RESEARCH ON MAIZE STEM RECOGNITION BASED ON MACHINE VISION

Minghao Liu[1], Xiangcai Zhang[1*], Xianliang Wang[1], Zhongcai Wei[1], Xiupei Cheng[1]

[1*]Corresponding author. School of Agricultural Engineering and Food Science, Shandong University of Technology/Zibo, China.
E-mail: zxcai0216@163.com | ORCID ID: https://orcid.org/0009-0003-9053-4398

ABSTRACT

Fertilization at the large bell stage of maize is the key to increasing maize yield and improving fertilizer use efficiency. To achieve fast and accurate recognition of maize stems by intelligent agricultural equipment in complex field environments, an improved YOLO v4 maize stem recognition model with an increased CBAM, which can achieve real-time identification and positioning of maize stems, is proposed. In this paper, first, we collected images of maize stems under different conditions in the field, expanded the maize stem images and produced a maize stem image dataset by adding Gaussian noise, changing the brightness and performing other data enhancement methods, and manually annotated the maize stem via LabelImg software. Second, a convolutional block attention module (CBAM) and SIoU loss function were added to the original YOLO v4 target detection network to obtain the CB-YOLO v4 target detection network. Last, this network was compared with the original YOLO v4, Faster-RCNN, SSD and YOLO v3 target detection networks, and it achieved 93.1%, 92.4% and 92.6% precision, recall and mAP (mean average precision), respectively, for maize root recognition, which is significantly better than the other algorithms and is suitable for practical maize interrow operation systems.

INTRODUCTION

As an important commodity grain in China, maize plays an important role in food, feed and other products (Yu, 2022). According to the data released by the National Bureau of Statistics, China's maize planting area increased from 41,284 thousand hectares in 2019 to 43,070 thousand hectares in 2022, and the proportion of maize in the total sown area of the country's grain increased from 35.57% to 36.40% (NBSPRC, 2022); this proportion ranked first among the four major grain crops in China. Moreover, maize is a very significant raw material for feed, industry and energy production processes (Hu et al., 2016).

With the development and application of machine vision technology, the recognition and positioning of targets through machine vision technology in the process of agricultural production has become the focus of research (de Lara et al., 2019; Maes & Steppe, 2019). Deep convolutional neural networks, which are more prevalent today, can be classified into one-stage target detection networks and two-stage target detection networks. The representative algorithms for two-stage target detection networks are the RCNN (region-based convolutional neural network) (Girshick et al., 2013) and Faster-RCNN (Ren et al., 2017). When a two-stage target detection network performs a target detection task, the first stage extracts candidate frames for the object to be detected and extracts target feature vectors from the candidate frames, and the second stage classifies and positions these target feature vectors. This class of algorithms has high recognition accuracy but slow real-time recognition. The representative algorithms of one-stage target detection networks are YOLO (Redmon et al., 2015), SSD (Single Shot MultiBox Detector) (Liu et al., 2016) and RetinaNet (Lin et al., 2020). The single-stage target detection network cancels the region candidate module and achieves target detection and positioning in a single stage. The number of parameters of this network model is greatly reduced, and the training and detection speed of the model is improved.

The YOLO algorithm is commonly employed in the field of deep learning, and the algorithm's recognition speed is faster while guaranteeing a certain recognition accuracy. Yuan et al. (Yuan & Tao, 2023) used the GCBlock structure

for the backbone network of YOLO v8, the neck end used the novel GSConv convolutional approach to reduce the computational effort, and the improved YOLO v8 accurately completed the detection and identification of fish in the monitoring data of commercial fishing vessels with lower computational effort. Based on the YOLO _ X target detection network, Wang Binbin et al. (Wang et al., 2022) used transfer learning to identify maize male ears and concluded that the detection effect varied for different maize male ear types and planting densities. Jiajun Liu et al. (Liu et al., 2023) used the YOLO v5 target detection network to detect and identify pests and diseases on maize leaves in natural environments and achieved an mAP of 71.5%, which is a good experimental result. Zhang Fan et al. (Zhang et al., 2023) improved the YOLO v3 target detection network to achieve 95% recognition accuracy on the maize leaf stomata dataset, which can automatically complete the recognition, counting and measurement of maize leaf stomata, resolving the inefficiency of the crosstalk stomata analysis method. Khan et al. (Khan et al., 2023) collected datasets of three maize crop diseases and applied YOLO v3-tiny, YOLO v4, YOLO v5s, YOLO v7s, and YOLO v8n target detection networks to detect these diseases; YOLO v4 achieved an accuracy of 97.5%, and the high-precision model was embedded into a mobile application, which can perform real-time maize disease detection in seconds.

With the continuous development of precision agriculture technology, precision fertilizer is gradually being widely utilized in agricultural production and has become an effective means of controlling the excessive use of chemical fertilizers (Quebrajo et al., 2015). In the case of maize, it is possible to locate and apply fertilizer by identifying and locating the stem of each maize plant in real time and then to carry out fertilizer application operations on the maize at the designated location, thereby reducing the amount of fertilizer applied and increasing fertilizer use efficiency. Therefore, the original YOLO v4 target detection network is improved, and the CBAM is added. Simultaneously, the loss function improves from the CIoU loss to the SIoU loss. Using the stem of maize as the research object, recognition and detection are carried out under natural conditions to provide target positioning for maize precision topdressing equipment.

MATERIAL AND METHODS

Image Acquisition

The field maize stem images in this paper were collected from 16 August to 23 August 2023 in a maize experimental field at the experimental base of the Research Institute of Agricultural Sciences, Zibo City, Shandong Province, which is 200 m long and 100 m wide, with maize planted in rows spaced 50 cm apart and plants spaced 10 cm apart. The maize was in the large flare period (V12), and the maize variety "Zhengdan 958" has long been used as a benchmark for the evaluation of regional maize variety trials (Tong, 2020). Maize images were acquired using Android smartphones with an image resolution of 2250 * 4000. The images were saved in *.jpg format, and most of the images focused on the maize stem. Images of maize stems were acquired under different weather conditions, such as sunny and cloudy days, with different light angles (downlight, backlight and photometry), as shown in Figure 1. A total of 450 images were collected, and 385 images were selected as training samples after filtering and removing the blurred images.

FIGURE 1. Some of the collected images.

Data sample preprocessing and data enhancement

In maize stem images under natural conditions, the color of the collected maize stem images was different under different weather conditions and different light angles. To obtain accurate data, the filtered maize stem image was manually labeled with the dataset through the labeling software LabelImg, as shown in Figure 2.

FIGURE 2. Dataset labeling.

Because the number of homemade dataset images was small, the training needs could not be met well. The maize stem dataset was expanded by data enhancement; 0.1-0.2 random Gaussian noise was added to the image, which randomly changed the contrast of the image (-5% ~ +5%) and randomly changed the brightness (-15% ~ +15%). To reduce the workload entailed by manual data labeling, the same position transformation method was used to label the target object in the initial data through the software to directly generate extended data containing labeling information. The number of images in the final dataset was 1540, with 80% used as the training set and 20% used as the test set.

Theoretical foundation

YOLO v4 target detection network

YOLO is a one-stage target detection algorithm that was pioneered by Redmon et al. in 2015 and is one of the most widely utilized target detection network models, and several versions have been developed. YOLO v4 is an upgraded version of YOLO v3, and YOLO v4 is an upgraded version of YOLO v3; YOLO v4 improves the detection accuracy and accelerates the detection speed.

The backbone feature extraction network of YOLO v4 (Bochkovskiy et al., 2020) uses CSPDarknet53, which is obtained by improving Darknet53, the backbone feature extraction network used in YOLO v3. CSPDarknet53 divides the feature mapping of the base layer into two parts, which are merged through the cross-stage hierarchy, reducing the amount of computation while ensuring the recognition accuracy. The network structure of the YOLO v4 target detection network mainly consists of four parts: the input, backbone, neck and head parts. Input is the input of the maize rootstock image. The input maize rootstock image is preprocessed via mosaic data enhancement, the cmBN strategy, SAT self-adversarial training and adaptive image scaling. The original maize stem image is uniformly scaled to a standard size. The backbone structure consists of a CBM module and a CSP module, the CSP module can effectively increase the depth of the network, and the feature extraction ability is enhanced. The neck structure adopts the SPP module, and the FPN+PAN structure. The SPP module is able to directly pool feature maps of arbitrary sizes with fixed sizes and obtain a fixed number of features; the FPN+PAN structure can convey both strong semantic features from the top-down and strong position features from the bottom-up to achieve bidirectional feature fusion for maize stems. In the head structure, the same multiscale idea as YOLO v3 is used for prediction, generating three feature layers with different scales, predicting their features and outputting the detection results. The structure of each module and the whole network is shown in Figure 3.
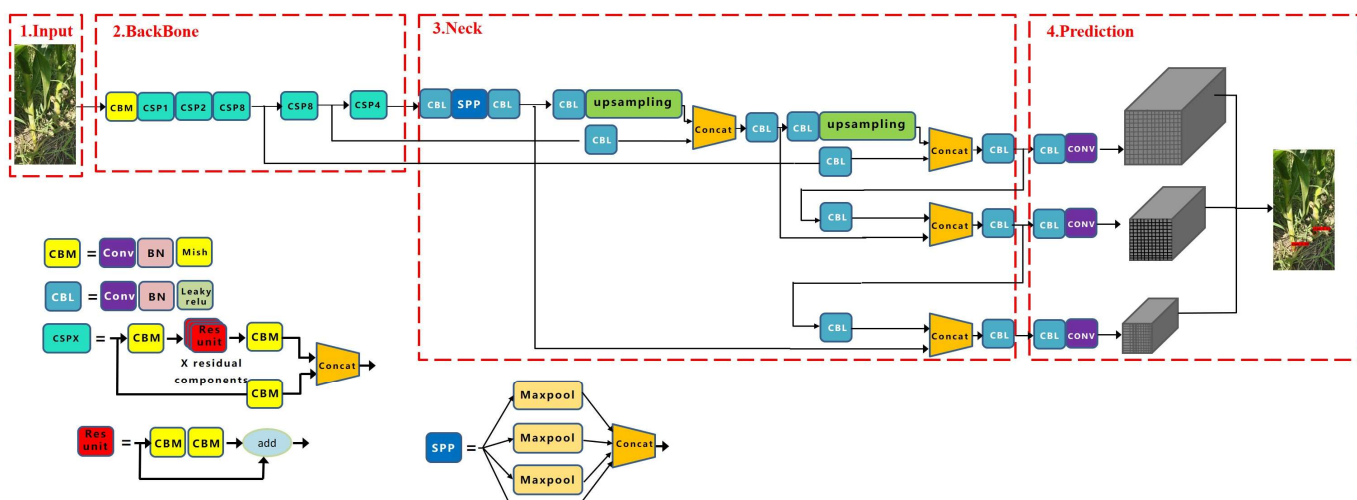

FIGURE 3. YOLO v4 network for maize stem recognition.

Attention mechanisms module

The attention mechanism in artificial neural networks is a resource allocation scheme to solve the problem of information overload; that is, in the case of limited hardware computing power, more computing resources are allocated to more important tasks. In the process of neural network learning, the greater the number of parameters of the model is, the greater the amount of information stored in the model, and the stronger the expression ability of the model; however, this leads to information overload. By adding the attention mechanism module to the network, we can pay attention to the more important information of the current task in the complex information and reduce the attention to other information to improve the accuracy and efficiency of the processing task. Therefore, attention mechanisms are widely employed in various fields of deep learning (Prinzmetal et al., 2010). Common attention mechanisms are divided into channel attention mechanisms and spatial attention mechanisms. Typical attention mechanism modules include squeeze-and-excitation networks (SENet) (Hu et al., 2020), efficient attention (ECA) (Wang et al., 2020) and spatial attention mechanism (SAM) (Zhu et al., 2019). The above attention mechanism modules are single attention mechanisms, and in practical tasks, channel attention mechanisms and spatial attention mechanisms are often combined.

In this paper, research on maize stem recognition showed that the similarity between the weed background in maize fields and the green maize stems is high, and information on channel characteristics and spatial characteristics is needed. The CBAM (Woo et al., 2018) combines the channel attention mechanism with the spatial attention mechanism. Given a spatial coordinate map as input, the CBAM infers the attention map along two channels and two dimensions of space and then multiplies the attention map and the input feature map for adaptive feature optimization. The CBAM is added before the SPP module of the neck structure of the YOLO v4 target detection network, which can be well integrated with the YOLO v4 target recognition network. The extracted features are more abundant and comprehensive and are more suitable for natural conditions in maize fields. The CBAM structure is shown in Figure 4.
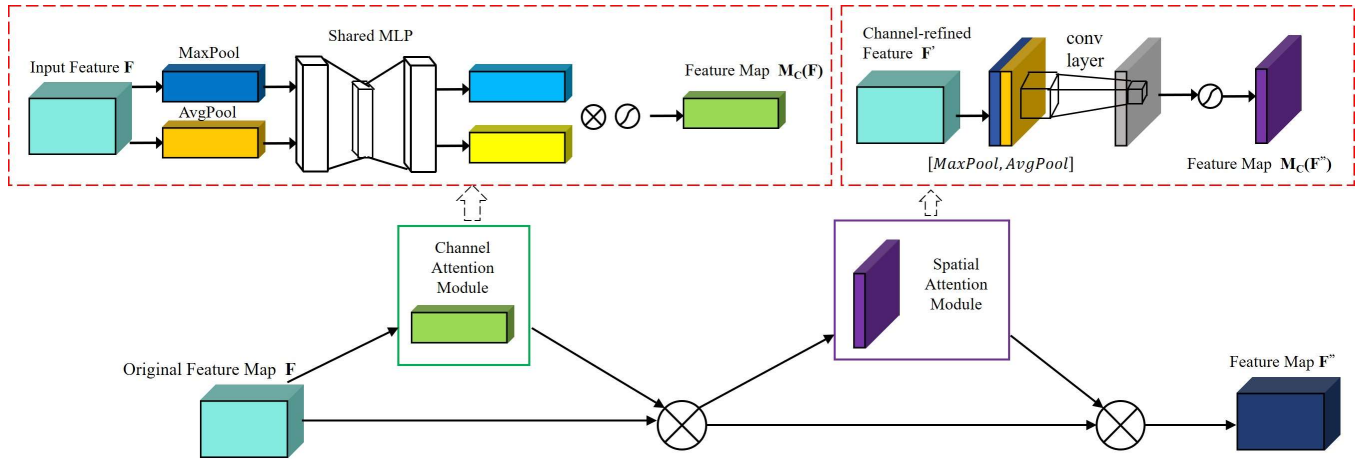

FIGURE 4. CBAM structure diagram.

The channel attention mechanism module of the CBAM can effectively detect the features of the target contour and obtain more target detection, which is calculated as follows (Woo et al., 2018):

$$M_C(F) = \sigma(W_1(W_0(F_{avg}^C)) + W_1(W_0(F_{max}^C))) \tag{1}$$

Where:

$M_C(F)$ is the channel attention output weight;

σ is the activation function;

F_{avg}^C is the spatial feature mapping after average pooling;

F_{max}^C is the spatial feature mapping after maximum pooling;

W_0 is the weight matrix of the first fully connected layer;

W_1 is the weight matrix of the second fully connected layer.

The spatial attention mechanism module of CBAM can effectively locate and detect the position of the target and improve the accuracy of target detection. The calculation method is described as follows (Woo et al., 2018):

$$M_S(F) = \sigma(f^{7 \times 7}([F_{avg}^S; F_{max}^S])) \tag{2}$$

Where:

$M_S(F)$ is the spatial attention output weight;

$f^{7 \times 7}$ is a 7×7 convolution operation filter;

F_{avg}^S is feature mapping after average pooling on the channel;

F_{max}^S is maximum pooling postfeature mapping on the channel.

In general, the input feature map F is multiplied by the channel attention mechanism module, the obtained feature result is multiplied by the spatial attention mechanism module, and the feature map is obtained after CBAM processing. The calculation method is presented as follows (Woo et al., 2018):

$$F^{'} = M_C(F) \times F \tag{3}$$

$$F^{''} = M_S(F) \times F^{'} \tag{4}$$

Where:

F is the input feature map;

$F^{'}$ is the feature map obtained by channel attention weighting;

$F^{''}$ is the feature map obtained by spatial attention weighting.

Loss function improvement

The loss function is a key indicator for measuring the quality of the network model. Most of the current target detection loss functions rely on the aggregation of bounding box regression indicators, such as the distance between the prediction box and the real box, the overlap area and the aspect ratio. The commonly employed loss functions are the GIoU (Rezatofighi et al., 2019), CIoU (Zheng et al., 2020) and EIoU (Zhang et al., 2021). The original YOLO v4 model uses CIou loss, which takes into account the distance, overlap rate and scale between the prediction box and the real box, making the target box regression more stable and increasing the consistency between the prediction box and the real box.

However, this approach does not consider the matching problem between the real box and the prediction box direction. Therefore, this paper introduces the SIoU loss (Zhora, 2022), which considers the angle of the vector between expectations and redefines the angle penalty metric, which is beneficial for improving the convergence speed and accuracy of target detection network training. Its calculation formula is (Zhora, 2022):

$$\Lambda = 1 - 2 \times \sin^2(\arcsin(\frac{c_h}{\sigma}) - \frac{\pi}{4}) \tag{5}$$

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho_t}) \tag{6}$$

$$\Omega = \sum_{t=w,h} (1 - e^{-w_t})^{\theta} \tag{7}$$

$$SIoU = IoU - \frac{\Delta + \Omega}{2} \tag{8}$$

$$L_{SIoU} = 1 - SIoU \tag{9}$$

$$\rho_x = (\frac{b_{c_x}^{gt} - b_{c_x}}{c_w})^2 \tag{10}$$

$$\rho_y = (\frac{b_{c_y}^{gt} - b_{c_y}}{c_h})^2 \tag{11}$$

$$\lambda = 2 - \Lambda \tag{12}$$

$$w_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})} \tag{13}$$

$$w_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \tag{14}$$

Where:

Λ is the angle loss;

Δ is the distance loss;

Ω is the shape loss;

IoU is the intersection and union ratio of the real box and the prediction box;

L_{SIoU} is the SIoU loss value;

c_h is the height difference between the center point of the real box and the prediction box;

σ is the distance between the center point of the real box and the prediction box;

$(b_{c_x}^{gt}, b_{cy}^{gt})$ are the center coordinates of the real box;

(b_{c_x}, b_{c_y}) are the center coordinates of the prediction box;

c_w, c_h are the width and height, respectively, of the minimum circumscribed rectangle of the real box and the prediction box;

w, h are the width and height, respectively, of the prediction box;

w^{gt}, h^{gt} is the width and height, respectively, of the real box.

Target detection network results and analysis

Test platform

The test software environment is a deep learning framework that uses Windows 10, Python 3.11.4, PyTorch 2.0.1 and Cuda 11.7. The test hardware environment (CPU) is an Intel I7-11700K, and the graphics card is an NVIDIA RTX A4000 24G. The maximum number of iterations is set to 1500, the batch size is set to 8, the Adam optimizer is selected, and the maximum learning rate is set to 0.001. Using the cosine annealing learning rate adjustment algorithm, the learning rate can be dynamically adjusted to achieve local convergence as soon as possible. The relevant model parameters are shown in Table 1.

TABLE 1. Test-related model parameters.

Parameter	Value/Type
Image size/(pixels×pixels)	320×640
Epoch	1500
Batch_size	8
Optimizer	Adam
Maximum learning rate	0.001
Modalities for adjusting learning rates	Cosine annealing

Model evaluation indicators

The effectiveness of the model is evaluated from a quantitative perspective. The selected indicators are the precision rate P, recall rate R and mean average precision (mAP). Its calculation formula is:

$$P = \frac{TP}{TP + FP} \times 100\% \qquad (15)$$

Where:

TP is the number of true positives, and

FP is the number of false positives.

$$R = \frac{TP}{TP + FN} \times 100\% \qquad (16)$$

In this case, FN is the number of false negatives;

$$AP = \int_0^1 P(R)dR \qquad (17)$$

$$mAP = \frac{\sum AP}{N} \qquad (18)$$

The AP is the average accuracy of a kind.

The size of the model is also an important evaluation index. A smaller model is conducive to the deployment of a later model. This paper evaluates the size of computer memory occupied by the model.

Ablation test

To verify the superiority of the YOLO v4 target detection network (CB-YOLO v4) with the CBAM module and improved loss function, an ablation experiment is carried out. By adding the CBAM and the SIoU loss function to the original YOLO v4 model, the effectiveness of each improvement point is verified. The ablation test results are shown in Table 2, where ' √ ' indicates that the improved method is used and ' - ' indicates that the improved method is not used.

TABLE 2. YOLO v4 ablation test results.

CBAM	SIoU	Precision ratio P/%	Recall ratio R/%	Average precision mean Map %
—	—	89.6	88.4	88.9
√	—	92.4	91.8	92.1
—	√	90.3	89.3	89.8
√	√	93.1	92.4	92.6

Table 2 shows that when the CBAM is added separately, the maize stem recognition accuracy increases by 2.8%, the recall rate increases by 3.4%, and the average accuracy increases by 3.2%. By separately improving the SIoU loss function, the accuracy of the model is increased by 0.7%, the recall rate is increased by 0.9%, and the average accuracy is increased by 0.9%. When the CBAM and the improved SIoU loss function are added to the YOLO v4 target detection network, the accuracy of the model increases by 3.5%, the recall increases by 4%, and the average accuracy increases by 3.7%. A comprehensive evaluation revealed that increasing the CBAM and improving the SIoU loss function are effective at improving the recognition accuracy of the model and that increasing the CBAM is more effective at optimizing the model.

Comparison Test with Other Algorithms

To further demonstrate the advantages of the proposed target detection network compared with other target recognition algorithms, SSD, Faster-RCNN and YOLO v3 are selected for performance comparison experiments. In the test, the four algorithms use the same dataset and training platform. Table 3 shows the detection indicators of the four detection models.

TABLE 3. Detection indices of the four models.

Detection model	Precision ratio P/%	Recall ratio R/%	Average precision mean Map %	Memory size /MB
CB-YOLO v4	93.1	92.4	92.6	250.26
SSD	80.3	81.8	80.7	42.82
Faster-RCNN	87.6	86.1	86.8	252.36
YOLO v3	84.3	82.6	83.1	240.69

The experimental results show that the accuracy, recall and average accuracy of the CB-YOLO v4 target detection network are 93.1%, 92.4% and 92.6%, respectively, which are significantly greater than those of other target detection networks. Compared with the SSD, which is also a single-stage object detection network, the accuracy, recall and average accuracy of CB-YOLO v4 are increased by 12.8%, 9.6% and 11.9%, respectively, but the memory consumption of the model is increased. Compared with Faster-RCNN, the accuracy, recall and average accuracy of CB-YOLO v4 are increased by 5.5%, 6.3% and 5.8%, respectively, and the memory consumption of the model is reduced by 2.1 MB. Compared with YOLO v3, the accuracy, recall and average accuracy of CB-YOLO v4 are increased by

8.8%, 9.8% and 9.5%, respectively, and the memory footprint of the model is increased by 9.57%. Although Faster-RCNN also has better detection accuracy, Faster-RCNN does not meet the task requirements of field maize stem recognition because the detection speed of the two-stage target detection network is significantly slower than that of the single-stage target detection network. Although SSD and YOLO v3 consume less memory than CB-YOLO v4, the other performance indicators are far lower than those of CB-YOLO v4. In summary, the CB-YOLO v4 target detection network has more advantages in practical applications.

To better evaluate the performance of the target detection network, the results of the four target detection networks are visualized; the results are shown in Figure 5.
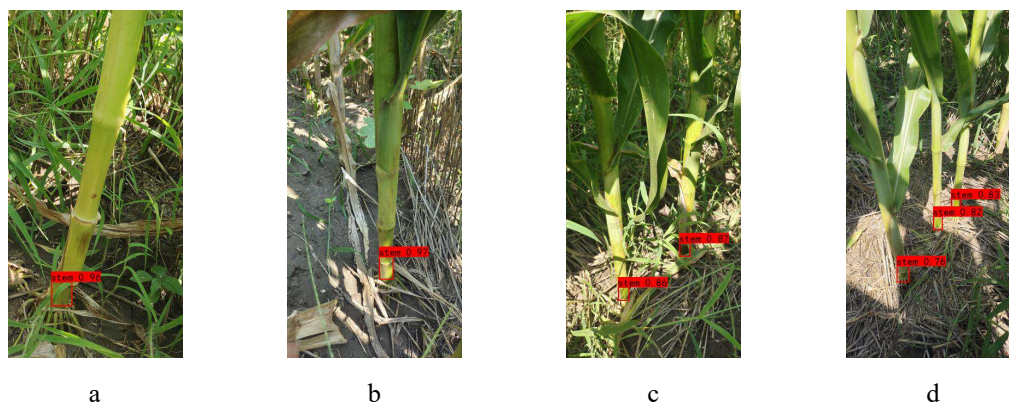
YOLO v3

a   b   c   d
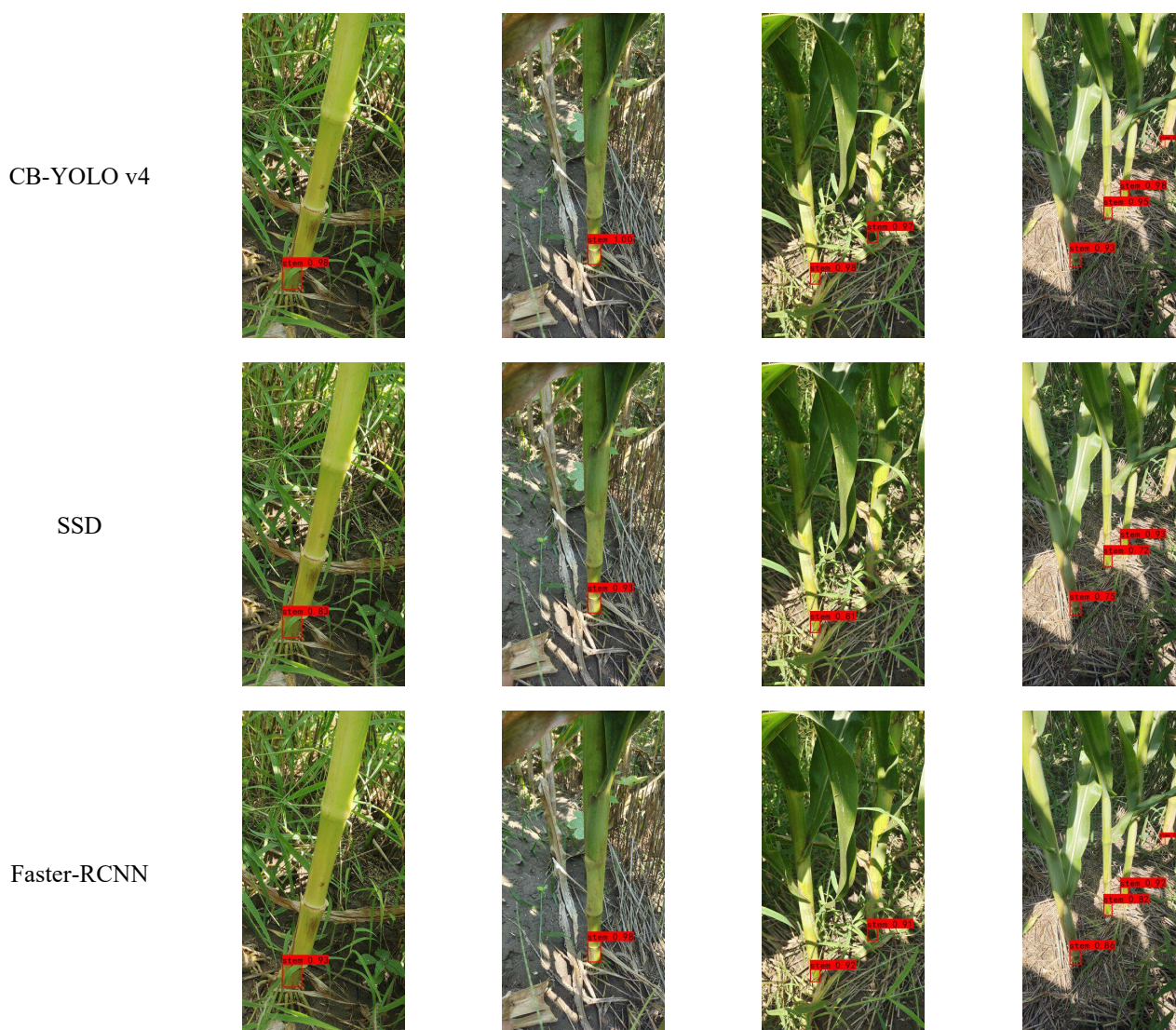
FIGURE 5. Comparison of the results of each target detection network.

Figure 5a and Figure 5b show the test results under normal light without weed interference. The four models can correctly identify maize stems, and CB-YOLO v4 has the highest credibility. Figure 5c and Figure 5d show that the CB-YOLO v4 and Faster-RCNN target detection networks can correctly identify maize stems under strong light conditions and weed interference. The SSD and YOLO v3 target detection networks can identify some maize stem targets, but some targets will be missed. CB-YOLO v4 has the highest credibility.

CONCLUSIONS

In this paper, an improved YOLO v4 target detection network (CB-YOLO v4) is proposed for maize stem recognition.

(1) Image acquisition was carried out in the maize planting area under natural conditions. The collected images included different weather conditions and different light conditions. The original image was transformed by adding Gaussian noise and changing the brightness and contrast of the image. The enhanced dataset included 1540 images. Through data enhancement, the application scenarios of maize stem recognition were effectively expanded, and the generalization ability of the model was improved.

(2) The CBAM is added to make the extracted features more comprehensive, and the CIoU loss function is replaced by the SIoU loss function, which improves the speed and accuracy of the target detection network training. The accuracy of the CB-YOLO v4 target detection network for maize stem recognition was 93.1%, the recall rate was 92.4%, and the average accuracy was 92.6%. Compared with those of the original YOLO v4, SSD, Faster-RCNN and YOLO v3 target detection networks, the performance indicators are better, and the proposed model is more suitable for identifying maize stems in maize fields.

FUNDING

REFERENCES

Bochkovskiy A, Wang CY, Mark L (2020) YOLOv4: Optimal speed and accuracy of object detection. ArXiv. https://doi.org/10.48550/arXiv.2004.10934

Girshick R, Donahue J, Darrell T, Malik J (2013) Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition 580-587. https://doi.org/10.48550/arXiv.1311.2524

Hu H, Zhang YF, Chen WZ, Zhao HB (2016) The development status and prospect of corn topdressing machinery in China. Corn Science 24(3): 147-152. https://doi.org/10.13597/j.cnki.maize.science.20160323

Hu J, Shen L, Sun G (2020) Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 42(8): 2011-2023. https://doi.org/10.1109/TPAMI.2019.2913372

Khan F, Zafar N, Tahir MN, Aqib M, Waheed H, Haroon Z (2023) A mobile-based systemfor maize plant leaf disease detection and classification using deep learning. Frontiers Plant Science 14: 1079366. https://doi.org/10.3389/fpls.2023.1079366

Lara A de, Longchamps L, Khosla R (2019) Soil water content and high-resolution imagery for precision irrigation: Maize Yield. Agronomy 9(4): 174. https://doi.org/10.3390/agronomy9040174

Lin TY, Priya G, Ross G, He KM, Piotr D (2020) Focal loss for dense object detection. 2017 IEEE International Conference on Computer Vision 42(2): 318-327. https://doi.org/10.48550/arXiv.1708.02002

Liu JJ, Su RQ, Wu Q, Xu JR (2023) Detection and identification of maize leaf diseases and insect pests based on YOLO model. Chemical Engineering and Equipment (06): 31-34. https://doi.org/10.19566/j.cnki.cn35-1285/tq.2023.06.026

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Bergm AC (2016) SSD: singleshot multibox detector. European Conference on Computer Vision. https://doi.org/10.1007/978-3-319-46448-0_2

Maes WH, Steppe K (2019) Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. Trends in Plant Science (24): 152-164. https://doi.org/10.1016/j.tplants.2018.11.007

NBSPRC - National Bureau of Statistics of China (2022) China statistical yearbook. Beijing. https://www.stats.gov.cn/

Prinzmetal W, Ha R, Khani A (2010) The mechanisms of involuntary attention. Journal of Experimental Psychology: Human Perception and Performance 36(2): 255–267. https://doi.org/10.1037/a0017600

Quebrajo L, Pérez-Ruiz M, Rodriguez-Lizana A, Agüera J (2015) An approach to precise nitrogen management using hand-held crop sensor measurements and winter wheat yield mapping in a mediterranean environment. Sensors 15(3): 5504-5517. https://doi.org/10.3390/s150305504

Redmon J, Divvala SK, Girshick RB, Farhadi A (2015) You only look once: unified, real-time object detection. IEEE Conference on Computer Vision and Pattern Recognition 779-788. https://doi.org/10.48550/arXiv.1506.02640

Ren SQ, He KM, Ross G, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence 39(6): 1137-1149. https://doi.org/10.48550/arXiv.1506.01497

Rezatofighi SH, Tsoi N, Gwak JY, Sadeghian A, Reid ID, Savarese S (2019) Generalized intersection over union: a metric and a loss for bounding box regression. IEEE/CVF Conference on Computer Vision and Pattern Recognition 658-666. https://doi.org/10.1109/CVPR.2019.00075

Tong PY (2020) Zhengdan 958 still plays a leading role in the market in 20 years. Seed Technology 38(21): 1-2. https://doi.org/10.3969/j.issn.1005-2690.2020.21.002

Wang BB, Yang GJ, Yang H, Gu JA, Zhao D, Xu SZ, Xu B (2022) Maize tassel detection based on YOLO_X and transfer learning. Journal of Agricultural Engineering 38(15): 53-62. https://doi.org/10.11975/j.issn.1002-6819.2022.15.006

Wang QL, Wu BG, Zhu PF, Li PH, Zuo WM, Hu QH (2020) ECA-Net: Efficient channel attention for deep convolutional neural networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition 11531-11539. https://doi.org/10.1109/CVPR42600.2020.01155

Woo S, Park J, Lee J, Kweon IS (2018) CBAM: convolutional block attention module. Computer Vision – ECCV 2018. https://doi.org/10.1007/978-3-030-01234-2_1

Yu XY (2022) Analysis of the current situation and driving factors of high-quality development of China's corn industry. Modern Marketing (late issue) (09): 53-55. https://doi.org/10.19932/j.cnki.22-1256/F.2022.09.05

Yuan HC, Tao L (2023) Fish detection and recognition based on improved Yolov8 Commercial fishing vessel electronic monitoring data. Journal of Dalian Ocean University 38(03):533-542. https://doi.org/10.16535/j.cnki.dlhyxb.2022-354

Zhang F, Guo SY Ren FT, Zhang XH, Li JP (2023) Automatic identification and measurement method of maize leaf stomata based on improved YOLO v3. Journal of Agricultural Machinery 54(2): 216 - 222. https://doi.org/10.6041/j.issn.1000-1298.2023.02.021

Zhang YF, Ren WQ, Zhang Z, Jia Z, Wang L, Tan TN (2021) Focal and efficient IOU Loss for accurate bounding box regression. Neurocomputing 506:146-157. https://doi.org/10.1016/j.neucom.2022.07.042

Zheng ZH, Wang P, Liu W, Li JZ, Ye RG, Ren DW (2020) Distance-IoU Loss: Faster andBetter Learning for Bounding Box Regression. Proceedings of the AAAI Conference on Artificial Intelligence 34(07):12993-13000. https://doi.org/10.1609/aaai.v34i07.6999

Zhora G (2022) SIoU Loss: more powerful learning for bounding box regression. ArXiv. https://doi.org/10.48550/arXiv.2205.12740

Zhu XZ, Cheng DZ, Zhang Z, Lin S, Dai JF (2019) An empirical study of spatial attention mechanisms in deep networks. IEEE/CVF International Conference on Computer Vision 6687-6696. https://doi.org/10.1109/ICCV.2019.00679