TOBACCO LEAF DETECTION MODEL BASED ON YOLOV7 AND MOBILENETV3+DCN FUSION

/

基于 YOLOv7 与 MobileNetV3+DCN 融合的烤烟检测模型

Jun XIAO[1], Lili ZHU[1*], Chengwei ZHANG[2], Hao JIANG[3], Liang ZHANG[2], Guoxin SHI[3]

[1] China Tobacco Guangxi Industrial Co., Ltd., Internet Research Center, Nanning, 530001, China;

[2] Jiangsu Newdee Digital Technology Co., Ltd., Xuzhou, 221000, China;

[3] Xuzhou Xinyun Institute of Public Credit, Xuzhou, 221000, China

**Coresponding author, Email: lilizhu0336@163.com*

DOI: https://doi.org/10.35633/inmateh-78-66

***Keywords:** YOLOv7; MobileNetV3; Tobacco leaf detection; DCN; CBAM*

ABSTRACT

To address the problems of low efficiency, strong subjectivity, and high cost associated with traditional tobacco leaf detection methods, an identification model suitable for tobacco leaves was developed to achieve rapid and non-destructive detection and to support the standardization of tobacco production. In this study, the Convolutional Block Attention Module (CBAM) was improved to enhance the feature extraction capability of tobacco leaves and highlight key feature information. MobileNetV3 and a Deformable Convolution Network were integrated to optimize the model structure, thereby reducing the number of parameters and computational complexity. Based on these improvements, a tobacco leaf detection model was constructed. Experimental results showed that the proposed algorithm achieved an accuracy of 97.01% with a loss value of 0.09, outperforming the Multi-Marker Similarity Assessment method and the Gorilla Troop Optimization Algorithm. The constructed model achieved an accuracy of 94.65%, a recall of 91.24%, and an F1 score of 93.54%. The model contains 9.36 M parameters and has a size of 50.69 MB, demonstrating better performance compared with the reference models. The results indicate that the improved tobacco leaf detection model can significantly enhance detection efficiency and accuracy. This study provides a useful approach for precise tobacco leaf detection in complex field environments and contributes to the development of modern intelligent agriculture.

摘要

针对烤烟检测传统方法存在的效率低下、主观性强以及成本高昂等问题，研究构建了一种适用于烟草叶片的鉴定模型，旨在实现烟草叶片的快速无损鉴定，推动烤烟生产的标准化。因此，本研究改进了卷积块注意力模块，优化烟叶特征提取能力，突出关键特征信息，并结合 MobileNetV3 和可变形卷积网络优化模型结构，降低参数和计算负荷。最后，构建了烟叶检测模型。实验结果表明，所提出的算法准确率达到 97.01%，损失值为 0.09，优于多标记相似性评估算法和猩猩部队优化算法。所构建模型的准确率为 94.65%，召回率为 91.24%，F1 得分为 93.54%。其参数和模型大小分别为 9.36 M 和 50.69 MB，表现优于对比模型。这些结果表明，改进后的烟叶检测模型可以提高检测效率和准确性。这项研究有助于在复杂的田间环境中精确构建烟叶检测模型，促进现代农业的发展。

1. INTRODUCTION

1.1 Background

Tobacco leaf, as an important economic crop, requires accurate recognition and status detection during the production process to improve its quality (*Moşteanu, 2025*). However, tobacco leaves vary in shape and posture, and shadows on leaves change with lighting conditions, affecting precise harvesting and intelligent field management, which leads to increased time waste and resource consumption. Therefore, effective recognition and detection of tobacco leaves is of great significance (*Bitzer et al., 2023*). Currently, traditional tobacco leaf detection methods still rely on manual operations and simple tools, which can easily cause missed or false detection. These methods have strong dependency on labor, low efficiency, high subjectivity, and high costs, showing certain limitations. Therefore, there is an urgent need for a method that can accurately and efficiently detect tobacco leaf quality (*Eği, 2023*). With the rapid development of computer technology, intelligent algorithm-based tobacco leaf detection models have become a research hotspot in recent years. Intelligent algorithms have advantages such as high efficiency, objectivity, and reduced dependence on human labor.

Among them, You Only Look Once version 7 (YOLOv7) ensures efficient gradient backpropagation and fully learns deep network features, while MobileNetV3 network applies comprehensive lightweight design to optimize model structure and reduce computational load (*Pramudhita et al., 2023, Brock, 2024*).

1.2 Related works

YOLOv7 and MobileNetV3 algorithms had advantages such as high efficiency and strong objectivity, and they solved complex problems in various fields. Mammeri S et al. proposed a lung nodule detection model based on YOLOv7 for lung nodule diagnosis. They captured important information by drawing bounding boxes and classified nodule types using transfer learning and the Visual Geometry Group 16-layer Convolutional Neural Network, which improved classification accuracy (*Mammeri et al., 2024*). *Yonglin et al*. adjusted the network structure of YOLOv7, expanded the receptive field, modified the model output parameters, and changed the feature scale relationships to solve difficulties in infrared image detection. They fused weight information with shallow networks to improve target localization accuracy (*Yonglin, Hengtao and Shang, 2024*). *Acikgoz H. et al.* collected defect images in photovoltaic modules and integrated a ghost module with global attention mechanism into YOLOv7. This enhanced the learning ability of the algorithm and reduced parameter computation through structural optimization, finally constructing a YOLOv7-based photovoltaic cell detection model (*Acikgoz, 2024*). *Shen J. et al.* (*2024)* proposed a lightweight model based on YOLOv8 and MobileNetV3 for insulator defect detection in power systems. They reduced model complexity using MobileNetV3 and introduced Deformable Convolution Network in the bottleneck structure to improve feature extraction for dense small targets. Results showed that the proposed model achieved superior recognition accuracy. *Prasad and Chandana (2023)* proposed a MobileNetV3 network model for facial expression classification. They applied mean filtering for image denoising, min-max normalization for data processing, used the binary dragonfly algorithm to extract facial features, and finally achieved high-precision facial expression classification (*Prasad 2023*).

Accurate detection of tobacco leaves significantly improved tobacco quality and played an important role in modern agriculture. The theory and practical application of tobacco leaf detection had become well established, and numerous scholars had carried out extensive investigations. *Thimmegowda et al.* proposed a color space model for image segmentation and preprocessing, combined flow histogram, motion boundary histogram, and oriented gradient optimal histogram to explore leaf movement trajectories, and analyzed spatial changes through the oriented gradient optimal histogram for tobacco crop classification (*Thimmegowda and Jayaramaiah, 2023*). *Kim et al. (2023)* developed a neural network model based on fuzzy clustering to address tobacco leaf quality identification. They fused multiple histograms for image processing, applied dimensionality reduction for feature extraction, and incorporated fuzzy clustering into the tobacco leaf classifier, achieving high-precision quality classification. *Wang and Qin, (2022)* extracted image feature information through denoising, standardized and cleaned the data, digitized labels, and input the processed data into a fusion algorithm combining Long Short-Term Memory (LSTM) and Extreme Gradient Boosting, constructing a tobacco leaf state prediction model. *Liu et al., (2023),* proposed a Raman spectroscopy-based tobacco leaf detection model for rapid quality identification. They highlighted key quality factors using multiple statistical methods, corrected multiple scattering effects with Raman spectroscopy, and analyzed functional groups to achieve rapid quality assessment. *Wei et al., (2022),* predicted tobacco leaf component content using near-infrared spectroscopy and applied deep transfer learning for online quality monitoring. They predicted seasonal and temperature changes and quantified chemical components through a deep neural network, ultimately constructing a tobacco leaf detection model based on deep transfer learning.

1.3 Research significance and innovation

Existing studies had made progress in tobacco leaf detection, but limitations remained, such as low feature extraction efficiency and high computational cost. Therefore, this study used the Convolutional Block Attention Module (CBAM) to enhance the model's ability to extract key feature information. MobileNetV3 was introduced to replace the backbone network of YOLOv7 to improve computational efficiency. On this basis, traditional convolution layers were removed, and a Deformable Convolutional Network (DCN) optimized by Simulated Annealing (SA) was introduced to dynamically adjust sampling positions. Finally, a tobacco leaf detection model was constructed, aiming to achieve efficient detection, reduce labor dependence, and minimize resource waste, meeting the needs of modern agriculture. The innovation of this study lay in combining CBAM with YOLOv7, using the improved DCN to further optimize MobileNetV3, enhancing detection efficiency, and constructing a tobacco leaf detection model based on the fusion of YOLOv7 and MobileNetV3, providing a new approach for related research.

2. MATERIALS AND METHODS

2.1 YOLOv7 algorithm design with CBAM integration

With the rapid development of modern agriculture, intelligent algorithms provide new technical paths for tobacco leaf detection (*Zhu, Zhang and Wang, 2025*). Traditional detection methods have limitations such as strong dependence on manual labor, low efficiency, high subjectivity, and high costs. To address these issues, this study applies the YOLOv7 algorithm to tobacco leaf detection, further improving detection capability in complex field environments and reducing computational load (*Zhai, Yu and Xia, 2023*). The YOLOv7 algorithm efficiently recognizes tobacco leaf shapes and postures, enhancing detection efficiency and generalization ability, with strong adaptability and high detection performance. The workflow of the YOLOv7 algorithm is shown in Figure 1.

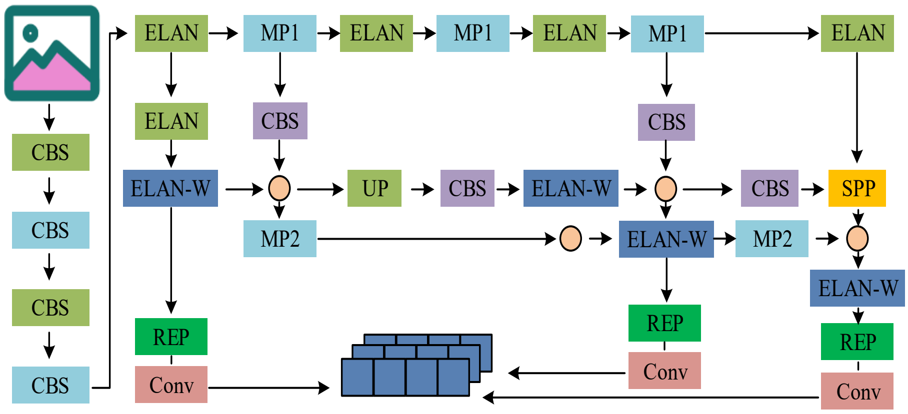


Fig. 1 - Operational flowchart of YOLOv7 algorithm

As shown in Figure 1, the CBS module in YOLOv7 consists of convolutional layers, batch normalization layers, and SILU activation functions. The CBS module efficiently extracts tobacco leaf feature information and improves convergence speed. Subsequently, the data is introduced into the efficient layer aggregation network module, the number of channels is changed, and the features are overlaid to capture the multi-level feature information of tobacco leaves. Extract the most significant features using downsampling operations, smooth the features and reduce noise during training, and perform weighted fusion on the matrix during inference. Finally, spatial pyramid pooling is utilized to enhance the receptive field, effectively processing leaf scale changes at different growth stages and shooting distances, and processing the data to reduce data volume, ultimately outputting characteristic data of tobacco leaves. The mathematical representation of the convolution operation was provided in Equation (1).

$$x_i^l = f(W_i^l * X^{(l-1)} + b_i^l) \tag{1}$$

In Equation (1), x_i^l is the i-th feature of the l layer, W_i^l is the weight matrix, f is the activation function, X is the output, and b_i^l is the bias term. Deep features of tobacco leaves are extracted through convolution and pooling operations. The extracted features are then processed by fully connected layers to further integrate the feature information and reduce information loss. Finally, the Softmax function is used to classify the tobacco leaf features. The probability output of each neuron is calculated as shown in Equation (2)

$$p(y_j) = \frac{\exp(y_j)}{\sum_{k=1}^{m} \exp(y_j)} \tag{2}$$

In Equation (2), $p(y_j)$ represents the Softmax output, $\exp(y_j)$ is the output, and m is the class value. The classification features extracted from tobacco leaf images by the convolutional network are then fed into the LSTM network. The LSTM dynamically adjusts the feature representation based on real-time input and output, selectively retaining historical information related to the evolution of tobacco leaf quality, thereby enabling effective control of neural information flow. The operation of the forget gate n_t is shown in Equation (3).

$$n_t = \sigma(\omega_n h_{t-1} + \omega_n x_t + b_n) \tag{3}$$

In Equation (3), σ represents the sigmoid function, ω is the weight matrix, h_{t-1} is the neuron output, and x_t is the input data at time t. The input gate controls information flow, filtering new input information and controlling neuron and cell state outputs. The output gate y_t operation is shown in Equation (4).

$$y_t = \sigma(\omega_t h_{t-1} + \omega_t x_t + b_t) * \tanh C_t \tag{4}$$

In Equation (4), $\tanh$ represents the hyperbolic tangent function, and C_t represents the cell state. YOLOv7 captures multi-level feature information to achieve accurate tobacco leaf detection, but in complex environments, it may lose key feature information, reducing feature extraction efficiency. The CBAM attention mechanism enhances feature maps by weighting features from different dimensions (*Guntara, 2023*). Therefore, this study introduces CBAM to improve YOLOv7's incomplete feature extraction. The CBAM workflow is shown in Figure 2.

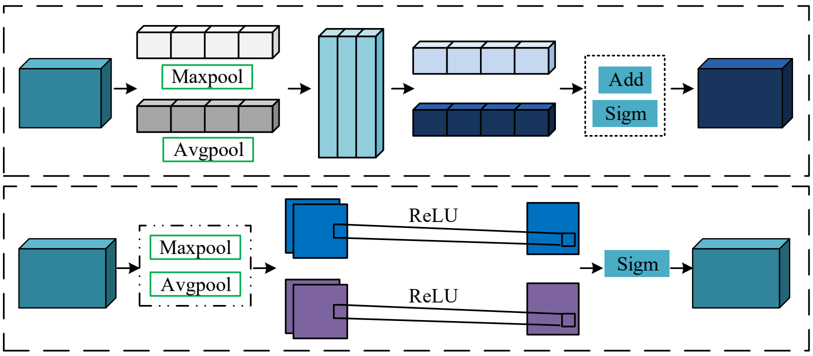


Fig. 2 - Operational flowchart of CBAM

As shown in Figure 2, the CBAM attention mechanism is divided into two modules: channel attention and spatial attention. Input the feature information into the channel attention module, adjust the weight of the original information in the channel direction, and enhance the response to key features such as leaf color, texture and disease. By utilizing global pooling to compress the information in the channel space, the obtained tobacco leaf feature information is fed into a shared neural network calculation to obtain attention weights. The expression of key features is improved by processing the weights. The spatial attention module generates a two-dimensional image by compressing the dimensions of the feature map, convolves the spatial information in the two-dimensional image, and finally weights the feature information to highlight leaf regions, lesions, or key areas of maturity. The channel attention is mathematically expressed in Equation (5).

$$M_c(F) = \sigma(MLP(avgpool(F)) + MLP(\max pool(F))) \tag{5}$$

In Equation (5), M_c denotes the output channel attention vector, and σ is the activation function. The obtained weights are applied to the channel attention, and the weighted channel attention is expressed in Equation (6).

$$M_{c1}(F) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{\max}^c))) \tag{6}$$

In Equation (6), M_{c1} represents the weighted channel attention vector, W_0 and W_1 represent different layer weights, and F_{avg}^c and $F_{\max}^c$ represent features after average pooling and max pooling. Spatial attention was applied to compress the channel-space dimension, with the channel attention calculation detailed in Equation (7).

$$M_s(F) = \sigma(f^{7\times7}([avgpool(F); \max pool(F)])) \tag{7}$$

In Equation (7), M_s represents the spatial attention vector, and $f^{7\times7}$ represents the convolution kernel after feature fusion. Process the spatial position information in the weights and adjust the expression of tobacco leaf features. The weighted spatial attention is shown in Equation (8).

$$M_{s1}(F) = \sigma(f^{7\times7}([F_{avg}^s; F_{\max}^s])) \tag{8}$$

In Equation (8), M_{s1} represents the weighted channel attention vector, and F^s_{avg} and F^s_{max} represent features after average pooling and max pooling. Finally, CBAM is integrated with YOLOv7 to form the fused algorithm CBAM-YOLOv7, as shown in Figure 3.

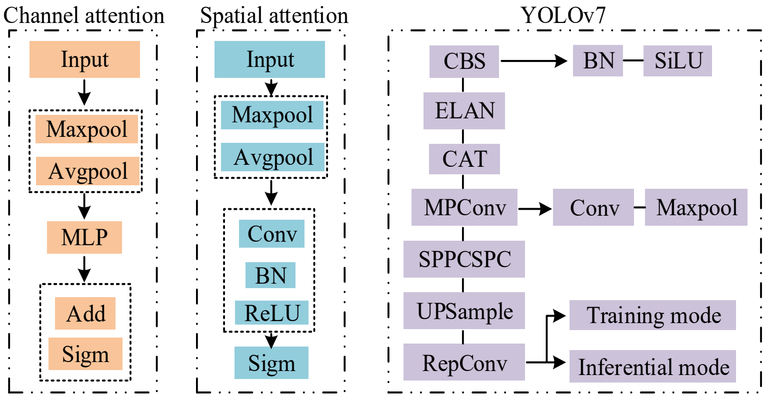


Fig. 3 - Operational flowchart of CBAM-YOLOv7

In Figure 3, the CBAM-YOLOv7 algorithm quickly extracts shallow features such as edges and textures of tobacco leaves through the CBS module, and captures multi-level feature information of tobacco leaves using an efficient layer aggregation network module. Subsequently, the CBAM module is introduced, which first utilizes global pooling to compress the information in the channel space, and introduces a shared neural network to enhance the key channel features of the blades. The spatial attention is compressed along the channel and convolutionally weighted to highlight key positions such as leaf regions and lesions. Extracting the most significant features of tobacco leaves through downsampling operation, while enhancing the receptive field through spatial pyramid pooling to reduce data volume, and finally outputting the feature representation of tobacco leaves.

2.2 Tobacco leaf detection model based on YOLOv7 and MobileNetV3

The CBAM-YOLOv7 improves the saliency of tobacco leaf feature extraction and reduces the loss of multi-level features, but its structure is complex, and the computation cost is high. This complexity often reduces the accuracy of information recognition and detection, which presents some limitations (*Ndovie and Masabo, 2024, Sriwiboon and Phimphisan, 2025*). To address the problems of excessive parameters and computational complexity in YOLOv7, this study introduces MobileNetV3 to replace the backbone network of YOLOv7 to improve computational efficiency. On this basis, the traditional convolution layer is removed, and DCN improved by SA is introduced to dynamically adjust the sampling positions. A tobacco leaf detection model based on YOLOv7 and MobileNetV3 (YOLOv7-MNDCN) is finally constructed. The improved DCN can adaptively learn complex geometric structures and enhance the recognition capability of the model. The ReLU activation function is defined as shown in Equation (9).

$$f(z) = \begin{cases} z & if\ z \geq 0 \\ az & otherwise \end{cases} \tag{9}$$

In Equation (9), a denotes the constant value 0.01. The detailed process of the improved DCN is shown in Figure 4.

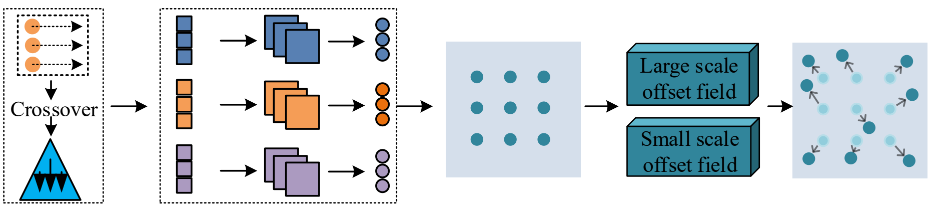


Fig. 4 - Operational flowchart of improved DCN

As shown in Figure 4, the SA algorithm updates parameter values through initialization, generates optimal parameter values through cross operation and temperature reduction processing, avoids module mixing and information decomposition errors, and improves the model's ability to identify key blade features. Subsequently, the SA processed data is introduced into the improved DCN. When the convolutional layer inputs the feature map, a set of offsets is generated to adjust the pixel positions of the feature map.

By using learnable offset parameters, the convolutional kernel can adaptively fit irregular shapes such as leaf edges and lesion areas, and expand the sampling coverage area. Subsequently, the distribution of sampling points in the convolution was further adjusted to make the tobacco detection model have geometric adaptive characteristics. The mathematical expression of the sampling position is shown in Equation (10).

$$R = \{(-1,-1),(-1,0),...,(0,1),(1,1)\} \tag{10}$$

In Equation (10), R denotes the sampling grid. After determining the receptive field size, a coordinate offset is added to the pixels, and the feature information is weighted through sampling to output the standard feature value. The mathematical expression of the feature value is shown in Equation (11).

$$y(p_0) = \sum p_{n\in R} w(p_n) \cdot x(p_0 + p_n) \tag{11}$$

In Equation (11), $y(p_0)$ denotes the feature output, p_n denotes the n-th offset, w denotes the learnable weight parameter, and x denotes the input feature. DCN adds the offset to the traditional feature computation, allowing dynamic adjustment of the sampling positions. The mathematical expression of the improved output feature value is shown in Equation (12).

$$y(p_{0a}) = \sum p_{n\in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \tag{12}$$

In Equation (12), Δp_n denotes the n-th offset parameter. The improved DCN is integrated into the MobileNetV3 to efficiently handle flexible targets and improve recognition and detection ability. The structure of the improved DCN fused with MobileNetV3 is shown in Figure 5.

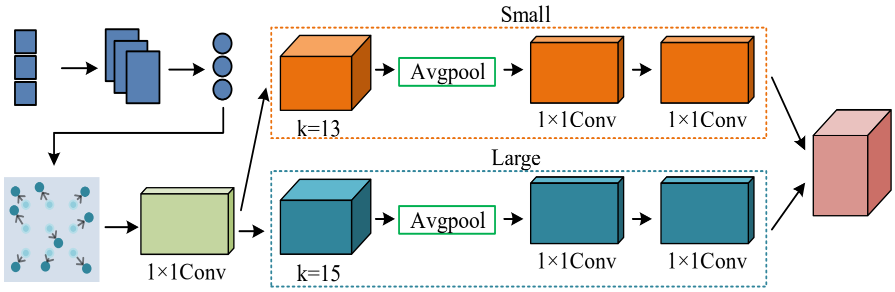


Fig. 5 - DCN and MobileNetv3 fusion network structure diagram

As shown in Figure 5, the improved fusion network of DCN and MobileNetv3 first generates learnable offsets through DCN, dynamically adjusts the sampling point positions, and enables the convolutional kernel to adaptively fit the leaf edges and lesion areas. Subsequently, the adjusted tobacco leaf feature map data was input into the MobileNetv3 network, and feature parameters were extracted through depthwise separable convolution. The multi-scale convolution was integrated into two versions, with the Small version set to 13 bottlenecks and the Large version set to 13 bottlenecks. The two versions were adapted to different scenarios. Finally, the information is compressed through average pooling and reduced in dimensionality through two 1x1 convolutional layers to output a lightweight feature representation of tobacco leaves. The operation equation of depthwise separable convolution is shown in Equation (13).

$$F_a = K^2 \times N \times W \times H + M \times N \times W \times H \tag{13}$$

Equation (13) defines K and N as the convolution kernel height and width, W and H as the feature map height and width, and M as the channel count. To enhance the nonlinear expression of the algorithm, this study introduces the Hard-Swish activation function. The mathematical equation of the Swish function is shown in Equation (14).

$$Swish(x) = x\sigma(\beta x) \tag{14}$$

In Equation (14), β denotes the learnable parameter. The enhanced Hard-Swish activation function, derived from the Swish function, was expressed in Equation (15).

$$Hard - Swish(x) = x\frac{ReLU6(x+3)}{6} \tag{15}$$

In Equation (15), $ReLU6$ denotes the variant of the $ReLU$ function. The improved DCN fused with MobileNetV3 can improve detection accuracy, reduce computation, and optimize the overall structure, while the improved YOLOv7 enhances the representation of feature information and regions, improving feature extraction ability. Therefore, the study will introduce the improved fusion network of DCN and MobileNetv3 into the improved YOLOv7 algorithm to construct a YOLOv7-MNDCN tobacco detection model. The operation process of the model is shown in Figure 6.

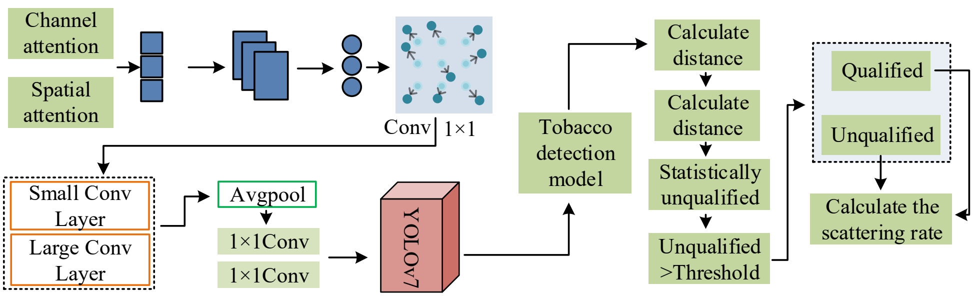


Fig. 6 - Operational flowchart of YOLOv7-MNDCN model

As shown in Figure 6, leaf texture and edge features are extracted through the CBS module, and multi-level semantic information is captured through an efficient layer aggregation network. Introducing the CBAM module, utilizing channel and spatial attention to enhance the response of key channels and lesion areas in leaves. Extract the most significant features using downsampling operation, smooth the features and reduce noise, and perform weighted fusion on the matrix. Subsequently, the pixel positions of the feature map are dynamically adjusted through offset to improve the distribution of sampling points in the convolution. Input the adjusted tobacco leaf data into the MobileNetv3 network, and use depthwise separable convolution to efficiently extract multi-scale features, and adapt to different detection scenarios using Small and Large versions. Finally, the calculation size is reduced through average pooling and two consecutive 1x1 convolutional layers, and the characteristics of tobacco leaves are outputted.

3. RESULTS

Research and collect a dataset of tobacco leaf images from Luozhuang Tobacco Station in Weifang City, Shandong Province in 2024. The variety is the locally cultivated tobacco variety NC55. On a sunny morning, simulated actual working lighting conditions were used for image acquisition, resulting in a total of 5000 tobacco leaf images with a size of 1.8TB. The image resolution is 6000 × 4000 pixels and the storage format is JPG. The dataset is randomly divided into training set, validation set, and testing set in a ratio of 6:2:2, and all images are manually annotated by experts. The study classified leaves into three categories based on maturity: immature, properly mature, and over-mature. As shown in Figure 7.

(a) Immature (b) Properly mature (c) Over-mature

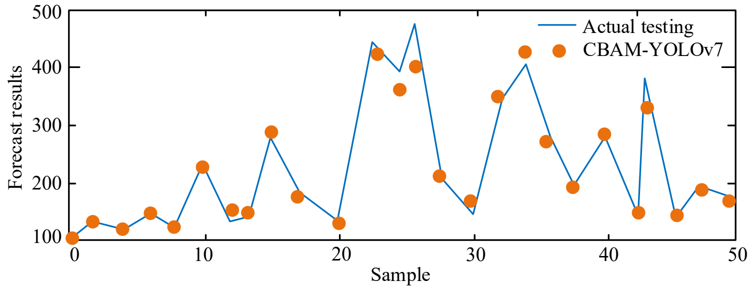
Fig. 7 - Tobacco leaf feature map

The research is based on the improved YOLOv7-MNDCN fusion model, and the trained model is deployed on edge computing devices or mobile terminal platforms. In practical applications, images of tobacco leaves are obtained through image acquisition devices, input into the model for end-to-end detection, and output the position, category, and confidence information of the leaves.

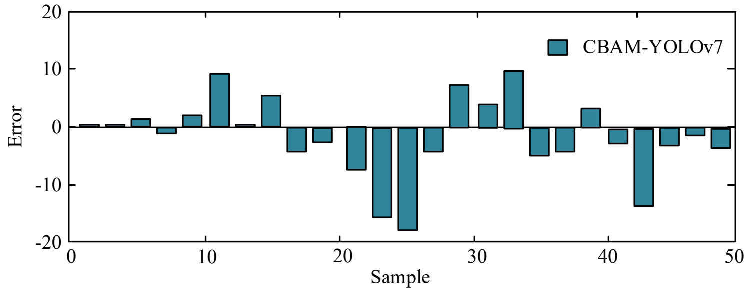
The research aims to achieve precise and rapid identification of tobacco leaves, assist in maturity judgment, reduce manual inspection costs, and improve the scientific level of tobacco quality control. By constructing a lightweight detection model and providing real-time feedback of recognition results to terminal devices, data support is provided for agricultural decision-making such as tobacco cultivation.

3.1 Effectiveness validation of CBAM-YOLOv7

To examine the efficacy of the proposed algorithm, it was compared with the Multi-Marker Similarity Assessment (MMSim) algorithm, the Gorilla Troops Optimizer (GTO), and the Bidirectional Long Short-Term Memory (BiLSTM) algorithm. The experiments were conducted on Windows 11 and Ubuntu 18.04 operating systems, with an Intel Core i7-5500 processor and 8 GB of RAM. Python 3.9 was used as the programming language, and PyTorch served as the deep learning framework. The algorithm employed a learning rate of 0.001, a weight decay of 0.0005, and ran for 100 iterations. To evaluate the prediction performance and stability of the CBAM-YOLOv7, its prediction results and errors were compared with the ground truth, as shown in Figure 8.



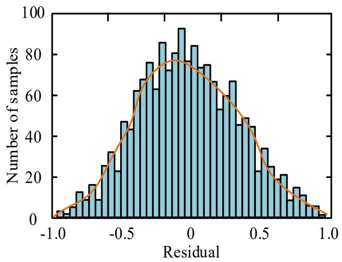
(a) CBAM-YOLOv7 algorithm prediction result chart



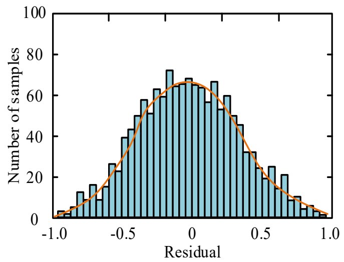
(b) CBAM-YOLOv7 algorithm prediction error result chart

Fig. 8 - Comparison of the prediction and error results of the true value

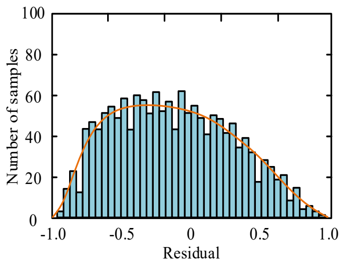
As shown in Figure 8(a), the prediction results of the CBAM-YOLOv7 were similar to the ground truth. The predictions matched closely for samples 0–10, and as the sample size increased, some deviation occurred. The overall prediction accuracy reached 95.23%, which was likely due to the attention mechanism enhancing feature learning and improving accuracy. As shown in Figure 8(b), the prediction error averaged 2 for 0–10 samples, peaked at an average of 28 for 20–30 samples, and had an overall mean error of 11. The peak error between 20–30 samples was likely caused by fluctuations from the rapid increase in sample size, after which the high adaptability of the algorithm led to stabilization. These results indicated that the algorithm achieved precise prediction performance and stable operation. To further validate the fitting performance of the CBAM-YOLOv7, its residuals were compared with MMSim, GTO, and BiLSTM, as shown in Figure 9.



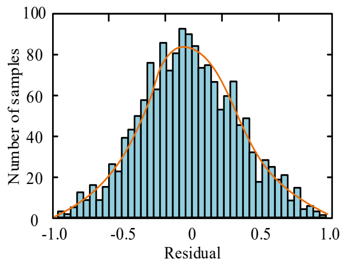
(a) Residual result graph of MMSim algorithm



(b) Residual result graph of GTO algorithm



(c) Residual result graph of BiLSTM algorithm



(d) Residual result graph of CBAM-YOLOv7 algorithm

Fig. 9 - Comparison of fitting effects of four algorithms

As shown in Figures 9(a), (b), and (c), the residual curves of MMSim, GTO, and BiLSTM showed varying degrees of deviation. MMSim and BiLSTM exhibited significant irregular changes, with fitting values of 0.4254 and 0.2156, respectively. Although GTO presented a normal distribution, its trend was not significant, with a fitting value of 0.5021. As shown in Figure 9(d), the proposed algorithm exhibited a clear normal distribution, with a fitting value of 0.8514, which was significantly higher than those of MMSim, GTO, and BiLSTM. This improvement was likely attributed to the ELAN module, which expanded operational width through group merging, enhancing the algorithm's reliability. These findings indicated that the algorithm achieved a good fitting degree and demonstrated superior reliability. To further evaluate its accuracy and loss, the algorithm was compared with MMSim, GTO, and BiLSTM, as shown in Figure 10.

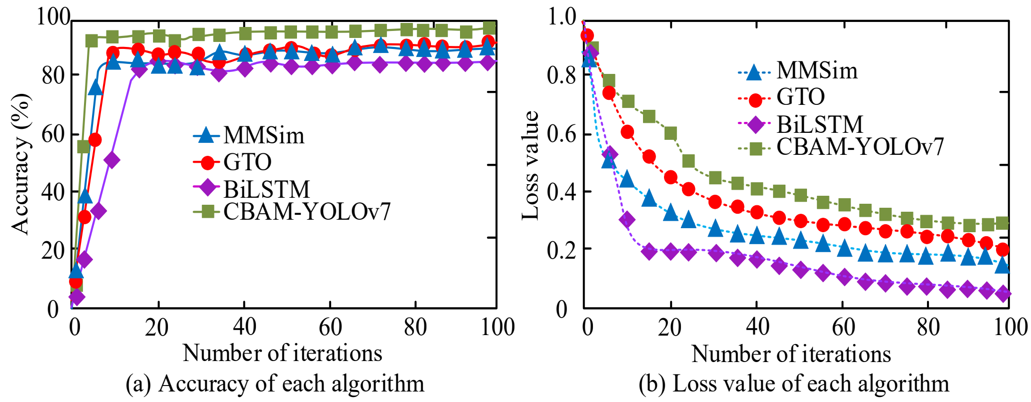


Fig. 10 - Comparison of accuracy and loss test results

As shown in Figure 10(a), the accuracy of the CBAM-YOLOv7 increased sharply within 0–15 iterations, reaching 95.39%, after which the growth became slower. Between 50–150 iterations, the accuracy fluctuated slightly and ultimately reached 97.01% after 500 iterations. As shown in Figure 10(b), the loss was 0.15 at 100 iterations and gradually decreased to 0.09 within 100–300 iterations, after which the trend stabilized. The accuracy and loss performance of the algorithm were better than those of the other three algorithms, indicating that CBAM optimization effectively addressed small-object missed detection, with faster and more stable convergence. This demonstrated that the algorithm had good optimization performance.

3.2 Application validation of the tobacco leaf detection model

After validating the performance of the CBAM-YOLOv7, the study further evaluated the practical application of the tobacco detection model based on the YOLOv7-MNDCN fusion algorithm by comparing it with MMSim, GTO, and BiLSTM models. The experiments were conducted on the AutoDL cloud platform using SGD as the optimizer. During the experiments, the learning rate was set to 0.001, the number of iterations to 100, the input image size to 640×640, the batch size to 8, the weight decay coefficient to 0.005, and the translation range to 0.1. To evaluate the mean Average Precision (mAP) at Intersection over Union (IoU) thresholds of 0.5 and 0.95, the model was compared with MMSim, GTO, and BiLSTM, and the test results are shown in Figure 11.

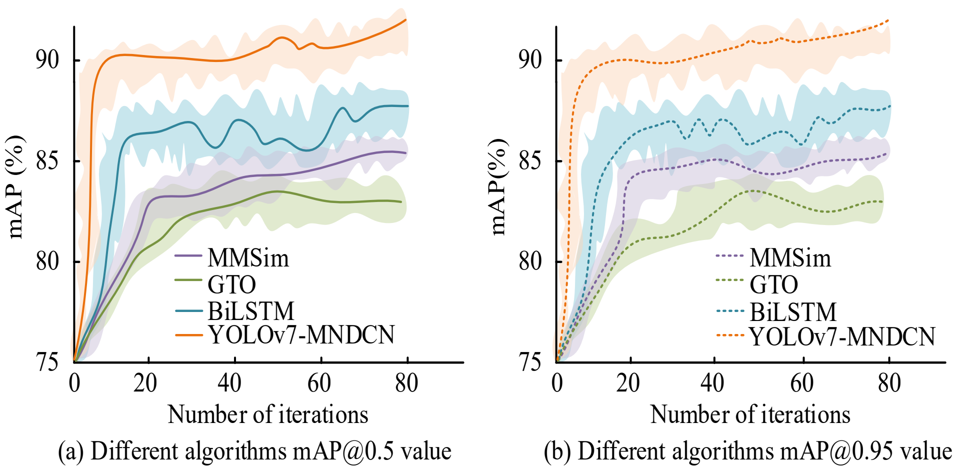


Fig. 11 - Comparison of mAP test results under different IoU

In Figure 11(a), at IoU=0.5, the mAP of the YOLOv7-MNDCN model increased sharply within 0–10 iterations, reaching 90.39%, and then grew more slowly. Between 20–80 iterations, the accuracy fluctuated slightly, and after 80 iterations, the precision reached 98.11%, which was an improvement of 10.54%, 14.45%, and 16.25% over MMSim, GTO, and BiLSTM, respectively. As shown in Figure 11(b), at IoU=0.95, the mAP also increased rapidly to 89.47% as the number of iterations increased, then slowed between 15–80 iterations, and finally reached 97.66% after 80 iterations, outperforming all three comparison models. This improvement was likely due to the integration of the MobileNetv3 network, which significantly reduced computational parameters and prevented missed detections of feature information. These results indicated that the model achieved high detection accuracy for tobacco leaves. To further evaluate its overall performance, the model was compared with MMSim, GTO, and BiLSTM, as shown in Figure 12.

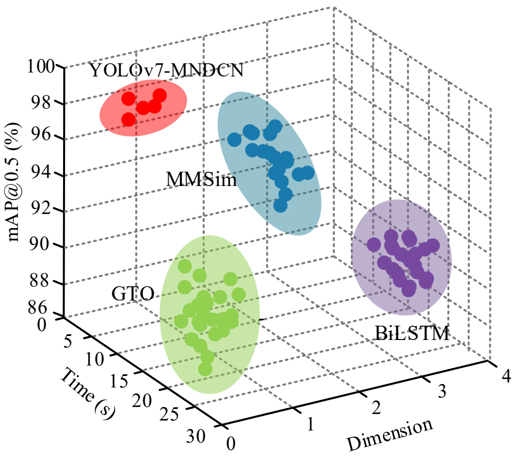


Fig. 12 - Comparison of overall performance test results

As shown in Figure 12, the number of scatter points represented computational cost, and the coverage area represented overall performance. The YOLOv7-MNDCN model achieved an mAP of 97.85% at a runtime of 5 s, outperforming MMSim, GTO, and BiLSTM. The model also had the lowest computational cost and the shortest runtime, indicating high overall performance. This was likely because the MobileNetv3 network optimized parameters through grouped convolution and limited the generation of large convolution kernels, improving operational efficiency and reducing computational complexity. These findings indicated that the model achieved lightweight performance and enhanced processing efficiency. To further analyze its detection performance, the model was compared with MMSim, GTO, and BiLSTM in terms of parameter size, accuracy, recall, F1 score, model size, and runtime, as shown in Table 1.

Table 1

Comparison of test results of multiple comprehensive indicators

Model	Parameters (M)	Accuracy (%)	Recall (%)	F1 score (%)	Dimension (MB)	Time (s)
MMSim	36.24	91.02	88.40	89.58	89.74	36.58
GTO	21.05	84.25	81.25	82.24	59.45	39.78
BiLSTM	13.69	89.23	86.55	88.65	77.14	25.44
YOLOv7-MNDCN	9.36	94.65	91.24	93.54	50.69	21.05

As shown in Table 1, the YOLOv7-MNDCN model achieved a tobacco detection accuracy of 94.65%, which was an improvement of 3.63%, 10.40%, and 5.42% over MMSim, GTO, and BiLSTM, respectively. Its recall and F1 score were 91.24% and 93.54%, both significantly higher than those of the other three models. The model parameters and size were only 9.36 M and 50.69 MB, indicating that the integration of the MobileNetv3 network optimized the model structure and reduced parameter computation through dynamic sampling adjustment. In terms of runtime, the model achieved 21.05 s, which was notably shorter than that of the other models. These results indicated that the model provided superior performance in tobacco detection.

4. CONCLUSIONS

To address the problems of high manual dependency and low detection efficiency in tobacco leaf detection, this paper built a tobacco leaf detection model based on the fusion of YOLOv7 and MobileNetV3. The model fused CBAM with YOLOv7 to enhance the representation of key features and regions and to improve feature extraction efficiency. MobileNetV3 replaced the backbone network of YOLOv7 to increase computational efficiency. On this basis, the traditional convolutional layers were removed, and an improved DCN was introduced to dynamically adjust sampling positions. As a result, a tobacco leaf detection model based on YOLOv7 and MobileNetV3 was constructed. Experimental results showed that the CBAM-YOLOv7 had superior prediction and fitting performance, with an accuracy of 97.01% and a loss value of 0.09. Its performance was better than MMSim, GTO, and BiLSTM. The final YOLOv7-MNDCN tobacco leaf detection model achieved a mAP of 98.11% at a 50% overlap threshold and 97.66% at a 95% overlap threshold, both surpassing the other three models. Overall, the model based on the fusion of YOLOv7 and MobileNetV3 efficiently extracted feature information, optimized the model structure, reduced the number of parameters and computation, and improved the accuracy of tobacco leaf detection. The experiment was only conducted in indoor environments and did not include outdoor conditions. Factors such as lighting variation and object interference were not tested. In the future, the model will be further studied in complex outdoor environments to evaluate its robustness in tobacco leaf detection.

STATEMENTS AND DECLARATIONS

Fundings

None.

Competing Interests

Author declares no conflict of interest.

Author Contributions Statement

All authors have agreed to publish this article.

REFERENCES

[1] Acikgoz H. (2024). An automatic detection model for cracks in photovoltaic cells based on electroluminescence imaging using improved YOLOv7. *Signal, Image and Video Processing*, 18(1): 625-635.

[2] Bitzer Z T, Mocniak L E, Trushin N, Smith M, Richie J P. (2024). Influence of tobacco variety and curing on free radical production in cigarette smoke. *Nicotine and Tobacco Research*, 25(7): 1400-1405.

[3] Brock S. (2023). Enhancing Steel Surface Defect Detection: A YOLOv8-DCNv3-FOCAL Integrated Approach. *Journal of Computer Science and Software Applications*, 4(5): 16-21.

[4] Eği Y. (2023). YOLO V7 and computer vision-based mask-wearing warning system for congested public areas. *Journal of the Institute of Science and Technology*, 13(1): 22-32.

[5] Guntara R G. (2023). Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(1): 55-60.

[6] Kim E H, Wang Z, Zong H, Jiang Z, Fu Z, Pedrycz W. (2023). Design of tobacco leaves classifier through fuzzy clustering-based neural networks with multiple histogram analyses of images. *IEEE Transactions on Industrial Informatics*, 20(3): 4698-4709.

[7] Liu F, Shen Y L, Zhan S W, Yang Y, Mou Y. (2023). Identification of Quality Characteristics of Flue-Cured Tobacco Based on Raman Spectroscopy. Journal of Applied Spectroscopy, 90(1): 108-115.

[8] Mammeri S, Amroune M, Haouam M Y, Bendib I, Silva A C. (2024). Early detection and diagnosis of lung cancer using YOLO v7, and transfer learning. *Multimedia Tools and Applications*, 83(10): 30965-30980.

[9] Moşteanu N R. (2025). Thriving in the entrepreneurial landscape of sustainability and intelligent automation era. *Green and Low-Carbon Economy*, 3(1): 18-25.

[10] Ndovie L K, Masabo E. (2024). Leveraging MobileNetV3 for In-Field Tomato Disease Detection in Malawi via CNN. *SAIEE Africa Research Journal*, 115(3): 74-85.

[11] Pramudhita D A, Azzahra F, Arfat I K, Magdalena R, Saidah S. (2023). Strawberry plant diseases classification using CNN based on MobileNetV3-Large and EfficientNet-B0 architecture. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, 9(3): 522-534.

[12] Prasad S B R, Chandana B S. (2023). Mobilenetv3: a deep learning technique for human face expressions identification. *International journal of information technology*, 15(6): 3229-3243.

[13] Shen J, Ye H, Tang C, Chen Y. (2024). InsuDet: a lightweight insulator defect detection algorithm based on YOLOv8. *International Journal of Advanced Mechatronic Systems*, 11(4): 179-191.

[14] Sriwiboon N, Phimphisan S. (2025). Efficient COVID-19 Detection using Optimized MobileNetV3-Small with SRGAN for Web Application. *Engineering, Technology & Applied Science Research*, 15(2): 20953-20958.

[15] Thimmegowda T G M, Jayaramaiah C. (2023). Cluster-based segmentation for tobacco plant detection and classification. *Bulletin of Electrical Engineering and Informatics*, 12(1): 75-85.

[16] Wang Y, Qin L. (2022). Research on state prediction method of tobacco curing process based on model fusion. *Journal of Ambient Intelligence and Humanized Computing*, 13(6): 2951-2961.

[17] Wei K, Bin J, Wang F, Kang C. (2022). On-line monitoring of the tobacco leaf composition during flue-curing by near-infrared spectroscopy and deep transfer learning. *Analytical Letters*, 55(13): 2089-2107.

[18] Yonglin C, Hengtao W, Shang Z. (2024). Lightweight Infrared Target Detection Algorithm Based on YOLO v7. *Infrared technology*, 46(12): 1380-1389.

[19] Zhai D H, Yu S, Xia Y. (2023). FANet: Fast and accurate robotic grasp detection based on keypoints. *IEEE Transactions on Automation Science and Engineering*, 21(3): 2974-2986.

[20] Zhu Z, Zhang J, Wang J. (2025). A prior knowledge-embedded deformable convolutional network for vision-based fabric defect detection. *Textile Research Journal*, 95(7-8): 783-795.