ISSN 2616-5775 Vol. 7, Issue 2: 72-78, DOI: 10.25236/AJCIS.2024.070210

A Lightweight Tomato Leaf Disease Detection System with Strong Generalization Ability Based on TMT-YOLOv5s

Yueyang Wu[1], Xinyao Zhang[2], Minhua Ye[3], Guiyuan Zhang[1], Xiang Zhang[1], Yize Tang[1], Xuewen Chen[1], Yongpei Cao[1], Derong Yang[1], Zhi Li[1,*]

[1]*School of Mathematics and Computer, Guangdong Ocean University, Zhanjiang, 524088, China*
[2]*College of Food Science and Technology, Guangdong Ocean University, Zhanjiang, 524088, China*
[3]*College of Ocean Engineering and Energy, Guangdong Ocean University, Zhanjiang, 524088, China*
**Corresponding author*

Abstract: *To address the issue of low accuracy and poor performance of existing crop disease detection methods in identifying various tomato leaf diseases, this study proposes an improved tomato leaf disease detection model, TMT-YOLOv5s, based on the YOLOv5 network model. Initially, a DCAM attention mechanism module is constructed within the Backbone of the original YOLOv5s model. This module enhances the model's capability to extract pathological features of tomato leaves by implementing dual-channel attention and spatial attention mechanisms and integrates with the BiFPN module to mitigate the influence of complex background features. The aim is to improve the model's accuracy and precision in detecting and classifying different types of diseases. Experimental results demonstrate that the TMT-YOLOv5s model achieves an average precision of 98.59% and a recall rate of 96.31%, marking an improvement of 2.79 percentage points and 2.51 percentage points, respectively, compared to the original YOLOv5s model. The model exhibits superior detection accuracy and effectiveness, accurately identifying various types of tomato leaf diseases. These findings provide valuable insights for the practical application of tomato leaf disease detection.*

Keywords: *Deep learning, YOLOv5s, Convolutional Neural Network, Disease detection, Tomato leaf diseases*

1. Introduction

Tomato, as one of the vegetables with the largest cultivation area and highest consumption worldwide, is prone to yield reduction due to susceptibility to diseases and pests during its growth [1]. Currently, the control of tomato diseases and pests mainly relies on manual methods, which are time-consuming, labor-intensive, and often suffer from delayed detection. Therefore, how to utilize intelligent pest detection technology, especially machine vision detection technology based on deep learning, has become one of the key focuses of research on disease and pest control among numerous scholars.

Methods for machine vision detection based on deep learning mainly include two-stage detection represented by Region-based Convolutional Neural Network (R-CNN), single-stage detection represented by the YOLO model, and the SSD model [2]. With the rapid development of machine vision, various machine vision detection algorithms have been widely applied in agricultural pest detection [3]. LIU et al. [4] proposed a pest detection method based on the YOLO convolutional neural network, but its average accuracy for tomato diseases and pests detection only reached 85.09%, indicating relatively low precision. WANG et al. [5], based on the YOLOv4 model, employed GhostNet as the backbone network and introduced new feature fusion methods and attention mechanisms for detecting lychee diseases and pests in complex natural environments, achieving an average detection accuracy of 95.31% on the training set, albeit with a larger model size. LIU et al. [6] proposed a YOLO-SL model for grape leaf disease and pest detection to meet the orchard's requirements, although the model's average detection accuracy reached 90.4%, its detection speed was slow and the model size was large.

Therefore, in this study, based on the YOLOv5s model, the DCAM attention mechanism is introduced to enhance the model's extraction of pathological features of tomato leaves, improving the classification accuracy of different types of diseases. Then, the C3STR module, fused with the Swin Transformer [7], is applied to strengthen the model's multi-scale modeling capability, enhancing the detection effect of small-

ISSN 2616-5775 Vol. 7, Issue 2: 72-78, DOI: 10.25236/AJCIS.2024.070210

scale diseases on leaves. Finally, the Weighted Bi-directional Feature Pyramid Network (BiF-PN) structure is incorporated to achieve bi-directional cross-scale connections for feature map weighted fusion, improving the learning efficiency of the network for disease features. This achieves an increase in model detection accuracy while adding only a small amount to the model size.

2. Improved Detection Model Based on YOLOv5s

2.1. YOLOv5 Object Detection Algorithm

YOLOv5 is a single-stage object detection algorithm composed of four parts: Input, Backbone, Neck, and Head. Its Input utilizes Mosaic data augmentation to enhance the training speed and network accuracy and employs an adaptive anchor box calculation and adaptive image scaling algorithm. The Backbone aggregates and constructs image features at different granularities. The Neck utilizes a Feature Pyramid Network (FPN) + Path Aggregation Network (PAN) structure to fuse image features, with FPN layer upsampling semantic features from top to bottom and PANet downsampling localization features from bottom to top. The Head generates predicted bounding boxes and predicted categories based on image features.

2.2. Constructing a Model for Tomato Leaf Disease Detection

2.2.1. DCAM Attention Mechanism

Due to the limited capability of the original YOLOv5s model in extracting features related to tomato leaf diseases, resulting in issues such as susceptibility to background interference and poor classification performance, this study proposes a Dual Channel Attention Module (DCAM) attention mechanism module. This module calculates the similarity between target features and features from other regions through the channel attention mechanism and spatial attention mechanism to generate target attention maps, thereby enhancing the model's focus on target features. It also computes the similarity between background features and features from other regions to generate background attention maps, mitigating the influence of background features on the model, thus effectively improving the model's feature extraction and classification capabilities. The structural diagram of the DCAM attention mechanism is shown in Figure 1.

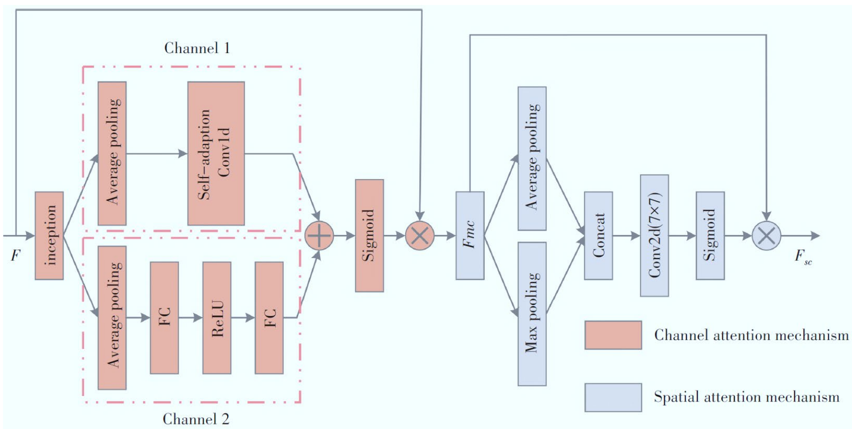


Figure 1: Structure of DCAM attention mechanism

2.2.2. BiFPN Weighted Bi-directional Feature Pyramid Network

Due to the original YOLOv5s model's utilization of uniform weights for merging feature maps of different scales, it exhibits lower accuracy in detecting various types of tomato leaf diseases. To address this issue, this study employs a Weighted Bi-directional Feature Pyramid Network (BiFPN) network structure to replace the original PANet network structure, enhancing the network's feature fusion capability through cross-scale feature fusion and learnable weight fusion across different hierarchical features.

The BiFPN (bidirectional feature pyramid network) network simplifies the network structure by removing nodes with only one input edge from the PANet network and adding an edge between input and output nodes at the same level to achieve feature fusion of different sizes.

ISSN 2616-5775 Vol. 7, Issue 2: 72-78, DOI: 10.25236/AJCIS.2024.070210

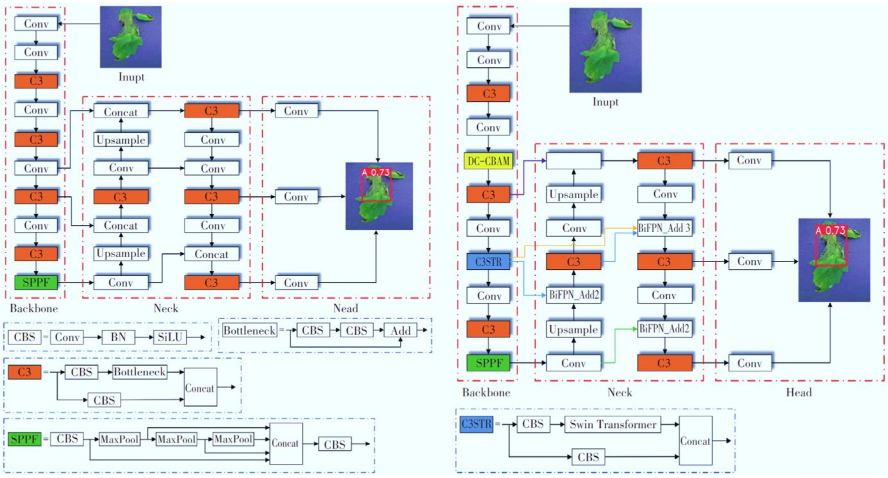
While the PANet network employs a balanced feature fusion mechanism to learn different input features, the BiFPN network constructs a weighted feature fusion mechanism to learn different input features and normalize the fused feature O, as shown in Equation (1). By continuously updating the weights through feature learning, the network acquires more feature information.

$$O = \sum_i \frac{w_i \cdot I_i}{\varepsilon + \sum_j w_j} \tag{1}$$

In the equation: w_i represents the learning weight of each layer in the network; w_j represents the learning weight of the j-th layer; I_i denotes the input feature, and $\varepsilon = 0.0001$.

2.2.3. Improved Network Model

In this study, to address the issues of low detection accuracy and poor performance of the original YOLOv5s model in detecting features of small-sized tomato leaf diseases, a DCAM (Dual Channel Attention Module) attention mechanism module is constructed at the 4th layer of the network. This aims to alleviate the overall influence of irrelevant information such as background features on network learning, thereby enabling the transmission of more tomato leaf disease feature information to the next module and the Neck. Finally, the PANet network is replaced by the BiFPN network to improve the model's receptive field and detection accuracy through multi-scale feature fusion and learnable weights. This leads to the construction of the TMT-YOLOv5s model, the schematic diagram of which is depicted in Figure 2.



a) Network structure of YOLOv5 *b) Network structure of TMT-YOLOv5s*

Figure 2: Comparison of network structure between improved model and original model

3. Experimental Results and Analysis

3.1. Experimental Environment

This study employed a computer system running Windows 10, equipped with 16 GB of memory operating at 3200 MHz, an Intel®Core™i7-11800H CPU@2.30GHz processor, and an NVIDIA GeForce RTX 3050 4GB graphics card. The deep learning framework utilized was PyTorch 1.12.1, with CUDA 10.2 and CUDNN 8.4.1, operating within a Python 3.9 environment. Parameter settings included a learning rate of 0.01, momentum parameter of 0.937, final learning rate of 0.1, batch size of 16, and 400 epochs. Additionally, a warm-up learning strategy with 3 epochs, a momentum parameter of 0.8, and a learning rate of 0.1 was employed. Confidence and IOU thresholds were set to 0.25 and 0.45, respectively, and image resolutions were uniformly adjusted to 640×640.

3.2. Experimental Dataset

To validate the effectiveness of the proposed model, this study primarily conducted research, training, and testing on the AI Challenger 2018 crop disease and pest dataset. After filtering out categories with severely imbalanced image counts, a dataset comprising a total of nine classes, including eight types of tomato leaf diseases and healthy leaves, was selected. Partial examples of tomato leaf cases are displayed in Figure 3.

ISSN 2616-5775 Vol. 7, Issue 2: 72-78, DOI: 10.25236/AJCIS.2024.070210

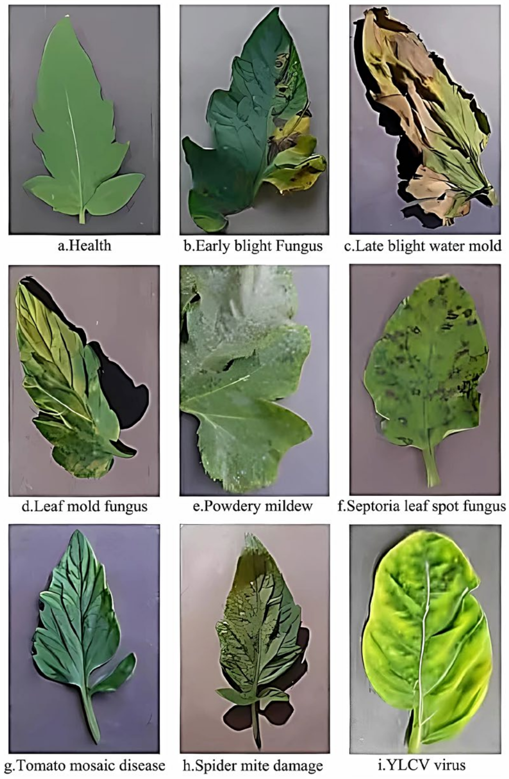


Figure 3: Tomato Leaf Example (Part)

3.3. Model Performance Evaluation Metrics

Single-label classification evaluation metrics include Accuracy (Acc), Precision (Prec), Recall (Rec), F1-score, Specificity (Spe), and confusion matrix. These evaluation metrics are calculated based on four fundamental indicators:

TP (True Positive): The number of images correctly classified as positive samples.

FN (False Negative): The number of images incorrectly classified as negative samples but are positive.

FP (False Positive): The number of images incorrectly classified as positive samples but are negative.

TN (True Negative): The number of images correctly classified as negative samples.

To objectively evaluate the model's performance in detecting tomato leaf diseases, this study employs Mean Average Precision (mAP), Recall (R), parameter count, detection rate (Frames Per Second, FPS), and model size as evaluation metrics. The formulas for mAP, R, and FPS are as follows (Equations (2) to (4)).

$$mAP = \frac{\sum AP}{N(Class)} \tag{2}$$

In the equations, "AP" represents Average Precision, and "N (Class)" denotes the number of images in the test dataset.

$$R = \frac{T_P}{T_P+F_N} \times 100\% \tag{3}$$

In the equations, "T_P" stands for True Positive, representing the number of samples correctly classified as positive by the model, and "F_N" denotes False Negative, indicating the number of samples incorrectly classified as negative by the model but are positive.

$$FPS = \frac{1000}{T_{pre}+T_i+T_N} \tag{4}$$

In the equations, "T_{pre}" refers to image preprocessing time, "T_i" represents inference time, which is the time taken for an image to go from input to output through the model after preprocessing, and "T_N" denotes post-processing time, which includes operations such as non-maximum suppression per image

ISSN 2616-5775 Vol. 7, Issue 2: 72-78, DOI: 10.25236/AJCIS.2024.070210

on the model output.

This study comprehensively evaluates the model from several aspects including mAP, R, FPS, and confusion matrix analysis.

3.4. Model Performance Experiments

To verify the detection performance of the proposed model, TMT-YOLOv5s, this study conducts comparative validations of its Mean Average Precision (mAP), Recall, parameter count, detection rate (Frames Per Second, FPS), and model size against Faster R-CNN, SSD, YOLOv5s, and YOLOv5m models on the tomato leaf disease dataset, as shown in Table 1. As indicated in Table 1, the mAP of the TMT-YOLOv5s model is 98.59%, with a Recall of 96.31%, parameter count of 7.28×106, detection rate of 67 FPS, and model size of 14.2 MB. Compared to Faster R-CNN, the mAP of the proposed model improves by 1.2 percentage points, Recall by 1 percentage point, with a reduction in parameter count by 4.89×107, an increase in FPS by 60, and a decrease in model size by 94.1 MB. Compared to SSD, the mAP improves by 2.6 percentage points, Recall by 2.1 percentage points, with a reduction in parameter count by 3.94×107, an increase in FPS by 24, and a decrease in model size by 77.4 MB. Compared to the original YOLOv5s, the mAP improves by 1.9 percentage points, Recall by 2.5 percentage points, with an increase in parameter count by 2.6×105, a decrease in FPS by 3, and an increase in model size by 0.4 MB. Compared to YOLOv5m, the mAP improves by 0.7 percentage points, Recall by 0.2 percentage points, with a reduction in parameter count by 1.35×107, an increase in FPS by 16, and a decrease in model size by 28.1 MB. Although the detection speed of the YOLOv5s model is slightly higher than that of the TMT-YOLOv5s model, the comprehensive comparison results indicate superior performance of the TMT-YOLOv5s model in terms of mAP and Recall. Despite the slight increase in parameter count and model size of the TMT-YOLOv5s model, its excellent mAP, Recall, and relatively high detection speed demonstrate its outstanding detection performance.

Table 1: Performance comparison of different models

Models	mAP	Recall	Parameters	Speed/FPS	Model size/MB
Faster R-CNN	96.52	95.31	5.62×10^{7}	7	108.3
SSD	95.14	94.16	4.67×10^{7}	43	91.6
YOLOv5	95.81	93.79	7.02×10^{6}	70	13.8
Yolov5m	97.02	96.11	2.08×10^{7}	51	42.3
TMT-YOLOv5s	98.59	96.31	7.28×10^{6}	67	14.2

To provide a more intuitive and clear assessment of the model's results, we present the confusion matrix plots comparing the models in Figure 4. The confusion matrix illustrates the areas of confusion in the predictive performance of the classification model [8]. It reveals the current errors in classification and the likelihood of future errors occurring. The values on the diagonal represent the TP values mentioned earlier. The deeper the blue color and the larger the numbers on the diagonal, the better the accuracy of correct classification and the better the overall classification performance. From the plots, it can be observed that VLCV exhibits the best classification performance as it has a large number of instances and relatively high accuracy. Powdery features are distinct, resulting in relatively high classification performance among the nine tomato leaf disease categories. Conversely, Septoria yields the lowest results among the nine leaf categories, which is somewhat mitigated by the use of the MobileNet-CAL model. The MobileNet-CAL model, constructed in this study by modifying the network structure and incorporating transfer learning principles, shows improvement over the original MobileNetV3 model in the classification tasks of the nine leaf categories. It demonstrates higher classification performance for Late, Leaf, Spider, Powdery, Tomv, and YLCY compared to the other seven networks. Through the comparison of confusion matrix results, it is evident that the proposed MobileNet-CAL model holds certain advantages in the classification and recognition of tomato leaf diseases and pests.

ISSN 2616-5775 Vol. 7, Issue 2: 72-78, DOI: 10.25236/AJCIS.2024.070210

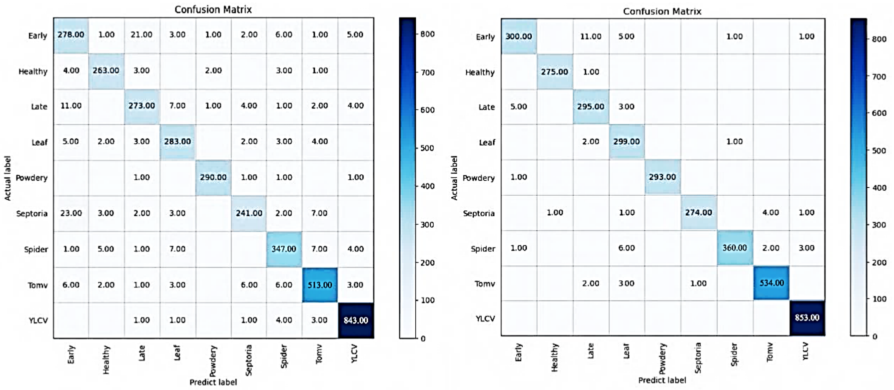


Figure 4: Comparison of improvements in MobileNet-CAL confusion matrices

4. Conclusions

This study proposes a tomato leaf disease detection model, TMT-YOLOv5s, based on the YOLOv5s model, aiming to address the challenges of low accuracy and poor performance in detecting tomato leaf diseases in complex real-world backgrounds observed in existing models. Comparative experiments are conducted with Faster R-CNN, SSD, YOLOv5s, and YOLOv5m models to validate the detection performance of the proposed model. Initially, this study constructs a DCAM attention mechanism module within the Backbone network to enhance the extraction of pathological features of tomato leaf diseases, thereby reducing interference from background features in target detection. Subsequently, the PANet path aggregation network in the Neck is replaced by the BiFPN network to improve the model's learning ability for various leaf disease features, enabling accurate detection of different types of tomato leaf diseases.

The comparative experimental results show that the TMT-YOLOv5s model achieves a mean average precision of 98.59% and a recall rate of 96.31%, with a detection FPS of 67 and a model size of 14.2 MB. Despite a decrease in detection FPS by 3 and an increase in model size by 0.4 MB compared to the original YOLOv5s model, the proposed model demonstrates significant improvements in detection accuracy and effectiveness, while still maintaining a relatively small model size. Furthermore, the effectiveness of using the TMT-YOLOv5s model for tomato leaf disease detection in complex real-world environments is validated, providing insights for its practical application in the field of agricultural intelligence.

Acknowledgments

We express our sincere gratitude to our menter, Dr. Ruihan Chen, for his invaluable guidance and unwavering support throughout our research journey. His continuous assistance and patience have played a vital role in ensuring the smooth progress of our experiments and the successful completion of this thesis. We deeply appreciate his dedication and expertise, which have greatly contributed to the overall success of our work.

This research was supported by a special grant from the National College Students Innovation and Entrepreneurship Training Program under Grant No. 202310566022, Guangdong University Students Innovation and Entrepreneurship Training Program under Grant No. S202310566045, Guangdong Provincial Science and Technology Innovation Strategy under Grant No. pdjh2022a0231, Guangdong Provincial Science and Technology Innovation Strategy under Grant No. pdjh2023b0247, and Guangdong Ocean University Undergraduate Innovation Team Project under Grant No. CXTD2023014.

References

[1] Abeid E E S ,Mosa A M ,Tabakh E M A M , et al.Antifungal activity of copper oxide nanoparticles derived from Zizyphus spina leaf extract against Fusarium root rot disease in tomato plants.[J]. Journal of Nanobiotechnology, 2024, 22(1):28-28.
[2] Ruihan C, Minhua Y, Zhi L, et al. Empirical assessment of carbon emissions in Guangdong Province within the framework of carbon peaking and carbon neutrality: a lasso-TPE-BP neural network

ISSN 2616-5775 Vol. 7, Issue 2: 72-78, DOI: 10.25236/AJCIS.2024.070210

approach.[J]. Environmental science and pollution research international, 2023, 30(58):121647-121665.
[3] Wen Y, Chen Y, Wang K, et al. A Review of Pest Detection Based on Machine Vision[J]. Journal of China Oils and Grains Society, 2022, 37(10): 271-279.
[4] Liu J, Wang X. Tomato Disease and Pest Recognition Algorithm Based on YOLO[J]. China Vegetables, 2020, 33(9): 18-22, 38.
[5] Wang W, Liu Z, Gao P, et al. A Lychee Disease and Pest Detection Model Based on Improved YOLOv4 [J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(5): 227-235.
[6] Liu G, Hu G, Gulibahaertuoheti, et al. Detection Method of Grape Leaf Diseases and Pests Based on Improved YOLOv3[J]. Microelectronics & Computer, 2023, 40(2): 110-119.
[7] Cai P ,Yang D ,Zou Y , et al.Reconsidering Multi-Branch Aggregation for Semantic Segmentation[J]. Electronics, 2023, 12(15).
[8] Luo J ,Wang J ,Wen B , et al.Accelerated genetic algorithm based on real number coding for solving mobile communication network site planning[J].Academic Journal of Computing Information Science, 2023, 6(7).