Detection and Identification of Daylily Maturity Based on YOLOv8

Fangyan Li

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222，China

E-mail: 3162129540@qq.com

www.tust.edu.cn

Abstract

To better apply object detection and identification techniques from deep learning to the field of agricultural automation, this paper focuses on the growth process of daylilies. It employs the state-of-the-art YOLOv8 model to achieve accurate assessment of daylily maturity. The backbone network of YOLOv8 draws inspiration from the CSPDarkNet network structure to extract image features. The introduction of cross-stage connections in the middle layers of the network enhances the efficiency of feature propagation. The neck section adopts the PAFPN bidirectional channel network to further process the features extracted by the backbone network, facilitating the smooth transmission of information from the bottom to the top layers. The Head section utilizes the Decoupled-Head structure to generate the output for object detection. The trained model had a mAP which reached up to 94%, an accuracy rate up to 95.6%, a recall rate up to 90.5%. Moreover, the identification speed is significantly improved.

Keywords: Deep Learning, Object Detection, YOLOv8, Daylily

1. Introduction

With the gradual maturity of AI technology, the focus of agricultural harvesting is gradually shifting to automation. The agricultural sector has been continuously exploring how to leverage modern computer vision technology to enhance crop production efficiency and quality. In the production and management of agricultural products, understanding the maturity of plants is crucial for achieving optimal harvesting and resource utilization. As an important edible plant, the timely monitoring and accurate assessment of the maturity of daylilies are essential for agricultural production [1]. Traditional harvesting of daylilies typically relies on manual inspection and experiential judgment, which poses challenges in terms of high labor costs and low efficiency. With the rapid development of computer vision and deep learning technologies, object detection techniques have gradually become an effective means to address this issue. For instance, Song [2] constructed and trained a Faster R-CNN model implemented by VGG16, which demonstrated good detection performance for fruit images under different time and lighting conditions. Liang [3] proposed improvements on the SSD network based on VGG16 and ZFNet for real-time mango detection. The enhanced algorithm achieved higher accuracy compared to Faster R-CNN, although the deep network structure led to slower computation speed. Zhen [4] introduced a vegetable image classification method, utilizing the Caffe open-source deep learning framework. They improved the VGG network by adding a Batch Normalization layer to the VGG-M network, enhancing the convergence speed and accuracy of the network (VGG-M-BN) with a vegetable classification recognition accuracy of up to 96.5%. Xue [5] aimed to improve the accuracy of immature mango detection in orchard scenes. They modified the YOLO V2 algorithm, developing a tiny-yolo network structure by combining multiple feature layers. This addressed the challenging issue of detecting overlapping and occluded parts of mangoes, achieving a detection accuracy of

97.02%. Zhao [6] proposed an apple localization method based on the YOLOv3 deep convolutional neural network. This method ensures the detection of apples in complex environments while balancing efficiency and accuracy.

The main contributions of this paper include the construction of a large-scale daylily image dataset, detailed annotation of daylily features at different growth stages, and the design of an architecture and training methodology based on the YOLOv8 model. Through experiments, the proposed approach has been demonstrated to exhibit high accuracy and practicality in the automatic detection and identification of daylily maturity.

2. Data Acquisition and Annotation

2.1. Data acquisition

The datasets for this experiment are all derived from manual field shooting with an f/1.6 wide-angle camera. Image categories encompass immature daylilies, mature daylilies, and overly mature daylilies. The dataset includes single and multiple targets, unobstructed and clustered daylily samples, captured under different weather conditions (sunny and cloudy) and at various time intervals. After acquiring a sufficient amount of video footage, the videos were decomposed frame by frame and converted into static image data. To curate the dataset, images with minimal variations in daylily features within the same second or in close proximity were excluded, as well as those that lacked specific images of daylily maturity flowers. The initial dataset was then refined by retaining images with distinct characteristics and clearly identifiable daylily buds. In order to expand the diversity of data, this paper uses the ACE image enhancement algorithm and histogram equalization algorithm to enhance the daylily dataset. In the end, we obtained a dataset of 2060 daylily images.

2.2. Images annotation and dataset production

To ensure the accuracy of training data, prior to experimentation, the daylily dataset was meticulously annotated, defining precise labels for optimal training outcomes. The Labelme annotation software was uniformly used for annotating daylily images. Based on characteristics such as color and shape of daylily flowers, the images were categorized into three classes: immature daylily, mature daylily, and overmature daylily. Immature daylilies, characterized by unopened flower buds, exhibit deep green or green coloration. These buds are relatively small and not ready for harvesting. Harvestable daylilies have bright yellow-green buds, larger in size, and fully developed, ready for picking. Overly mature daylilies have either opened or wilted buds, displaying bright yellow or dark yellow coloration with petal-like structures. The annotation process is shown in Fig. 1. The marked data is a JSON file, which is converted to a txt file . The dataset was split into 80% for training and 20% for testing and validation collectively. The testing set was utilized to assess the model's generalization capabilities post-training, while the validation set was employed for fine-tuning hyperparameters and conducting preliminary assessments of the model's performance.

Fig. 1 Dataset labels

3. Methodology

This study selected YOLOv8n as the foundational framework for research. YOLOv8, being at the forefront of object detection technology, is recognized for its efficient detection speed and accuracy across various object detection tasks. The algorithm takes the input image and resizes it to a fixed size, dividing the resized image into a grid of cells. Each cell is responsible for predicting objects that fall within its boundaries. For each cell, the algorithm predicts B bounding boxes and the probability C for each category, each containing 5 values: (x, y, w, h, c). These (x, y) values represent the center of the bounding box relative to the cell boundary, and (w, h) represent the width and height of the bounding box. This C value indicates the confidence of the algorithm in the prediction. Non-maximal suppression (NMS) is applied to remove redundant bounding box predictions. NMS compares the overlap of the predicted bounding box, leaving only the one with the highest confidence, and IoU evaluates the accuracy of the bounding box prediction from 0 to 1, with higher values indicating better overlap between the predicted bounding box and the true bounding box. The

offering an innovative and intelligent solution for agriculture.

The YOLOv8 model comprises three main components: the backbone, neck, and head network.CSPDarknet53 serves as the backbone network for YOLOv8, primarily responsible for extracting high-level features from input images. It extracts features at multiple levels and outputs feature maps at different scales. These feature maps possess varying receptive fields and semantic information, aiding the model in better handling targets of different sizes and complexities. In contrast to earlier versions of YOLOv5, YOLOv8 eliminates the less user-friendly Focus module, and the initial network layers are directly accomplished by straightforward and conventional convolutions. CSPDarknet53 is an improvement upon Darknet53, introducing the CSPNet (Cross Stage Partial Network) structure. This structure involves cross-stage partial connections, enhancing information propagation efficiency. It divides feature maps into two parts, with one part directly passed to the next stage, and the other part subjected to further processing. This contributes to strengthening information propagation and feature fusion within the network.

The neck still employs the PaFPN (Path Aggregation Feature Pyramid Network) structure to construct the YOLO feature pyramid, facilitating comprehensive fusion of multi-scale information. In comparison to YOLOv5, the only difference lies in the reduction of a 1x1 convolutional layer in the upsampling operation during the top-down process. Additionally, the C3 module has been replaced with the C2f module. Notably, the channel numbers returned from the three scales are kept equal to the channel numbers output by the backbone for the three scales.

The Head section adopts a decoupled structure, separating the classification and detection heads. Two parallel branches independently extract category and position features, followed by a 1×1 convolution layer for each to perform classification and localization tasks. Simultaneously, the use of anchor boxes is discarded.

The loss calculation process comprises two main components: positive and negative sample assignment strategy and loss computation. Considering the superior performance of dynamic assignment strategies, YOLOv8

final output is the detection result, including the category label, bounding box coordinates, and confidence score.

Leveraging its capabilities, the study applies YOLOv8 to the detection of daylily maturity,

algorithm directly incorporates TOOD's Task Aligned Assigner. It selects positive samples based on the weighted scores of classification and regression. Loss computation includes classification and regression. Binary Cross Entropy (BCE) Loss is still employed for classification. For the regression branch, it is bound to the integral representation introduced in Distribution Focal Loss, thus utilizing Distribution Focal Loss and CIoU Loss. The three losses are weighted using a certain proportion. The final training results are generated accordingly.

4. Experiments

4.1. Experimental environment

To ensure the rigor of the experiments, both the training and testing processes in this paper were conducted on the same experimental platform. The experimental environment was built and configured based on Ubuntu 18.04, with a NVIDIA GeForce GTX 1070 graphics card, Intel Core i5-9400F CPU, and programming platform Anaconda 2022.05. The CUDA version used was 11.2, and the deep learning framework employed was PyTorch 1.10 +CUDA11.1 support. The programming language used was Python 3.7.

4.2. Model training and evaluation

Through multiple experiments in the early stage, the maximum convergence of daylily target detection model training was obtained. The optimal number of iterations was 400 times, and 400 iterations were used for model training to ensure the consistency of experimental parameters. The basic parameters of the model are set to epoch=400 and batch size=16, and the initial learning rate is set to 0.01 using the SGD optimizer.

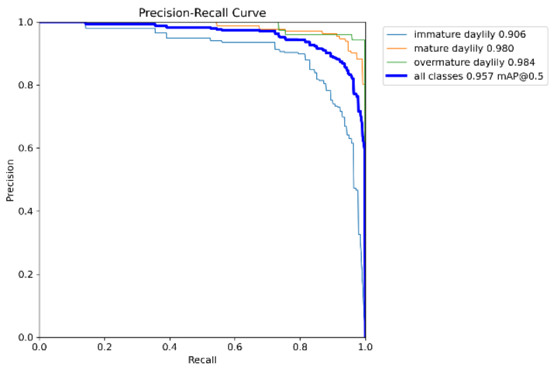
In this paper, an objective criterion was used to evaluate the daylily identification line. we use precision (P), recall (R), and mAP (mean average precision) to evaluate the model trained in this paper. As shown in Equation (1).

$$P = \frac{TP}{TP+FP}$$
$$R = \frac{TP}{TP+FN} \quad (1)$$
$$mAP = \frac{1}{C}\sum_{k=i}^{N} P(k)\Delta R(k)$$

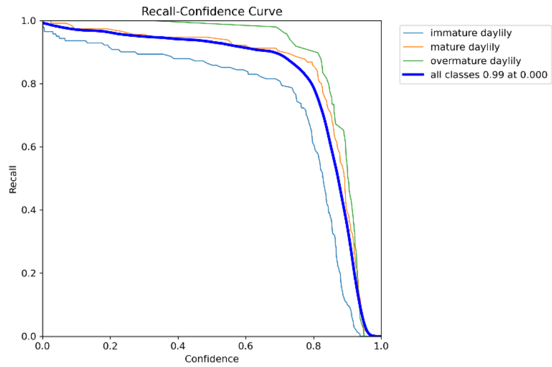
where TP is the true number of positive samples and FP is the false number of positive samples, FN is the number of spurious negative samples, C is the number of categories, and N is the reference threshold. The number of values, k is the threshold, P(k) is the accuracy, and R(k) is the recall.

5. Result

The results of different species of daylily tested on this model are shown in Fig. 2. The P-R curve and R curve of the model are shown in Fig. 3.



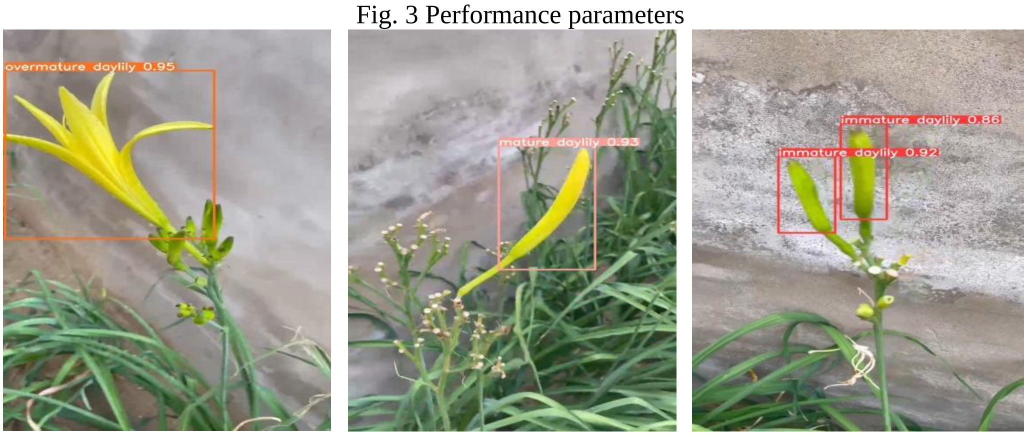
(a) P-R Curve



(b) R-Curve

Fig. 2 Detection result of the algorithm

Fig. 3 Performance parameters



6. Conclusion

After completing the model training, we conducted validation on daylilies at different growth stages. Experimental results indicate that the proposed method exhibits high accuracy and robustness in detecting and identifying the maturity of daylilies. The accuracy achieved is above 95.6%, with a recall rate of 90.5% and a mean Average Precision (mAP) of 0.94. Furthermore, our model demonstrates stability under varying lighting conditions and significant detection performance across daylilies with different shapes and colors.

Our model not only provides a reliable solution for the intelligent cultivation of daylilies but also offers effective support for agricultural decision-making. The stability of performance in diverse lighting conditions and its ability to detect daylilies of varying shapes and colors make it a valuable tool for farmers. This research lays the foundation for agricultural intelligence, enhancing the productivity and decision-making capabilities of farmers.

References

1. Zhang X, Ma H, Wei S, et al. Design of Day-Lily Robot Recognition System Based on ZYNQ. *International Conference on Networking, Communications and Information Technology (NetCIT).* IEEE, 2021: 154-157.
2. SONG ZH ZH，FU L SH，WU J ZH，et al. Kiwifruit detection in field images using Faster R -CNN with VGG16. IFAC-PapersOnLine，2019，52 (30) : 76-81.
3. LIANG Q K，ZHU W，LONG J Y，et al. A real-time detection framework for on-tree mango based on SSD network. *International Conference on Intelligent R obotics and Applications. Springer，Cham，* 2018: 423-436.
4. Zhenbo Li, Fei Li, Ling Zhu, et al. Vegetable Recognition and Classification Based on Improved VGG Deep Learning Network Model. International Journal of Computational Intelligence Systems, 2020, 13(1) : 559-564.
5. XUE Y J，HUANG N，TU SH Q，et al. Immature mango detection based on improved YOLOv2. Transactions of the Chinese Society of Agricultural Engineering，2018，34(7) : 173-179.
6. ZHAO D AN，WU R D，LIU X Y，et al. Apple positioning based on YOLO deep convolutional neural network for picking robot in complex background. Transactions of the Chinese Society of Agricultural Engineering，2019，35(3) : 164-173.

Authors Introduction

Ms. Fangyan Li

She is the scond-year graduate student of Tianjin University of Science and Technology. Her research direction is machine Learning.