ITC 1/55
Information Technology and Control
Vol. 55 / No. 1/ 2026
pp. 86-99
DOI 10.5755/j01.itc.55.1.42841

Tobacco Plant Counting Based on Improved YOLOv8 and UAV Remote Sensing Images

Received 2025/09/20 | Accepted after revision 2025/12/23

HOW TO CITE: Yang, C., Huang, F. (2026). Tobacco Plant Counting Based on Improved YOLOv8 and UAV Remote Sensing Images. *Information Technology and Control*, 55(1), 86-99. https://doi.org/10.5755/j01.itc.55.1.42841

Tobacco Plant Counting Based on Improved YOLOv8 and UAV Remote Sensing Images

Changping Yang

College of Artificial Intelligence, Yango University, Fuzhou 350015, Fujian, China; e-mails: 1997changping@sina.cn (C.Y.); fhhuang@ygu.edu.cn (F.H.)

Fenghua Huang*

College of Artificial Intelligence, Yango University, Fuzhou 350015, Fujian, China; e-mails: 1997changping@sina.cn (C.Y.); fhhuang@ygu.edu.cn (F.H.)

Fujian Key Laboratory of spatial information perception and intelligent processing, Yango University, Fuzhou 350015, Fujian, China; e-mail: fhhuang@ygu.edu.cn (F.H.)

Fujian University Engineering Research Center of Spatial Data Mining and Application, Yango Univesity, Fuzhou 350015, Fujian, China; e-mail: fhhuang@ygu.edu.cn (F.H.)

Corresponding author: fhhuang@ygu.edu.cn

The tobacco plant counting is an important aspect in tobacco production management, traditional manual methods are time-consuming, labor-intensive and inaccurate, failing to meet the efficiency demands of modern agriculture. To enhance the accuracy and efficiency of tobacco plant counting in the field environment, this study utilizes high-resolution remote sensing imagery collected by drones to construct a sample dataset and proposes an improved YOLOv8-based target detection model (YOLOv8-CSD). YOLOv8-CSD model, based on YOLOv8, incorporates a coordinate attention mechanism (CA) to improve the extraction ability of the model to tobacco plant features. It also optimizes the feature pyramid network (FPN) and adds a small target detection layer to enhance the detection ability for the small target tobacco plants. Additionally, the shape intersection over ratio (SIoU) loss function is used to accelerate model convergence, and the slice-assisted hyper inference (SAHI) strategy is introduced to improve the accuracy and inference efficiency of small target detection by slicing high-resolution images. The experimental results show that the YOLOv8-

CSD model achieves a precision of 97.96%, a recall rate of 97.93%, and an average accurate mean (mAP0.5) of 99.32%, significantly outperforming the original YOLOv8 and other 5 commonly used target recognition models. In addition, the efficiency of YOLOv8-CSD model is only lower than YOlOv11, indicating good overall performance. The YOLOv8-CSD model has good adaptability and robustness in tobacco plant detection at different growth stages, with a low missed detection rate, and the YOLOv8-CSD model can effectively meet the requirements of tobacco plant counting in complex field scenarios.

KEYWORDS: Deep learning, UAV remote sensing imagery, Tobacco plant counting, Improved YOLOv8 algorithm.

1. Introduction

With the rapid development of modern agricultural technology, crop growth monitoring and fine management has become important means to improve crop yield and quality [21]. Tobacco, as a crop with high economic value, and growing tobacco has become one of the important sources of income for farmers in some areas of China. Information on Tobacco cultivation is an essential basis for tobacco information management. Accurate measurement the number of tobacco plants can provide reference basis for tobacco management department to timely and accurately grasp tobacco growth trend, adjust tobacco guiding planting plan and evaluate the implementation of relevant policies. For tobacco farmers, identifying the tobacco plants and estimating their number through intelligent algorithms can effectively monitor tobacco growth and predict yield. Through better fertilizing, weeding, controlling pests and replanting tobacco plants for the dead ones in tobacco fields, the tobacco yield and economic income can be effectively improved. Traditional methods for verifying the number of tobacco plants mainly rely on manual field measurements, which are time-consuming, labor-intensive, and cannot guarantee accuracy.

With the rapid development of remote sensing technology, the combination of machine learning and remote sensing imagery to extract tobacco planting information has become a hot topic for many researchers at home and abroad. Liu Mingqin et al. [10] estimated the area of inter-planted tobacco in mountainous areas based on the remote sensing imagery of ZU-3 satellite, and the precision of extracted tobacco area after classification processing reached 94.63%; Zhu et al. [15] proposed a method combining supervised classification and morphology to extract the tobacco field based on unmanned aerial vehicle imagery, with an accuracy of 95.93%; Fu Jing et al. [9] combined morphological methods with the Otsu algorithm to extract the number of tobacco seedlings in the UAV remote sensing imagery, with the accuracy of 93.88%. Especially in recent years, with the continuous improvement of computer computing power, deep learning technology has been widely applied in various fields. Target detection combined with unmanned aerial vehicle remote sensing technology has gradually emerged in precision agriculture. Representative related studies include: Pu et al. [17] combined UAV aerial images with an improved Tassel-YOLO algorithm to achieve real-time corn ear detection and counting, with a counting accuracy of 97.55%; Xiao Hengshu et al. [20] combined UAV aerial images with an improved YOLOv8 algorithm to achieve high-throughput tobacco leaf detection and counting. with an average accuracy of 93.6% and a model parameter of only 2.5M; Geng Lichuan et al. [3] proposed an improved YOLOv7-CSD optimization model based on YOLOv7 for tobacco plant detection. In this method, a small target detection layer is introduced to enhance enhancement algorithm. achieving an average precision mean of 94.5%. It can be seen that the existing research applies the improved YOLO network structure model to corn ear counting can achieve high accuracy and good results, but there is a problem of insufficient accuracy in application to tobacco plant recognition and counting.

Inspired by existing research results, this study proposes an improved target detection model based on YOLOv8 (YOLOv8-CSD) to enhance the precision of tobacco plant counting in open field scenarios. The main features of the model are as follows:

1 A coordinate attention (CA) mechanism module [4] is embedded in the backbone network, this module can effectively capture the interaction relationship between channel information and position features, enhance the ability of the model to identify obscured tobacco plants under different scales, and thus significantly improve the detection accuracy;
2 The YOLOv8 network structure is optimized by adding a new small target detection layer and adjusting the feature pyramid network to enhance the utilization of high-resolution features, thereby improving the detection capability and precision of the model for small tobacco plant targets;
3 The SIoU loss function is introduced to accelerate the convergence of the model, and improve the recognition accuracy of overlapping targets and enhance the adaptability of the model to complex scenarios in the tobacco field.

2. YOLOv8-CSD Model Construction

2.1. Traditional YOLOv8 Model

YOLOv8 (You Only Look Once version 8) is the latest target detection algorithm model launched by Ultralytics in 2023, aiming to improve the performance of tasks such as object detection, image classification and instance segmentation. This algorithm incorporates many innovations based on previous generations of YOLO models, including new network structure, improved loss functions and efficient training strategies. YOLOv8 introduces C2f module instead of C3 module in YOLOv5 to improve the richness and computational efficiency of gradient flow. Meanwhile, it adopts a decoupled head and anchor-free strategy simplifying the target detection process and reducing reliance on prior boxes. In addition, YOLOv8 introduces VFL Loss (Varifocal Loss) and DFL Loss (Distribution Focal Loss) to improve the convergence speed and detection accuracy of the model. The algorithm supports many methods in data augmentation, and introduces the Task-Aligned Assigner strategy for positive sample assignment in the training process, thereby enhancing overall model performance. YOLOv8 supports traditional target detection tasks, but also covers some visual tasks such as image classification and instance segmentation, exhibiting good flexibility and scalability.

Despite its strong performance and broad application prospects in the field of target detection, YOLOv8 still has certain limitations when dealing with high-density and complex scenes. Firstly, due to the dense growth and planting of tobacco plants, there are overlapping and occluded target scenes, and the feature extraction capability of YOLOv8 is relatively insufficient, making it difficult to effectively capture the interaction between channel information and spatial position information, resulting in missed or false detection of some targets. Secondly, some tobacco plants are short due to some reasons, the existing network structure has limited capability of feature extraction and expression for high-resolution images, and the design of feature pyramids does not support small targets sufficiently, affecting detection accuracy. In addition, the default loss function of YOLOv8 has a relatively average optimization effect in the high-density target scene, and the regression ability of the bounding box of the occlusion target is not ideal, and the adaptability of the model in complex scene still has room for improvement. The above problems pose challenges to the practical application of the model in the detection of high-density small targets in tobacco fields, and it is urgent to improve the performance by optimizing the network structure and improving the training strategy.

2.2. Improvement Ideas for YOLOv8

2.2.1. CA Attention Mechanism

The CA (Coordinate Attention) mechanism significantly improves the target recognition ability of deep learning models in complex scenario by embedding position information into channel attention. Traditional attention mechanisms, such as the SE (Squeeze-and-Exitation) attention mechanism [12], which improves the attention of the model to the key information by adaptively adjusting the importance of each channel. It generates channel descriptors through Global Average Pooling (GAP) and evaluates channel weights using fully connected layers, thereby strengthening the information of key channels and suppressing unimportant features, as shown in Figure 1(a). However, the SE mechanism only focuses on channel information and ignores spatial position information, which makes it difficult for the model to accurately capture the spatial features of tobacco plant targets in complex back-

ground or dense targets in tobacco field scene, affecting detection accuracy.

To compensate for the shortcomings of SE, the CBAM (Convolutional Block Attention Module) attention mechanism [7] combines channel attention and spatial attention, generates channel attention map through global pooling, and generates spatial attention map by convolution, which improves the processing capacity of spatial and channel information, as shown in Figure 1(b), where "GAP" and "GMP" refer to global average pooling and global maximum pooling, respectively. Although CBAM has some improvement in spatial information processing, it does not directly embed spatial information into channel features like the CA mechanism, so it still has limitations in complex background or dense target detection in tobacco scenario.

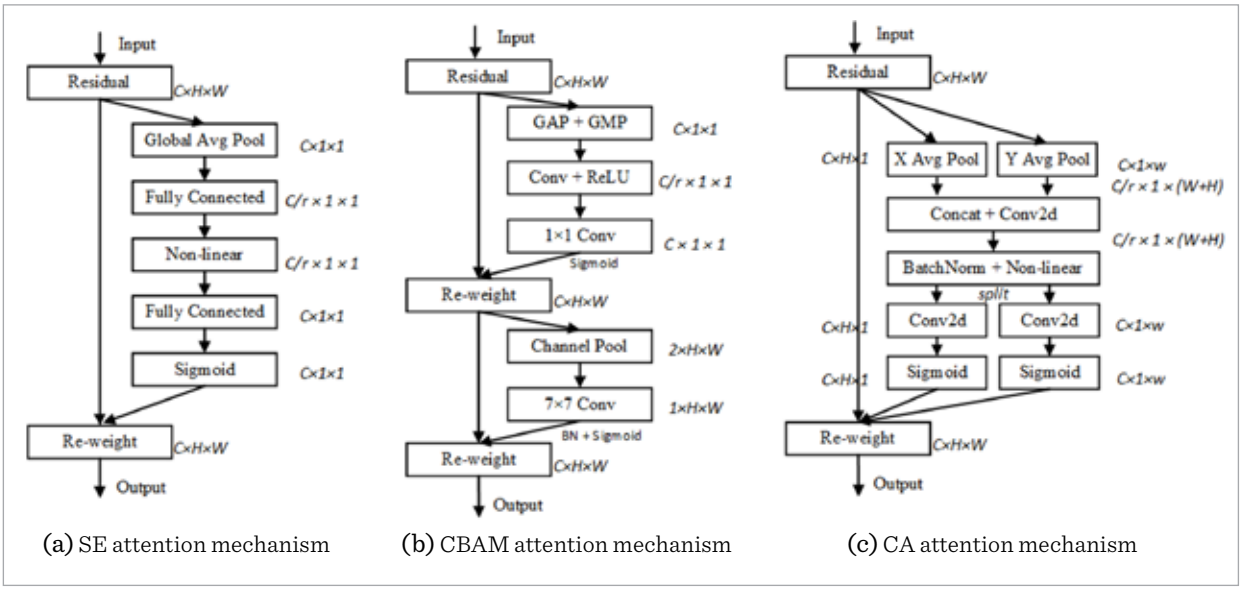
The CA (Coordinate Attention) mechanism combines spatial information with channel information, which significantly enhances the target recognition ability of the model in complex scenario. The CA performs a one-dimensional global pooling operation on the input feature map, extracting spatial features along the width and height directions, and generating corresponding attention vectors, as shown in Figure 1(c), where "X Avg Pool" and "Y Avg Pool" respectively refer to one-dimensional horizontal global pooling and one-dimensional vertical global pooling. In this way, the CA attention mechanism can not only capture the relationship between the channels, but also can effectively retain spatial information, enabling the network to more accurately identify tobacco plant targets in a complex field environments. Especially for detecting small stunted tobacco plants, it significantly improves recognition accuracy.

2.2.2. Optimization of Loss Function

IoU (Intersection over Union) can be regarded as one of the evaluation metrics and loss function in target detection, that is, calculating the ratio of the area of the intersection to the area of the union between the predicted box and the ground truth box. However, the IoU has drawbacks as it does not consider the scale difference and center point drift between the predicted box and the ground truth box. This defect may cause unstable optimization phenomenon in bounding box regression, especially in target detection with overlapping targets and largescale changes.

To alleviate this phenomenon, YOLOv8 adopts the CIoU (Complete Intersection over Union) loss function, which further considers the size difference and the center point distance between the predicted box and the ground truth box when calculating IoU and introduces a penalty term to comprehensively mea-

Figure 1
Comparison of different attention mechanisms.



(a) SE attention mechanism (b) CBAM attention mechanism (c) CA attention mechanism

sure the overlap degree, center point distance and aspect ratio matching of the target box, effectively alleviating scale mismatch and unstable regression of the boundary box. However, CIoU does not consider the direction matching between the predicted box and the ground truth box. this neglect may result in slower convergence of the model in the bounding box regression process, thereby affecting the overall performance of target detection.

Due to the complex background in tobacco fields and the overlap among tobacco plants, which easily leads to a decline in detection accuracy, to accelerate the regression process of the predicted box to the ground truth box and improve the recognition effects of tobacco plants, this study adopts the Shape Intersection over Union (SIoU) loss function [16] to replace the original CIoU loss function of YOLOv8. Formula (1) shows the SIoU loss function, which includes the angle loss Λ in Formula (2), the distance loss Δ in Formula (3) and the shape loss Ω in Formula (4). The composition principle of the SIoU loss function is shown in Figure 2. Here, B and B^{GT} represent the center points of the predicted box and the ground truth box, respectively; H and W represent the difference in the horizontal and vertical coordinates of B and B^{GT}; σ represents the distance between the center of the predicted box and ground truth box; and α represents the the angle between the line connecting the two center points and the horizontal line. SIoU optimizes the regression process of the bounding box by comprehensively considering the direction, position and scale information of the target box so as to improve the accuracy of tobacco plant detection and the convergence speed of the model.

$$Loss_{sIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \tag{1}$$

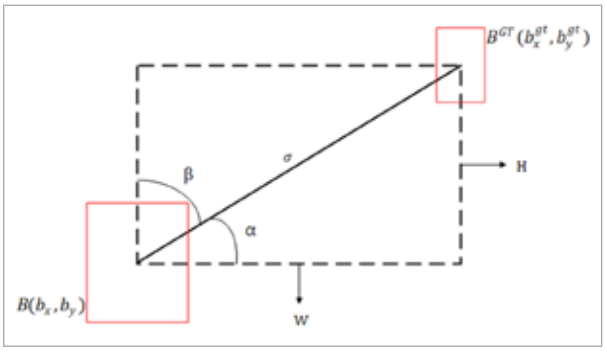
$$\wedge = \cos\left(2\left(\arcsin\left(\frac{H}{\sigma}\right)\right) - \frac{\pi}{4}\right) \tag{2}$$

$$\Delta = 2 - e^{-\gamma p_x} - e^{-\gamma p_y} \tag{3}$$

$$\Omega = \sum_{t=w,h} (1 - e^{-W_t})^{\theta} \tag{4}$$

Figure 2

Schematic Diagram of SIoU Loss Function Composition.



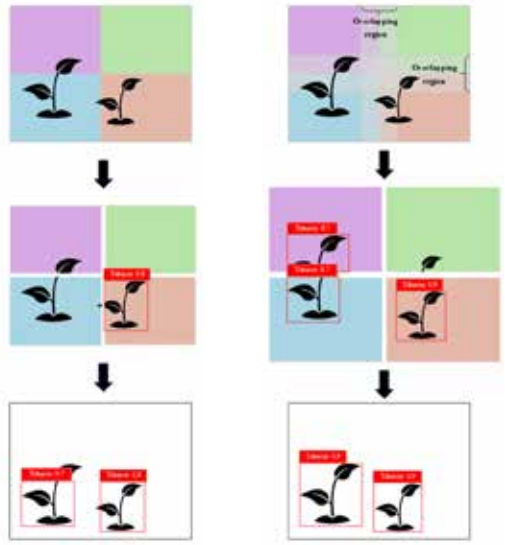
2.2.3. Addition of small object detection head

The size of UAV remote sensing imagery is usually large. Take the aerial image taken by P1 camera in DJI as an example, the size of single image is 8192×5460(unit: pixel, the same below). In the above image, the typical size of tobacco plant target is usually less than (32×32) pixels, and some slow-growing plants are even smaller in size, far smaller than the conventional tobacco plant target. Due to the small size and dense distribution of targets, traditional detection models show certain limitations in extracting and utilizing the relevant features of high-resolution UAV remote sensing imagery. Therefore, to solve this issue, this study adjusts and optimizes the network structure of YOLOv8 by adding a small object detection head with dimensions of 160×160 to the original structure's head. This enhancement strengthens the extraction and fusion capabilities of high-resolution image features, thereby improving the model's accuracy and robustness in detecting small tobacco plant targets.

2.2.4. Slice-Assisted Super Inference

Due to the fact that UAV remote sensing imagery cover multiple fields with a large number of small-sized tobacco plant targets, directly inferring on the whole image can leads to image size compression, small targets are easy to be ignored, and the large size leads to slow inference speed and excessive memory consumption. In order to solve this issue, SAHI (Slice Assisted Super Inference) is introduced to improve the detection accuracy and inference efficiency of the model. First, the original remote sensing image is sliced into fixed sizes (such as 640×640), with a reasonable overlap rate set during slicing according

Figure 3
Schematic diagram of block overlap inference.



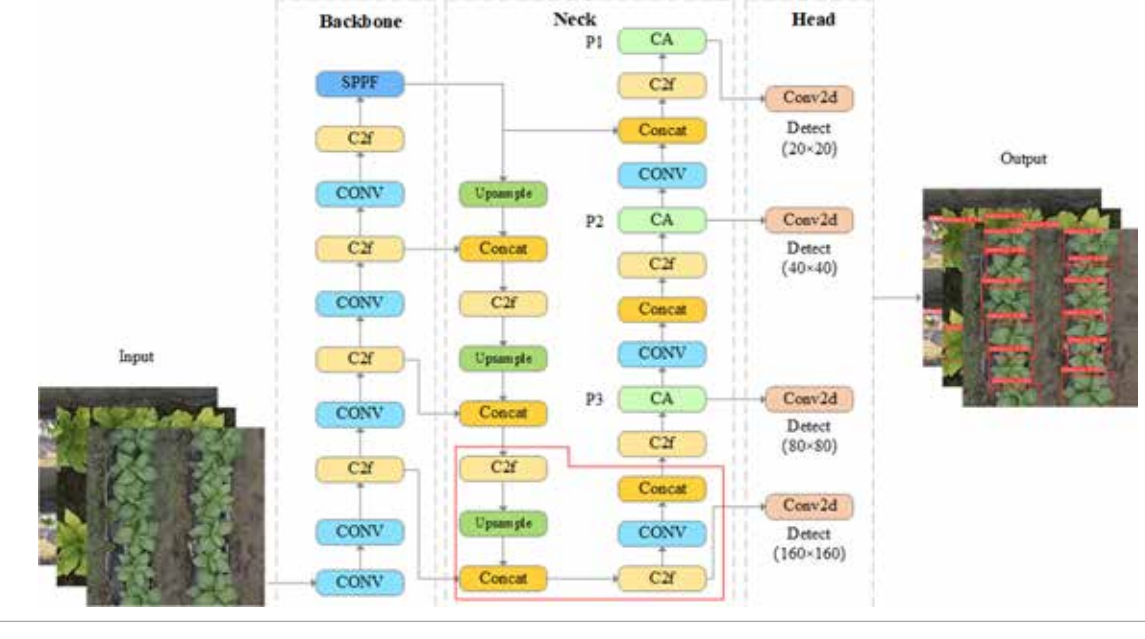
(a) block non-overlap (b) block overlap

to the size of the tobacco plants and their spacing to avoid losing edge information due to the slicing line passing through the target. The blocking strategy is shown in Figure 3. Then, the improved YOLOv8 model is used to perform independent reasoning on each slice to generate local detection results. And finally, Non-Maximum Suppression (NMS) is employed to merge targets in overlapping areas, combining the inference results from multiple slices to form the overall detection result.

2.3. YOLOv8-CSD Model Design

In summary, this study improves the YOLOv8 model from three aspects: feature extraction, bounding box regression optimization and inference strategy, constructing the YOLOv8-CSD model. Through introducing CA attention mechanism, YOLOv8-CSD model enhances the expression capability of spatial and channel information; adopts SIoU loss function to improve the positioning accuracy and convergence speed of bounding box; adds a small target detection head to enhance high-resolution feature extraction capacity; and combines with SAHI strategy to effectively alleviate the problems of small target missed detection and inference resource consumption. The network structure of the YOLOv8-CSD model is shown in Figure 4. Here, the red solid line box indicates the addition position of the small target detection layer, while positions 1, 2 and 3 are the addition positions of the CA attention mechanism.

Figure 4
Network Structure Diagram of YOLOv8-CSD Model.



YOLOv8-CSD is an improved architecture based on YOLOv8, which combines Coordinate Attention (CA) module and multi-scale feature fusion mechanism to improve the detection accuracy of small tobacco plant targets. The network consists of three main modules: Backbone, Neck and Head. In the backbone network, the multi-scale features of the image are extracted step by step through spatial pyramid pooling (SPPF) and multi-layer convolution (Conv). Then, the neck network utilizes the CA module to introduce spatial attention mechanism to enhance the focus on small tobacco plant targets, combined with upsampling (Upsample) and feature concatenation (Concat) to further enrich the feature map. In the head part, multi-scale detection boxes (20x20, 40x40, 80x80, 160x160) are used to predict targets of different sizes, thus achieving precise detection of various scale objects. Finally, through this multi-scale fusion and attention mechanism design, YOLOv8-CSD shows significant performance improvement in processing small target detection tasks of tobacco plants.

3. Materials and Methods

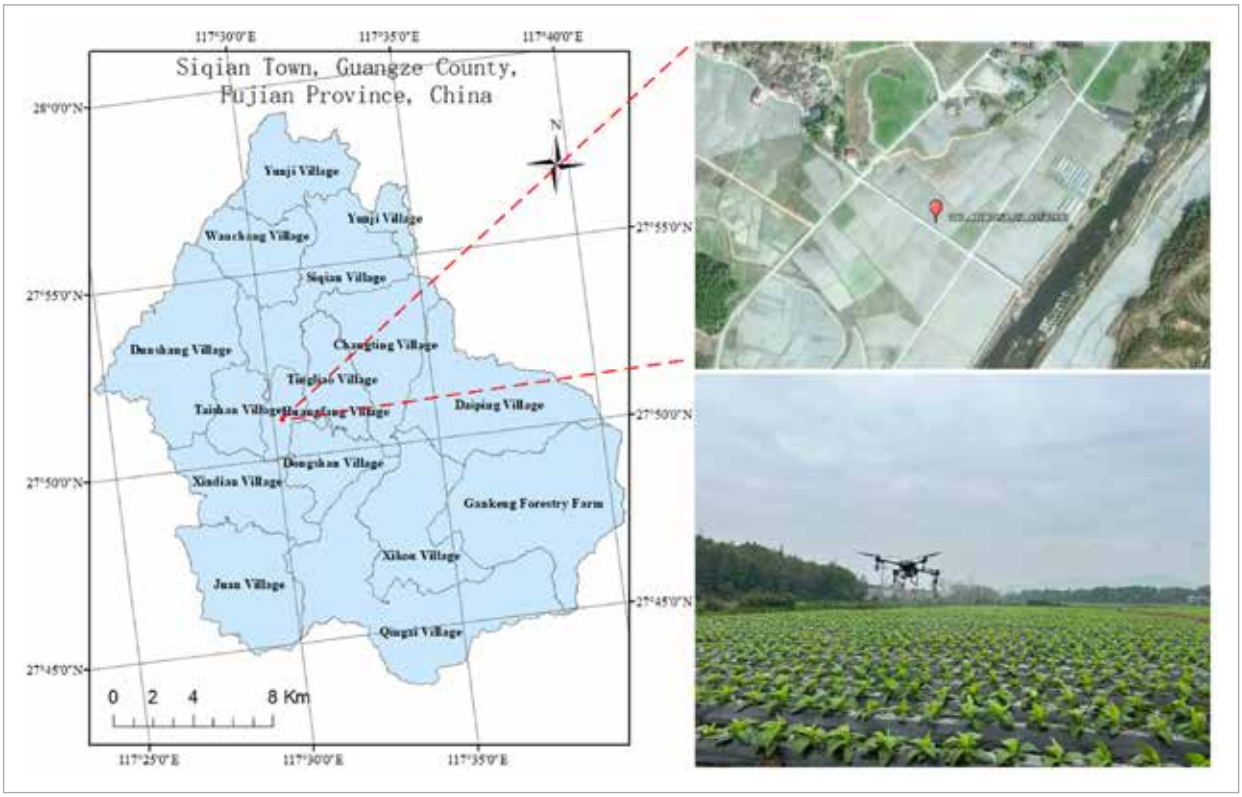
3.1. YOLOv8-CSD Model Design

The data set used in this experiment was collected from the tobacco planting area in Siqian Township, Guangze County, Nanping City, Fujian Province (with the central geographic coordinates at 27°51′N and 117°30′E, and an altitude of 332 meters). The climate in this area is mild, making it highly suitable for tobacco cultivation. Additionally, the area features flat and expansive terrain, providing convenient conditions for large-scale data collection using UAV. The geographical position of the experimental site is shown in Figure 5.

3.2. Dataset Construction

An UAV (model: DJI-M350 RTK) equipped with a high-resolution aerial camera (model: DJI Zenmuse P1) was used to collect orthogonal images of the tobacco fields under favorable weather conditions, low wind speed, no strong direct sunlight, and with-

Figure 5
Geographical location of the experimental region.



out shadow shielding, The UAV autonomously flew along predetermined flight paths at an altitude of 40 meters, with continuous image capture set at an interval of 2 seconds. The collected images were then preprocessed to obtain orthogonal images of the tobacco fields in the experimental area, with a spatial resolution of 3.5 centimeters per pixel.

Since the spectral and texture features of the tobacco fields vary greatly during different growth stages. To improve the accuracy of the tobacco plant recognition model, experiments were conducted during three key growth periods of the tobacco plants: the root extension stage, the vigorous growth stage, and the maturity stage. This was done to establish a recognition model for accurately counting the number of tobacco plants in complex field scenarios. In this study, 200 RGB images (600 in total) of tobacco fields from the experimental area were collected during the root extension stage, the vigorous growth stage, and the maturity stage to form a sample dataset. Part of the tobacco plant image samples collected at each growth stage are shown in Figure 6.

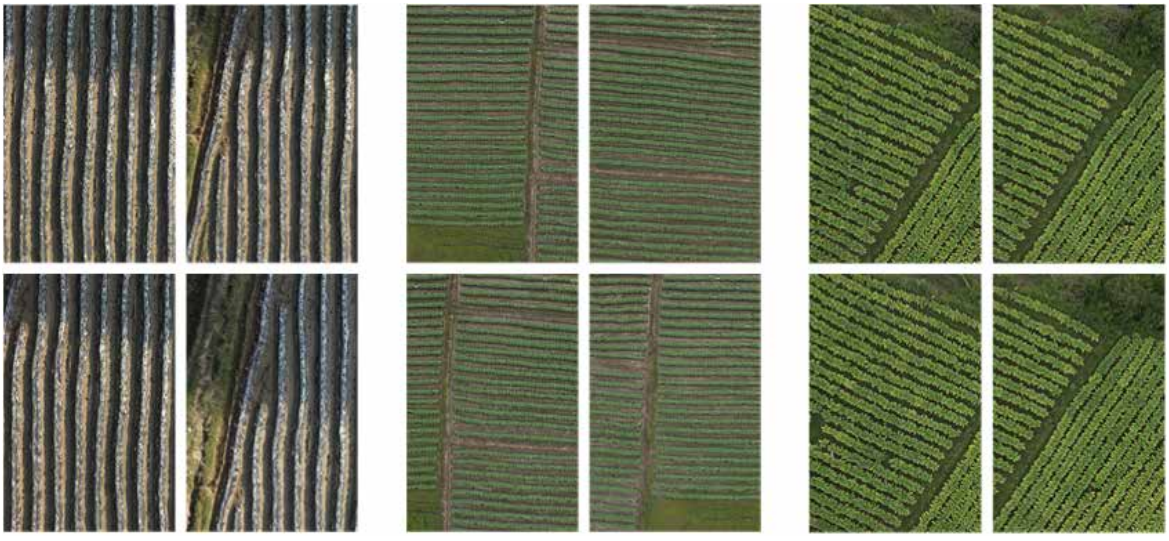
To accommodate hardware constraints and experimental parameters, as well as to simplify the model training process, this study first randomly cropped the 600 collected images and then removed some images that did not contain any tobacco plant targets. finally, datasets consisting of 10,000 640×640 RGB images of tobacco plants at each growth stages were obtained. Labellmg tools was used to label tobacco plants, with the target label set as “tobacco”. After selecting the position of the maximum bounding box, the file was saved in XML format, which contains the pixel coordinates of target tobacco leaves and their corresponding labels in the image. Then, the datasets collected in each growth stage were divided into training sets, validation sets, and test sets. Specially, the training set consisted of 8,000 images (80%), while the validation set and the test set each consisted of 1,000 images (10%). This process ultimately resulted in the construction of three datasets, each containing 10,000 tobacco plant images.

4. Experimental Analysis

4.1. Experimental Environment and Parameter Configuration

The relevant experiments of this study were carried out on Windows 10 operating system, the host computer was equipped with an Intel (R) Core (TM) i7-12700F CPU, 32GB of RAM, and an NVIDIA GeForce RTX 4090 24GB graphics card; Python was used as

Figure 6
Part of tobacco plant image samples.



(a) root extension stage (b) vigorous growth stage (c) maturity stage

the programming language, and PyTorch was used as the deep learning framework; 640×640×3 was used as the training image size. The model training consisted of 200 epochs, with the batch size automatically adjusted. All other configuration parameters were set to their default values.

4.2. Evaluation Metrics

To measure the accuracy of the model, this study adopted the following evaluation metrics: Precision (*P*), Recall (*R*), Average Precision (*AP*) and Mean Average Precision (*mAP*), which are calculated as shown in Equations (5)-(8).

$$P = \frac{TP}{TP + FP}. \tag{5}$$

Precision, which represents the proportion of all samples that were correctly predicted as positive among all samples predicted as positive, where, *TP* (True Positive): number of samples

$$R = \frac{TP}{TP + FN}. \tag{6}$$

Recall rate, which represents the proportion of all samples that were correctly predicted as positive among all actual positive samples. Where, *TP* (True Positive): the number of samples correctly predicted as positive; *FN* (False Negative): the number of samples incorrectly predicted as negative.

$$AP = \int_0^1 P_{smooth}(r)\,dr. \tag{7}$$

Average accuracy, which represents the weighted average of the accuracy at different recall rates (calculated by smoothing the accuracy curve), where, P_{smooth} (w): the smoothed precision rate value as a function of the recall rate *r*.

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K}. \tag{8}$$

The average Accuracy of all categories, which reflects the overall performance of the model in multi-category tasks. where *K* represents the total number of categories in the model evaluation. AP_i: Indicates the average precision of the *i*-th class.

In additon, this work uses the prediction time (*T*) to evaluate the efficiency of the proposed YOLOv8-CSD model.

4.3. Ablation Experiment

Based on the previous analysis, this study improved the YOLOv8 algorithm to obtain the YOLOv8-CSD algorithm. To verify the performance improvement of the YOLOv8-CSD algorithm, an ablation experiment was set up. In the experiment, the dataset was randomly selected from the sample image sets of the root extension stage, the vigorous growth stage, and the maturity stage, with 5000 images from each stage (totaling 15,000 images). All models used the same dataset and training parameters. The attention mechanism module, replacement loss function and small object detection head were added to YOLOv8 in turn. Through comparison of multiple groups of experiments, the final experiment results are shown in Table 1, where √ indicates that the module is used and indicates that the module is not used. As shown

Table 1
Ablation Experiment Results.

S/N	Add attention mechanism	Replace loss function	Add small object detection head	P/%	R/%	$mAP_{0.5}$/%
1	×	×	×	95.40	94.62	96.23
2	√	×	×	95.73	95.44	97.30
3	×	√	×	96.76	96.71	97.32
4	×	×	√	96.57	96.98	97.29
5	√	√	√	97.96	97.93	99.32

in Table 1, after CA attention mechanism is added to the original YOLOv8 network structure, compared with the original model, the accuracy rate, recall rate and $mAP_{0.5}$ are increased by 0.33%, 0.82% and 1.07%, respectively; after the original CIoU loss function is replaced by SIoU, the accuracy rate, recall rate and $mAP_{0.5}$ are improved by 1.36%, 2.07% and 1.09%, respectively; after a small object detection head is added in the head, the accuracy, recall rate and $mAP_{0.5}$ are respectively increased by 1.17%, 2.36% and 1.06%, respectively; the accuracy, recall rate and $mAP_{0.5}$ of the improved YOLOv8-CSD model for tobacco plant recognition in this study are 97.96%, 97.93%, 99.32%. Compared to the original YOLOv8 model, the accuracy, recall rate, and $mAP_{0.5}$ increased by 2.56%, 3.31%, and 3.09%, respectively.

4.4. Comparison Experiment of Recognition Performance of Different Network Models

To verify the effectiveness and superiority of the improved YOLOv8 algorithm (YOLOv8-CSD) proposed in tobacco plant recognition and counting, it was compared with five other commonly used object detection algorithms. The five algorithm models for comparison include YOLOv5, YOLOv7, YOLOv11, SSD (Single Shot MultiBox Detector) [8, 11, 18, 22] and Faster R-CNN [1, 2, 5, 6, 11, 13, 14, 19]. During the comparative experiment, both YOLOv8-CSD and the aforementioned 5 algorithms were trained using the same dataset as in the ablation experiment Section 3.3, with optimally configured training parameters, and the performance evaluation metrics mentioned in Section 3.2 were used for evaluation. The specific experiment results are shown in Table 2.

Table 2
Comparison results of detection performance of different models.

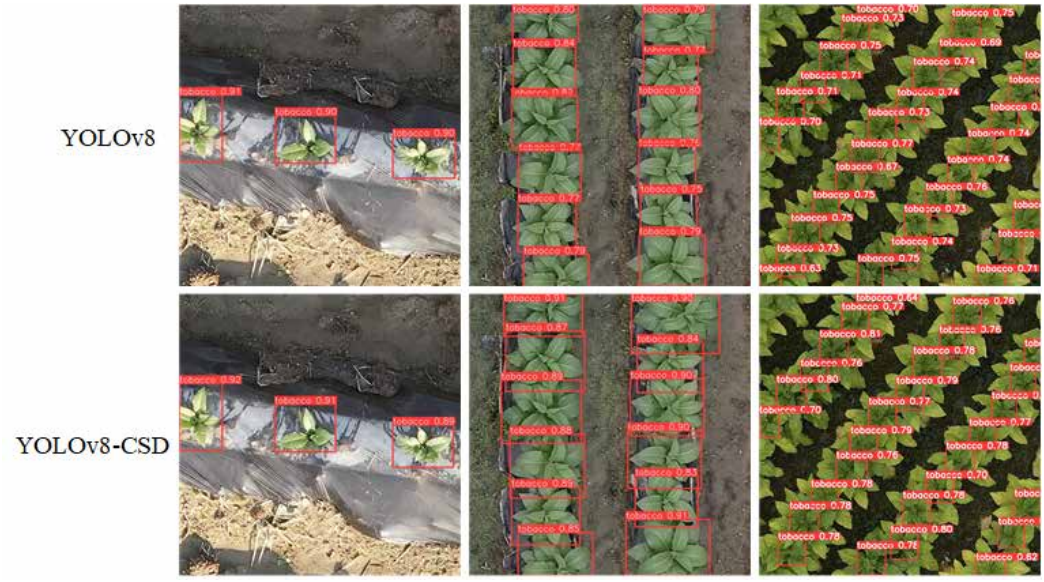
Model	*P*/%	*R*/%	$mAP_{0.5}$/%	*T*/*s*
YOLOv5	93.8	91.7	97.1	100.43
YOLOv7	94.1	93.0	97.5	70.34
YOLOv11	93.7	92.7	97.7	48.11
SSD	87.5	75.6	82.1	80.92
Faster R-CNN	75.6	60.9	72.6	69.13
YOLOv8-CSD	97.6	97.9	99.3	61.54

It can be seen from Table 2 that compared with the earlier YOLOv5 and YOLOv7 algorithms, the YOLOv8-CSD algorithm proposed in this study has improved the accuracy, recall rate and $mAP_{0.5}$ by 3.8%, 6.2% and 2.2%, and 3.5%, 4.9% and 1.8%, respectively; even compared with the upgraded YOLOv11 algorithm of YOLOv8, YOLOv8-CSD also performs better, with the accuracy, recall rate and $mAP_{0.5}$ higher by 3.9%, 5.2% and 1.6%, respectively. The detection accuracy of YOLOv7 is second only to that of YOLOv8-CSD. Faster R-CNN performed the worst in all metrics, followed by SSD, whose detection accuracy was higher than Faster R-CNN but lower than both YOLOv11 and YOLOv5. This is due to the complex environment of the tobacco fields, dense plant targets, and small target sizes. The above comparison experiment results show that the accuracy of YOLOv8-CSD algorithm is superior to the other 5 algorithms in complex back-ground, fully demonstrating the high applicability of this model for tobacco plants in complex tobacco field environments. In addition, according to Table 2, although the YOLOv8-CSD model has the highest prediction accuracy, its time consumption is only higher than YOlOv11, indicating good overall performance.

4.5. Comparison of Detection Effect in Different Growth Stages Before and After Model Improvement

Due to the asynchronous planting time of tobacco leaves, there are differences in the growth forms and plant densities of tobacco plants at different periods, making manual plant count time-consuming and laborious. In order to test the effect of the improved tobacco plant detection model in this study, the UAV remote sensing imagery of tobacco plants in three different growth stages, i. e. root extension stage, vigorous growth stage and maturity stage, were selected as the effect test data, and input into YOLOv8 and YOLOv8-CSD for inference, respectively. The inference results are shown in Figure 7. It can be seen from Figure 7 when using the above two models for inference, the detection effects in the three different growth stages are ranked basically the same, that is, the confidence in the root extension stage is the highest, followed by the vigorous growth stage, and the maturity stage is the worst. Confidence is a probability estimation of the model of whether a certain prediction box contains the

Figure 7
Recognition effects of YOLOv8 and YOLOv8-CSD on tobacco plants in different growth stages.



(a) root extension stage (b) vigorous growth stage (c) maturity stage

tobacco plant target. The higher the value, the more the model considers that the target does exist in the predicted box; although there are differences in the recognition confidence of YOLOv8-CSD for tobacco plants in different growth stages, the recall rate is kept at a high level. The high recall rate indicates that the model can effectively reduce missed detection, which means that the model has higher reliability in counting the number of tobacco plants. Meanwhile, the YOLOv8-CSD model proposed in this study effectively highlights important channel features by introducing the CA (coordinate attention) mechanism, enhancing the model's ability to focus on the key information. By adding a small object detection head, the model becomes more sensitive to details, allowing for more accurate capture of the characteristics of the tobacco plant. The experimental results provide a solid technical support for the subsequent statistics of the number of tobacco plants.

4.6. Tobacco Plant Counting Effect in the Field Scenario

The counting of tobacco plants can provide a reference for tobacco management departments to timely and accurately grasp the situation of tobacco production, adjust the guiding planting plan of tobacco and evaluate the implementation of relevant policies. An unmanned aerial vehicle (UAV) is used for flight route planning in an experimental area, orthophoto images of a tobacco field are continuously taken. Then, the above photos were stitched into a single orthogonal image of the experimental area as a whole. Finally, the YOLOv8-CSD model was used for inference and counting. The effect of tobacco plant recognition and counting in a large tobacco field (vigorous growth stage of tobacco plants) in the experimental area is shown in Figure 8. The area of this large tobacco field plot is approximately 958 square meters., and the number of tobacco plants counted by the YOLOv8-CSD model is 1219, and compared with the manual counting result of 1226 plants, the missing detection rate of the tobacco plants was 0.5%; in addition, using the same method to count the number of tobacco plants in the above mentioned large tobacco field in the root extension stage and maturity stage, the missing detection rates are about 0.4% and 1%, respectively. The results show that the YOLOv8-CSD model proposed in this study can meet the needs of counting the number of tobacco plants in complex tobacco field scenarios.

Figure 8
Statistical effect of YOLOv8-CSD tobacco plant number in field scenario.

5. Conclusion

This study addresses the limitations of existing deep learning network model in accurately counting the number of tobacco plants in complex scenarios by proposing a tobacco plant recognition and counting method based on YOLOv8-CSD. The YOLOv8-CSD model improves the perception of tobacco plant characteristics by embedding a CA attention mechanism; replaces the original loss function to optimize the penalty metrics vector angle; adds a small object detection head to more accurately capture tobacco plant targets in high-resolution remote sensing imagery; introduces a slice-assisted hyper-inference (SAHI) strategy to solve the problems of wide image coverage, dense tobacco plants and small targets, thereby improving detection accuracy and inference efficiency. The results showed that the accuracy and recall rate of YOLOv8-CSD model were better than those of Fast R-CNN, SSD, YOLOv5, YOLOv7, YOLOv11 and original YOLOv8 models, and the efficiency of YOLOv8-CSD model was only lower than YOlOv11, indicating good overall performance. During the three different growth stages of tobacco plants, YOLOv8-CSD effectively identifies tobacco plants with high confidence and robustness, enabling accurate counting. YOLOv8-CSD can effectively solve various issues existing in traditional manual tobacco plant counting method, improve the automation level in the tobacco cultivation, and provide an important reference for tobacco management departments to accurately grasp the production situation, adjust the planting plan and evaluate the implementation of policies. Although the proposed YOLOv8-CSD model still exhibits minimal missed detections in practical tobacco plant counting, future work will focus on further optimizing and lightweighting the model to provide a more accurate and efficient solution for tobacco plant counting. Especially, we plan to further improve the YOLOv8-CSD network structure or replace YOLOv8-CSD with higher version of YOLO (such as YOLOv9 or YOLOv11), and introduce lightweight network architecture, such as the Generalized Efficient Layer Aggregation Network (GELAN) based on gradient path planning, so as to reduce the parameter count of the new YOLO model and improve the recognition and quantity statistics accuracy of tobacco plants. In addition, the YOLOv8-CSD model proposed in this work is specifically designed

for tobacco plant quantity statistics, so its generalization ability for other crop categories still needs to be further verified. In subsequent research, training samples of other crop categories can be supplemented and the network structure can be improved to construct an improved YOLOv8-CSD model to enhance its generalization ability.

Funding

This research was funded by Fujian Province Science and Technology Plan Project (Guided Project, 2023N0021), the National Natural Science Foundation of China (NSFC, 41501451), China Postdoctoral Science Foundation (2015M571963).

References

1. Chen, Z. L., Guo, Y., He, J., Wang, L. G., Zeng, K., Wang, A. G., Yang, X. Z., Zhang, D. Y., Niu, Z. J. Research on Lightweight Detection and Recognition of Tobacco Disease Based on RT-YOLOv10 and UAV Remote Sensing Images. Shandong Agricultural Sciences, 2025, 57(09), 154-168. https://doi.org/10.2139/ssrn.5366881
2. Fu, B. H., Huang, L. Extraction of Tobacco Planting Area from UAV Images Based on Deep Semantic Segmentation. Communications Tech-nology, 2022, 55(02), 181-186. https://doi.org/10.3969/j.issn.1002-0802.2022.02.007
3. Geng, L. C., Wang, Z. F., Qin, Y. Z., Ma, L. To-bacco Plant Quantity Detection Based on YOLOv7-Sim and UAV Remote Sensing Image-ry. Chinese Tobacco Science, 2023, 4, 94-102. https://doi.org/10.13496/j.issn.1007-5119.2023.04.013
4. Hou, Q., Zhou, D., Feng, J. Coordinate Attention for Efficient Mobile Network Design. Proceed-ings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021), Nashville, USA, June 20-25, 2021, 13713-13722. https://doi.org/10.1109/CVPR46437.2021.01350
5. Hu, Z. H., Wang, C. F. Improved Fast R-CNN Algorithm for Remote Sensing Imagery Small Target Detection. Computer Engineering and Science, 2024, 6, 1063-1071. https://doi.org/10.3969/j.issn.1007-130X.2024.06.013
6. Jian, R. B., Cai, Z. Y., Yang, Z. S., Wang, W. Z., Liu, Y., Chen, J. F., Wang, M. L. Research on Image Segmentation Method of Field Wheat Harvest Boundary Based on U-Net. Journal of Henan Agricultural University, 2023, 57(03), 444-450. https://doi.org/10.16445/j.cnki.1000-2340.20230322.001
7. Jiang, P., Zhang, J. Z., Chen, J. J. Enhanced Rain Removal Network with Convolutional Block At-tention Module (CBAM): A Novel Approach to Image De-Raining. EURASIP Journal on Ad-vances in Signal Processing, 2025, 9(2025). https://doi.org/10.1186/s13634-025-01212-z
8. Jiang, S., Xue, B. SSD Object Detection Algo-rithm Based on Multi-Step Multi-Scale Feature Fusion. Computer and Digital Engineering, 2024, 10, 2972-2976.
9. Li, Q. X., Zhou, Z. F., Qian, Y. Z., Yan, L. H., Huang, D. H., Yang, Y., Luo, Y. N. Accurately Segmenting/Mapping Tobacco Seedlings Using UAV RGB Images Collected from Different Ge-omorphic Zones and Different Semantic Seg-mentation Models. Plants, 2024, 13(22), 3168-3193. https://doi.org/10.3390/plants13223186
10. Lin, H., Chen, Z. Q., Qiang, Z. P., Tang, S. K., Liu, N., Pau, G. Automated Counting of To-bac-co Plants Using Multi-Spectral UAV Data. Agronomy, 2023, 13(12), 2861-2881. https://doi.org/10.3390/agronomy13122861
11. Ling, Y. J. Research on Optimization of Video Surveillance Target Detection Algorithm Based on YOLO, SSD and Fast R-CNN. Modern Com-puter, 2024, (21), 49-53, 58. https://doi.org/10.3969/j.issn.1007-1423.2024.21.009
12. Ma, Y. Y., Li, S. T., Zhou, S. L., Wang, X. Y. Per-formance Degradation Prediction of Proton Exchange Membrane Fuel Cells Based on CNN-LSTM Network with Squeeze-and-Excitation At-tention Mechanism. Energy, 2025, 335. https://doi.org/10.1016/j.energy.2025.138127
13. Magdy, A., Moustafa, M. S., Ebied, H. M., Tolba, M. F. Lightweight Faster R-CNN for Object De-tection in Optical Remote Sensing Images.

Scien-tific Reports, 2025, 15(1), 16163-16176. https://doi.org/10.1038/s41598-025-99242-y

14. Miao, R., Li, Y., Zhou, K., Zhang, Y. N., Chang, R. R., Meng, G. Research on Improved Faster R-CNN Multi-Target Detection Model for Remote Sensing Imagery. Computer Engineering, 2024. https://doi.org/10.19678/j.issn.1000-3428.0068856
15. Murphy, K. M., Ludwig, E., Gutierrez, J., Gehan, M. A. Deep Learning in Image-Based Plant Phe-notyping. Annual Review of Plant Biology, 2024, 75, 771-795. https://doi.org/10.1146/annurev-arplant-070523-042828
16. Peng, H., Yu, S. A Systematic IOU-Related Method: Beyond Simplified Regression for Better Localization. IEEE Transactions on Image Pro-cessing, 2021, 30, 5032-5044. https://doi.org/10.1109/TIP.2021.3077144
17. Pu, H. L., Chen, X., Yang, Y. Y., Tang, R., Luo, J. W., Wang, W. C., Mu, J. Tassel-YOLO: A New High-Precision and Real-Time Method for Maize Tassel Detection and Counting Based on UAV Aerial Imagery. Drones, 2023, 8, 492-510. https://doi.org/10.3390/drones7080492
18. Sun, R. K., Ying, P., Li, Z. N., Xu, X. Z. Small and Weak Target Detection Based on Light-weight SSD. Computer Simulation, 2024, 10, 355-361. https://doi.org/10.3969/j.issn.1006-9348.2024.10.068
19. Tian, T., Wang, D., Wang, Z., Li, H. Precise Classification of Crops in Complex Planting Structure Area Based on Deep Learning Model. Chinese Journal of Agricultural Resources and Regional Planning, 2022, 43(12), 147-158.
20. Xiao, H. S., Li, J. Y., Liang, H., Ma, E. D., Zhang, H. Detection of Tobacco Plant Numbers in Large Fields Based on Improved YOLOv8 and UAV Remote Sensing Imagery. Electronic Measure-ment Technology, 2024, 9, 163-171. https://doi.org/10.19651/j.cnki.emt.2415549
21. Zhang, R., Wu, X., Li, J., Zhao, P. Y., Zhang, Q., Lige, W., Zhang, D. H., Zhang, Z. J., Yang, L. N. A Bibliometric Review of Deep Learning in Crop Monitoring: Trends, Challenges, and Future Per-spectives. Frontiers in Artificial Intelligence, 2025, 8. https://doi.org/10.3389/frai.2025.1636898
22. Zheng, X. Y., Zhao, X. L. Single Shot MultiBox Detector-Based Feature Fusion Model for Building Object Detection. Journal of Applied Science and Engineering, 2025, 28(2), 391-398.