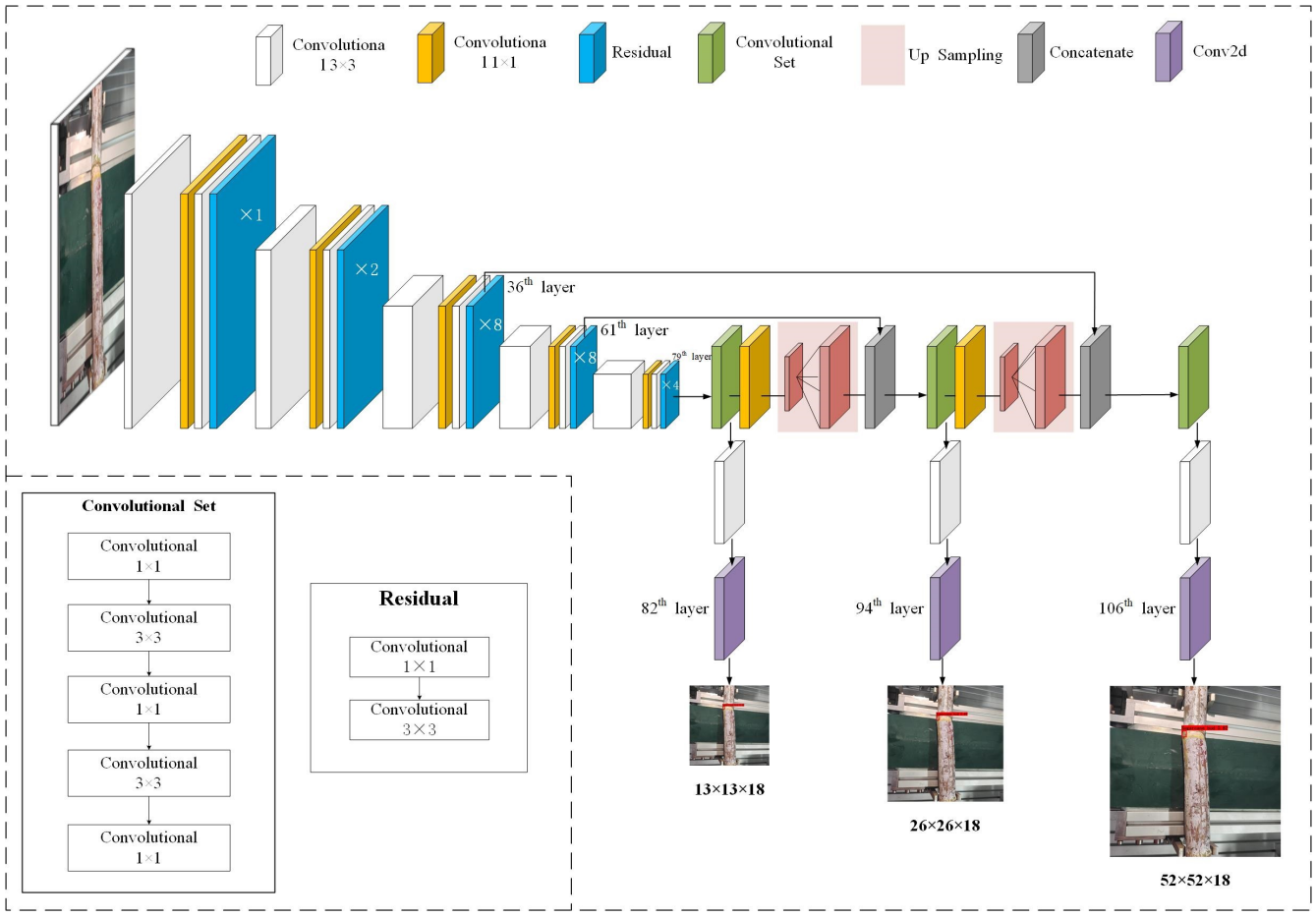

Figure 4. Framework of YOLOv3 network.

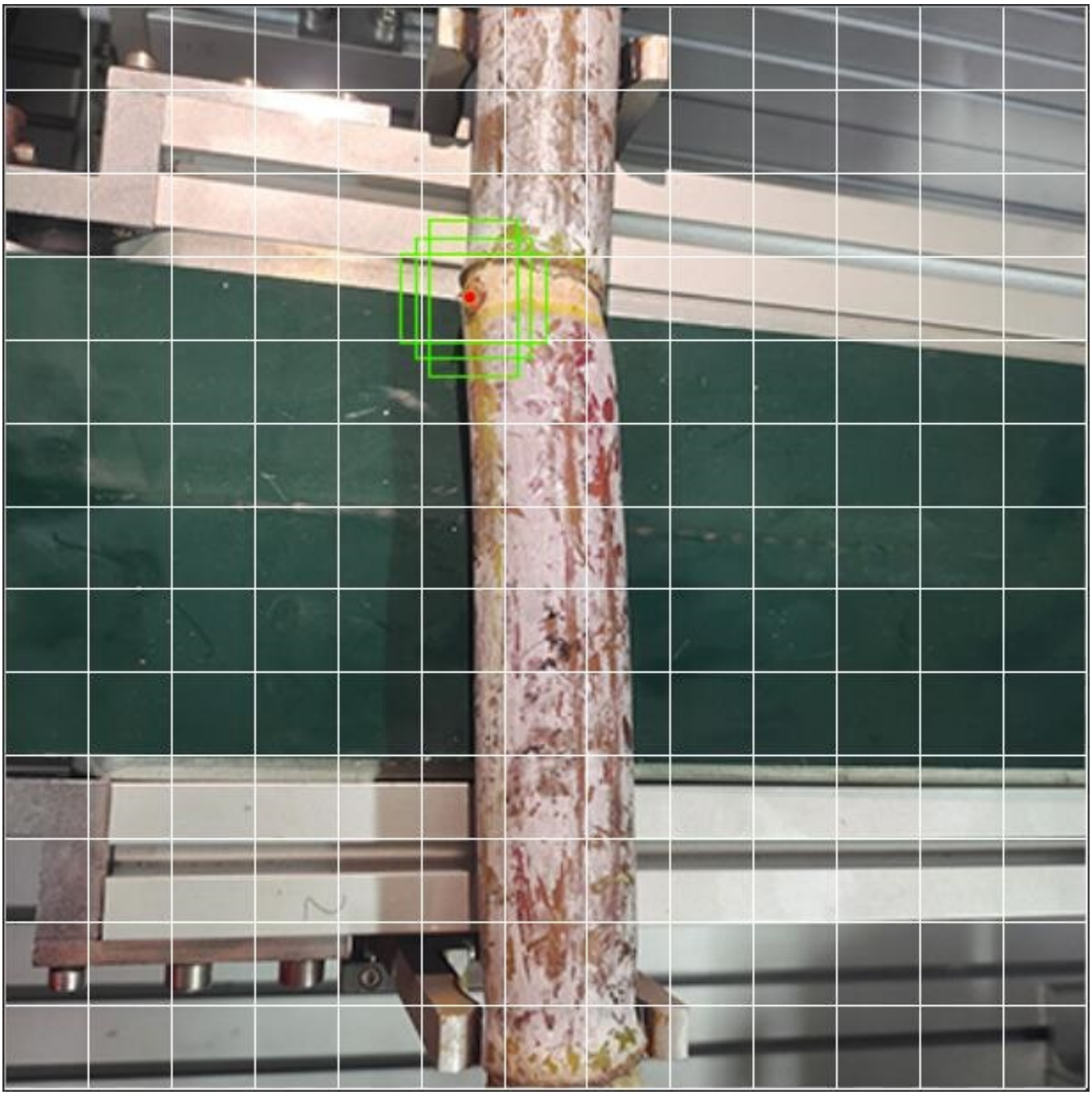

Figure 5. Prediction of bounding boxes on 13×13 grid.

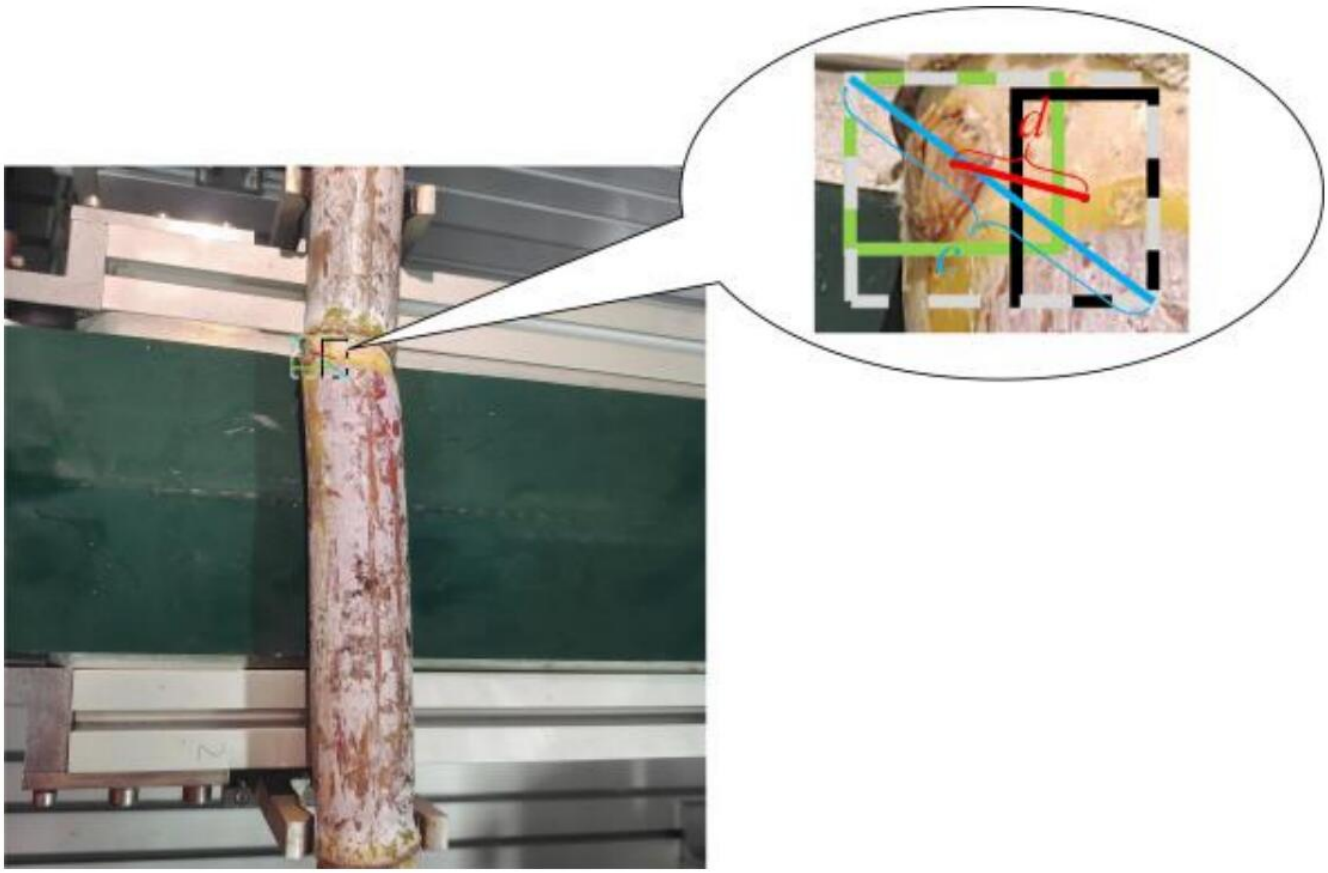

Figure 6. Normalized distance between prediction box and ground truth box.

Figure 7. CutMix workflow.

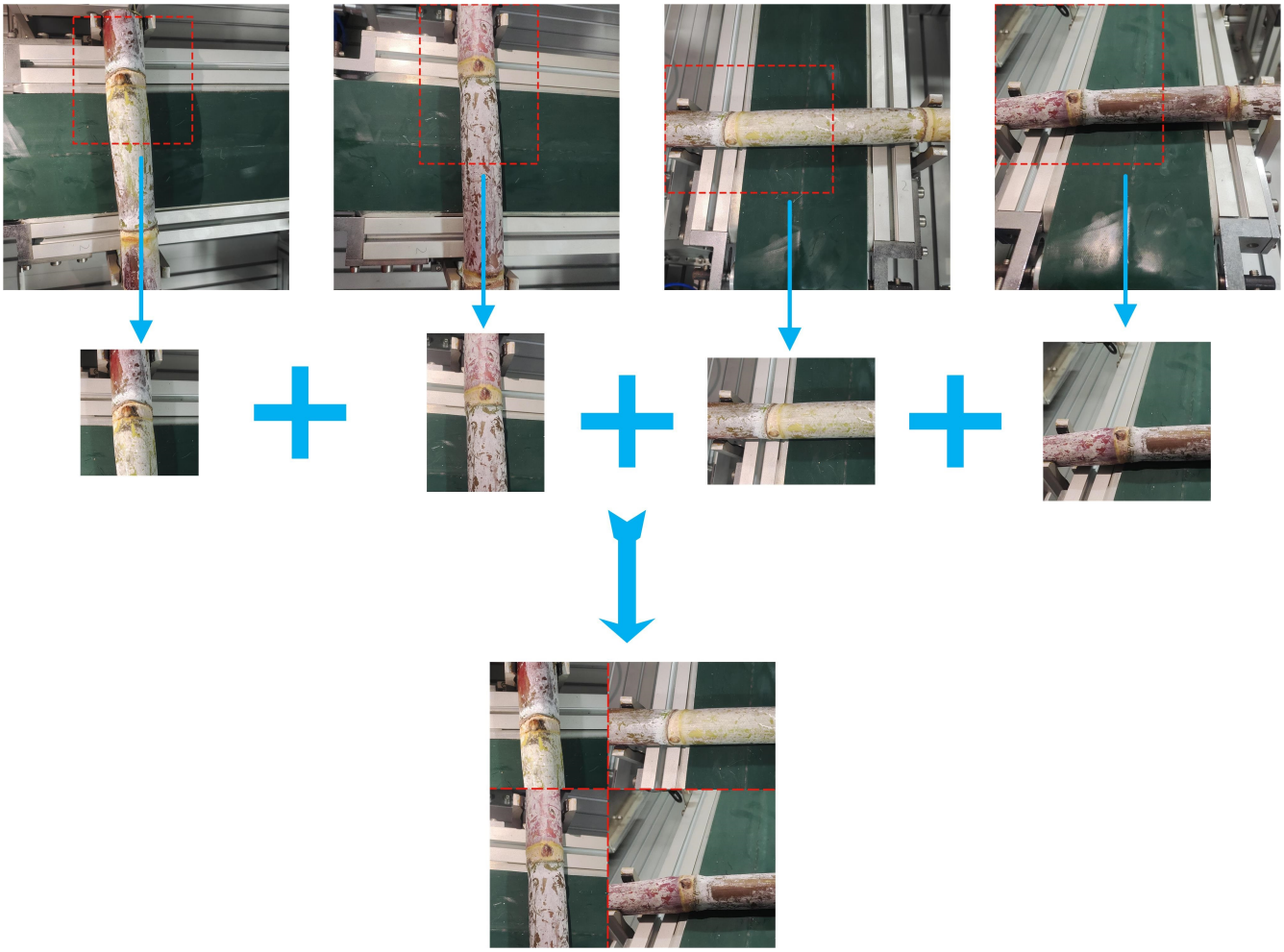

Figure 8. Mosaic data augmentation effect.

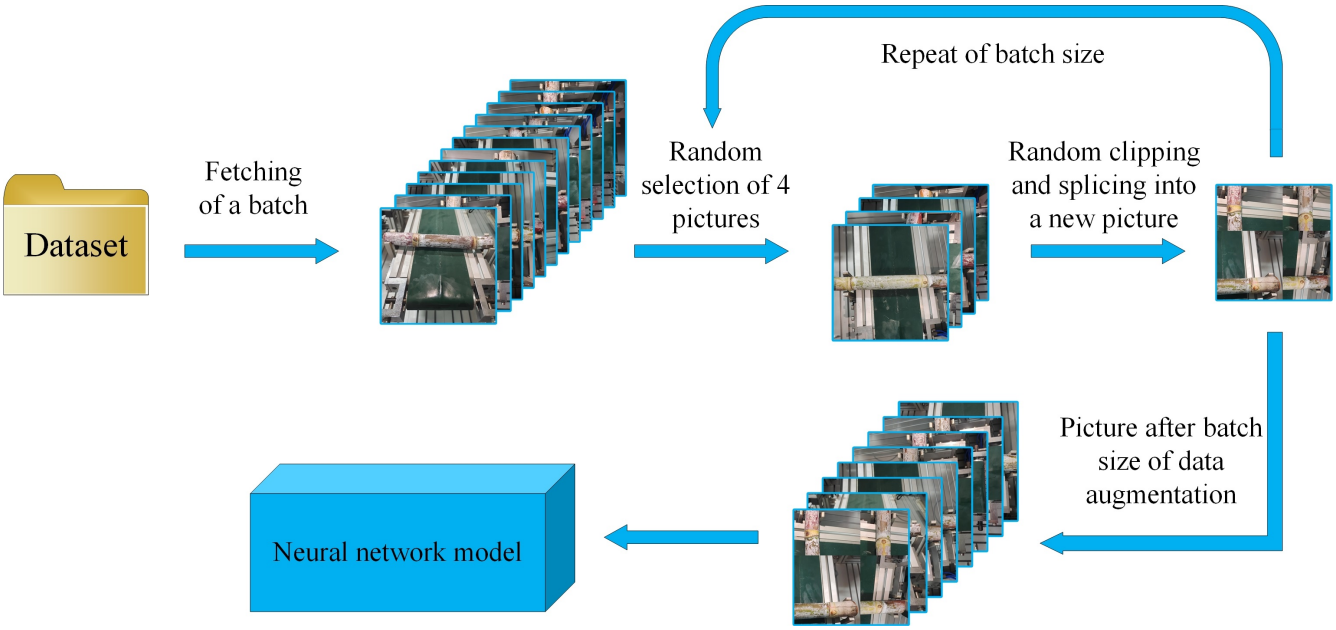

Figure 9. Mosaic workflow.

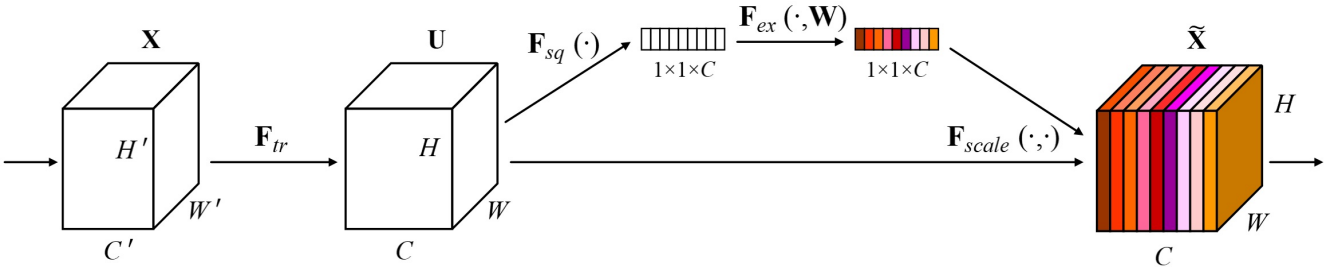

Figure 10. SENet workflow.

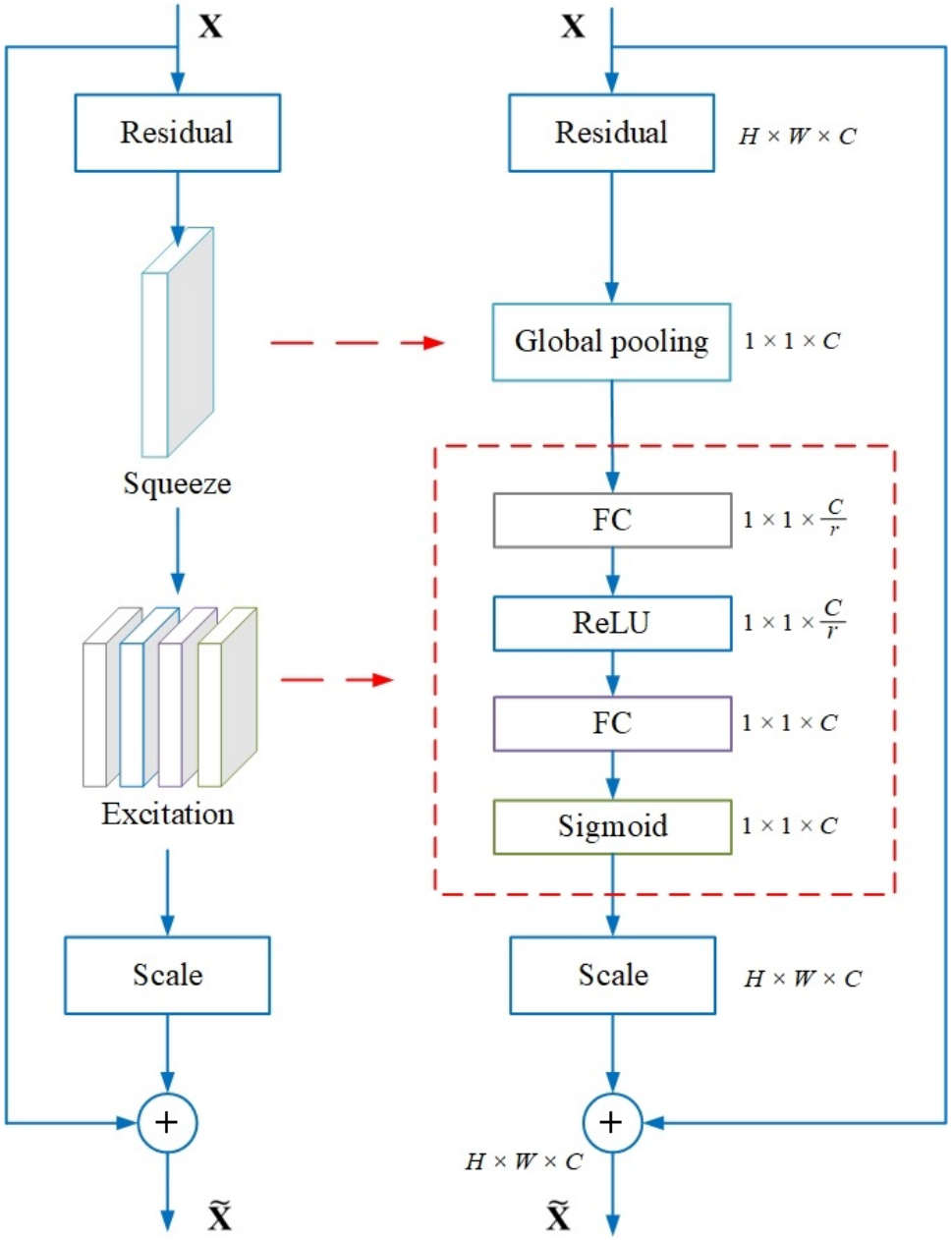

Figure 11. SE-ResNet module (left) and FC layer (right).

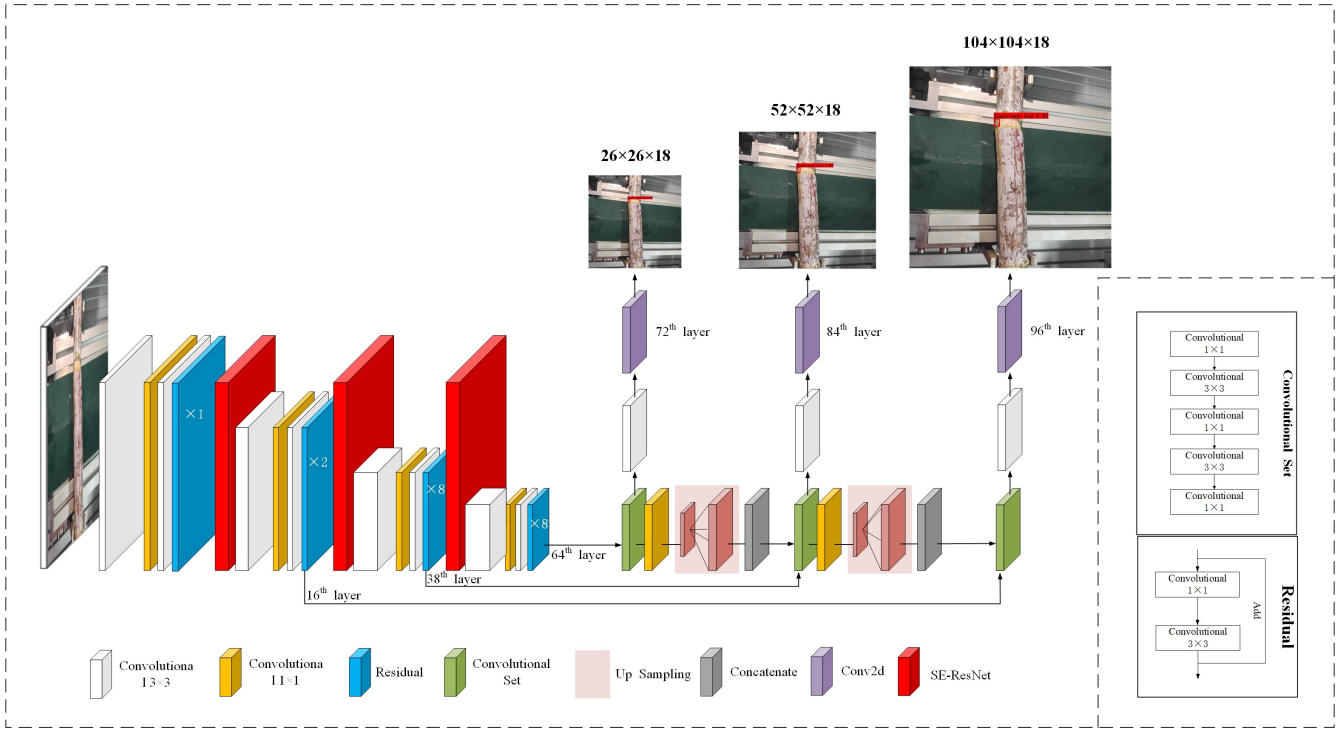

Figure 12. Structure of YOLOv3-CSE network.

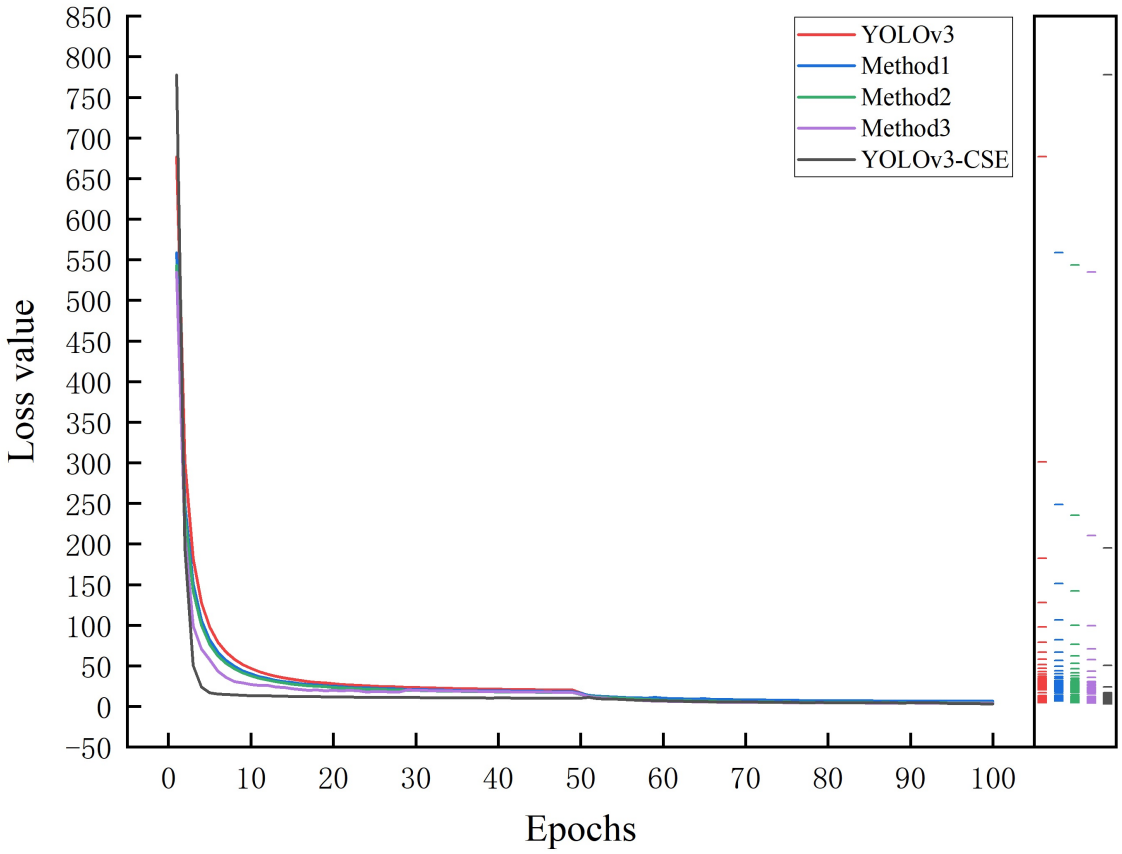

Figure 13. Change curves of training loss value of five different network models.

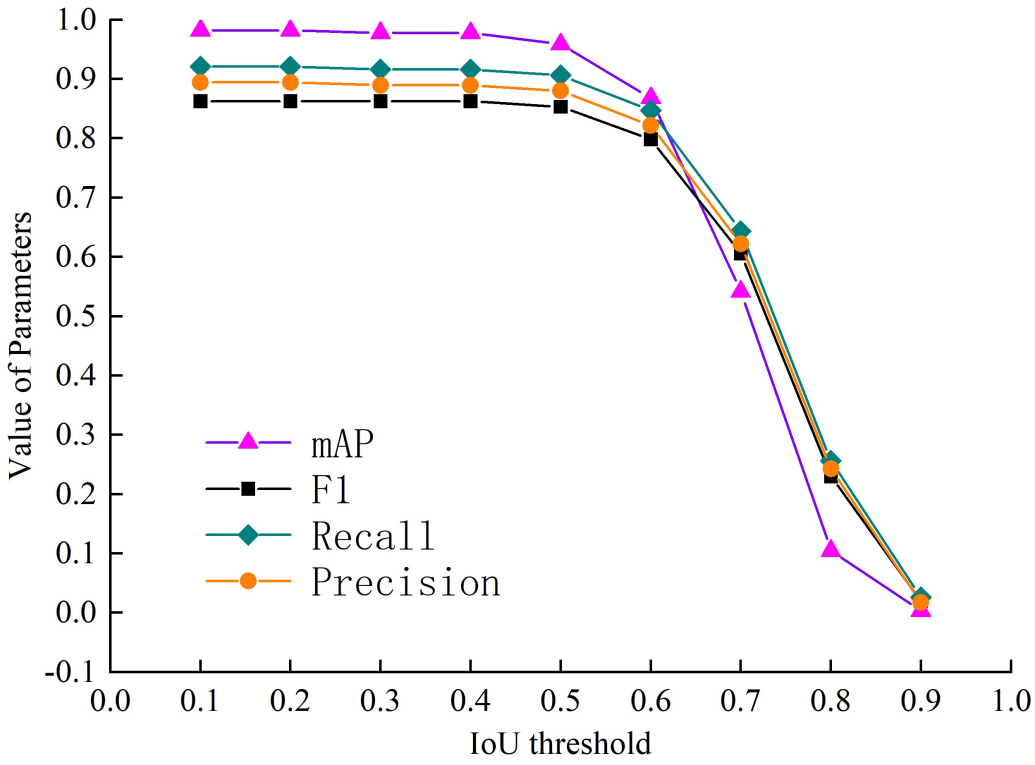

Figure 14. Changes of each indicator under different IoU thresholds.

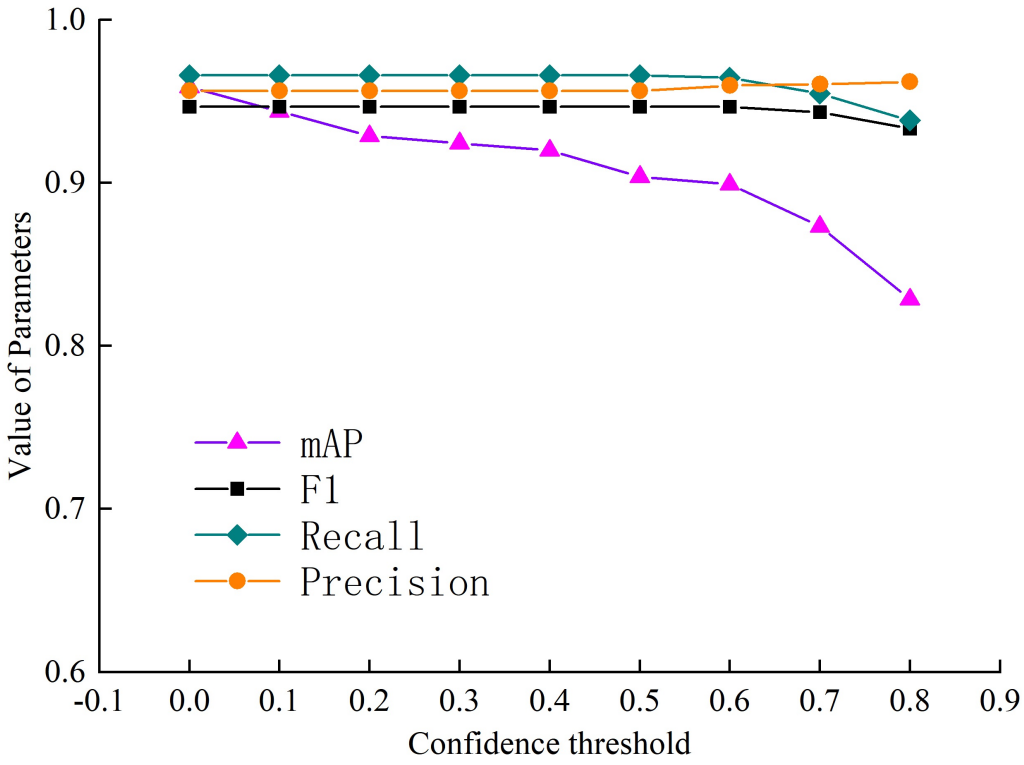

Figure 15. Changes of each indicator under different confidence thresholds.

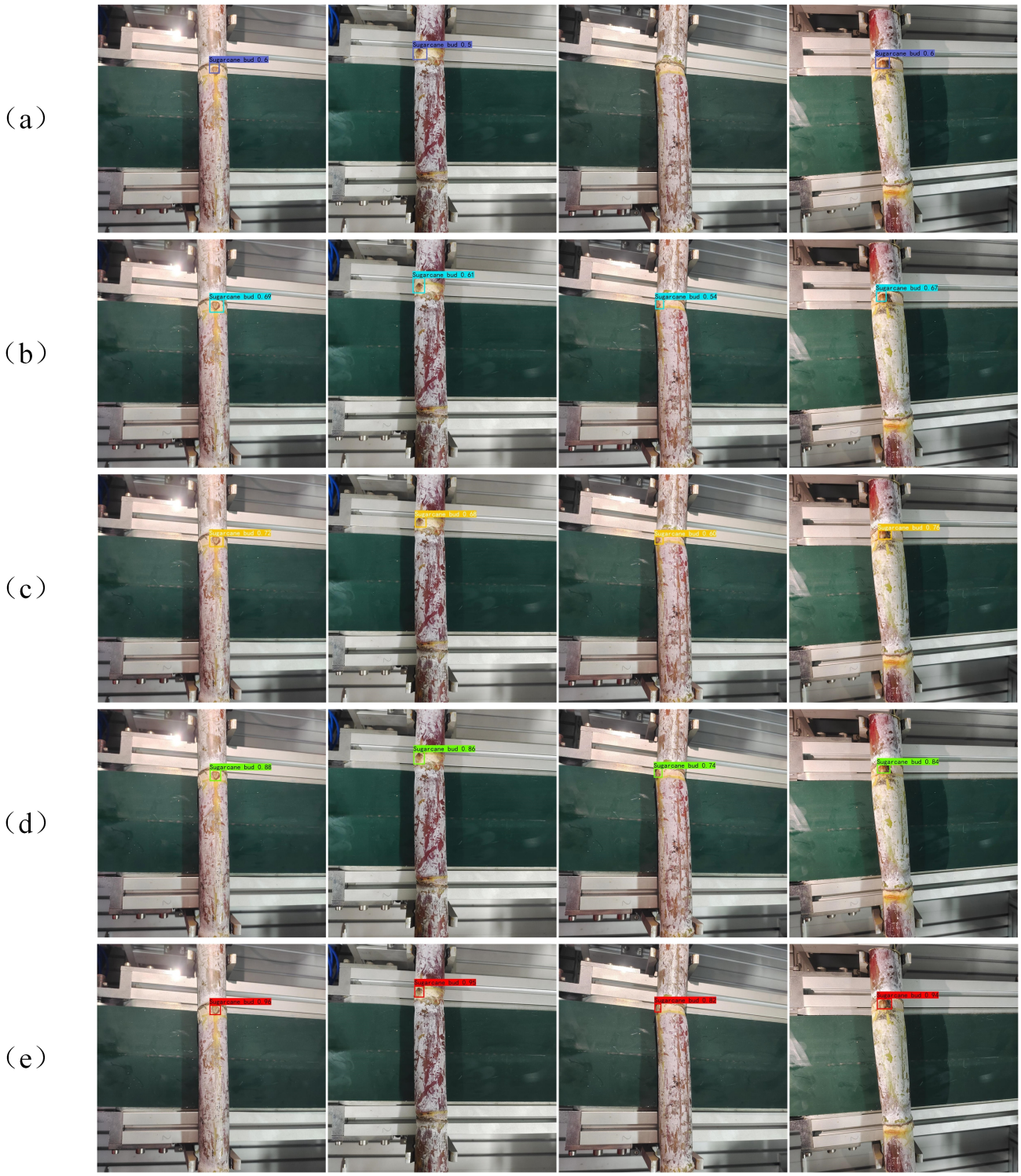

Figure 16. Sugarcane bud identification effect of different network models; (a) CenterNet; (b)Faster RCNN(VGG16); (c)RetinaNet(ResNet50); (d)YOLOv4; (e)YOLOv3-CSE.