2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT) | 979-8-3503-9905-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICEICT57916.2023.10245026

Detection and Recognition of Traditional Chinese Medicine Slice Based on YOLOv8

Yaying Su[1], Baolei Cheng[1], and Yijun Cai[1,*]

[1]Smart Sensing Integrated Circuit Engineering Research Center of Universities in Fujian Province, Xiamen University of Technology, Xiamen 361024, China

*yijuncai@foxmail.com

***Abstract*—Traditional Chinese medicine (TCM) herbal pieces are a type of TCM preparation with complex production processes and a wide variety of species. It is a challenging task for ordinary people to identify TCM herbal pieces. To solve this problem, this paper proposes an object detection method for TCM herbal slice based on YOLOv8.The YOLOv8 model is trained on a Chinese herbal slice (CHS) dataset that is collected and labeled from the internet, and with the Mosaic data augmentation, the average precision of YOLOv8 model on all categories can reach 95.8%, with a model size of 5.94MB, and can achieve 30~45 FPS on the MX150 graphics card environment. Compared with image classification mode, the object detection mode of YOLOv8 can better perform efficient and accurate detection and classification of CHS, which can effectively improve the recognition rate and popularization of CHS, and enable more people to benefit from the culture of TCM.**

***Index Terms*—deep learning, YOLOv8, object recognition, TCM detection**

I. Introduction

Traditional Chinese medicine (TCM) has been widely recognized for its unique therapeutic effects and is increasingly accepted worldwide. TCM preparations, such as Chinese herbal slice (CHS) and decoctions, have become a popular choice for those seeking natural remedies for health problems. Among these TCM products, CHS are widely used due to their convenience and easy storage. However, with the increasing demand for TCM products and the rise of e-commerce, the identification and quality control of CHS have become more challenging, and even different processing methods of the same species of herbs can result in significant variation [1]. It is usually difficult for ordinary consumers to identify different types of CHS, which can lead to improper usage and even health risks.

To address this issue, image recognition and detection techniques have been introduced to the field of medicinal plant identification, among which deep learning-based detection methods have been more widely applied in recent years. Sun et al. used a convolutional neural network (CNN) with added Triplet Loss for fine-tuning, achieving an average recognition accuracy of 71% and average retrieval accuracy of 53% on all 95 medical categories of images [2].Marwaha et al. utilized four models, including VGG16 and MobileNet, to perform accurate classification of Indian herbal plant powder micrographs. The MobileNet model demonstrated the fastest processing speed, while VGG16 achieved the highest detection accuracy of up to 96.4%.[3]. Xing et al. improved the training of the DenseNet model by utilizing dropout and introducing an adaptive average pooling layer. Based on transfer learning, the model achieved the highest recognition rate of up to 97.34% in a dataset containing 80 types of TCM images [4]. Wang et al. proposed a network called CCSM-Net, which integrates channel attention and spatial attention modules. The CCSM-Net can effectively identify CHS in 2D images, and the proposed algorithm achieved an accuracy of up to 99.33% with good performance [5].

Object detection algorithms based on deep learning, such as YOLO(You Only Look Once), have shown high efficiency and accuracy in image detection. Among various detection methods, YOLO has become an excellent choice for TCM identification. Lv Bijun et al. achieved an average accuracy of 94.33% on a dataset of 100 TCM images using the original YOLOv5 model [6]. With the development of deep learning technology, the latest YOLOv8 version has made some improvements in terms of object detection performance compared to the v5 version, making it an excellent choice for identifying CHS.

Therefore, this study proposes a CHS recognition method based on YOLOv8. We trained the model using a self-labeled CHS image dataset and evaluated its performance in terms of accuracy and other metrics. The YOLO series of algorithms can quickly and accurately identify different types of CHS. In this study, we mainly selected the YOLOv8n model, which is the lightest model in the v8 series, to reduce the computational cost of hardware devices, and compared it with YOLOv8's image classification mode.

II. YOLOv8

YOLO, proposed by Redmon et al. in 2016, is a regression-based object detection algorithm. The algorithm divides an input image into $S \times S$ grid cells and predicts B bounding boxes and their corresponding class probabilities for each cell. By transforming the object detection problem into a regression problem, YOLO achieves fast and accurate object detection[7].

Since its inception, the YOLO algorithm has undergone continuous improvements and optimizations, from v1 to v5 and now to the latest version, YOLOv8, which was introduced in early 2023. YOLOv5 employs a more lightweight network architecture and several new techniques, such as adaptive convolutional layers, the use of the Swish activation function, and the PANet architecture for multi-scale feature fusion. YOLOv5 has achieved improvements in both detection

accuracy and speed while maintaining low computational complexity.

YOLOv8 consists of five models: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x, and provides convenient and efficient API calls using Python's command line. Although there is no formal paper yet, the network structure of YOLOv8 has already been partially explained. The head of YOLOv8 uses a Decoupled-Head with different branches for computation, which helps improve performance. Moreover, the anchor-based approach used in previous versions is replaced with the anchor-free approach. The backbone still employs the CSP technique and the SPPF module from YOLOv5, but the C3 module is replaced with the C2f module, which has a richer gradient flow, resulting in further lightweighting. YOLOv8 removes the convolutional structure in the PAN-FPN upsampling phase of YOLOv5 and replaces the C3 module with the C2f module.

In addition, the data augmentation part of YOLOv8's training introduces the operation of turning off the Mosaic augmentation for the last 10 epochs, which effectively improves accuracy. The classification loss is changed to VFL Loss, and the CIOU Loss is added with DFL (Distribution Focal Loss) as the regression loss. YOLOv8 also replaces the previous IOU matching or one-sided proportion allocation with the Task-Aligned Assigner matching approach.

III. Data Construction

In this study, we downloaded the initial Chinese herbal slices (CHS) classification dataset from Baidu AI Studio and used labelimg to annotate bounding boxes in the YOLO format, converting it into an object detection dataset. The dataset contains 902 images of five different types of CHS, namely Lilii Bulbus, Codonopsis pilosula, Chinese wolfberry, Pagodatree Flower,and Honeysuckle, which were divided into training and validation sets in an 8:2 ratio. The distribution of the dataset classes is shown in Fig.1.

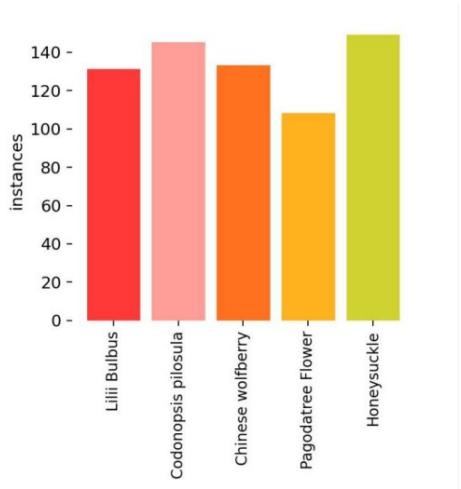

Fig. 1. Distribution of the Datasetand Figure.

IV. Experiment

A. Computer Configuration

The computer configuration used in this study for training is as follows: Windows 10 operating system; CPU is Intel Core i5-10400F, 4000 MHz (40 x 100), with 8 GB of RAM; GPU is RTX3060 (12G / ASUS); the training environment software is Pycharm, python3.7, CUDA11.6, and the deep learning open-source framework is PyTorch11.3.

The computer configuration used in this study for inference is as follows: Windows 10 operating system; lntel(R) Core(TM) i7-8550U CPU @1.80GHz 1.99 GHz processor; NVIDIA GeForce MX150 graphics card; the training environment software is Pycharm, python3.9.7, CUDA11.6, and the deep learning open-source framework is PyTorch0.12.0.

B. Evaluation Metrics

Commonly used evaluation metrics for models include recall (R), precision (P), F1-score, average precision (AP), mean average precision (mAP), FLOPs, model parameters, etc. Their analytical expressions are shown in Equations (1) - (5).

$$R = \frac{TP}{TP + TN} \tag{1}$$

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$F1 = \frac{2PR}{P + R} \tag{3}$$

$$AP = \int_0^1 P\, dR \tag{4}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{5}$$

Here, TP refers to the number of true positives in Chinese herbal slices (CHS) that are predicted as positive samples, FN refers to the number of true positives in CHS that are predicted as negative samples, and FP refers to the number of false positives in CHS that are predicted as positive samples. AP refers to the area under the PR curve (generally, the larger the area under the PR curve, the better the model), and mAP refers to the average value of various APs. N refers to the total number of categories in the dataset.

In this study, P, R, and mAP@0.5 are used as model evaluation standards, and the model size and FPS (Frames Per Second) on hardware with poor performance are examined to evaluate the performance advantages of the model.

C. Training Parameters

The parameters or hyperparameters used in the experiments are shown in Table I. In this study, the Batch Size was set to 32, epoch was set to 80, and the optimization method used was SGD. Additionally, a learning rate warmup was employed to accelerate the convergence speed of the model and improve its performance.

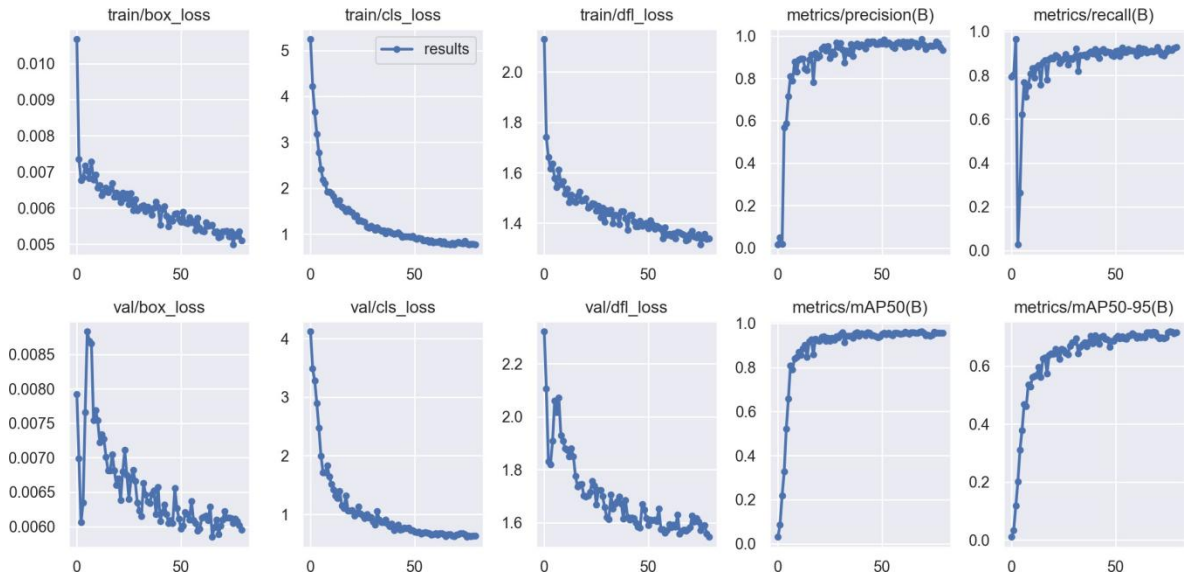

Fig. 2. The training results of the YOLOv8n model.

TABLE I. HYPERPARAMETERS

Parameter	*Value*
img-size	(640,640)
lr0	0.0016
lrf	0.236
momentum	0.965
weight decay	0.00043
warmup epochs	2.96
warmup momentum	0.799
warmup bias 1r	0.1
mosaic	0.198

D. Experimental Results and Analysis

The YOLOv8n model was trained using the pre-trained weights on ImageNet, and the training results at an IoU of 0.7 are shown in Fig.2. The precision and recall started to converge after about 20 epochs of training, and the precision had stabilized at over 90% after 50 epochs. The precision at different IoU thresholds continued to increase with the increase of mAP50-95[8]. The PR curve on the validation set is shown in Fig.3, with an average precision of 95.8% for all categories and a highest precision of 98.9% for a single category. The YOLOv8n model has a size of 5.94MB and 3.2M parameters.

In this study, we also conducted image classification training on YOLOv8 using the same dataset. The results are shown in Fig.4, where the loss began to converge after approximately 25 epochs, and the metric accuracy top1 converged at around 80% after 60 epochs. After conducting experiments on a small local CHS test set with the object detection model, the results shown in Table II indicate that YOLOv8 achieved higher precision in object detection mode than in image classification mode on our dataset.

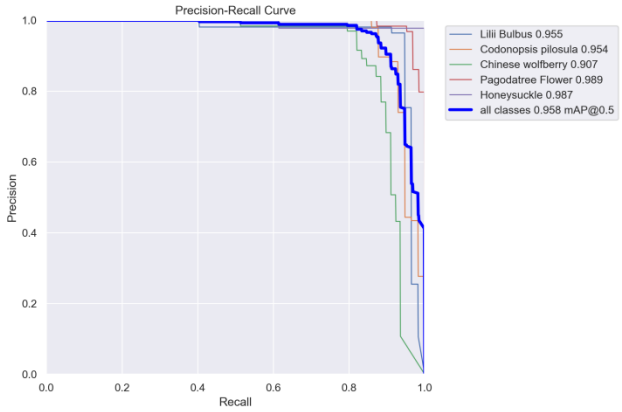

Fig. 3. The PR Curve.

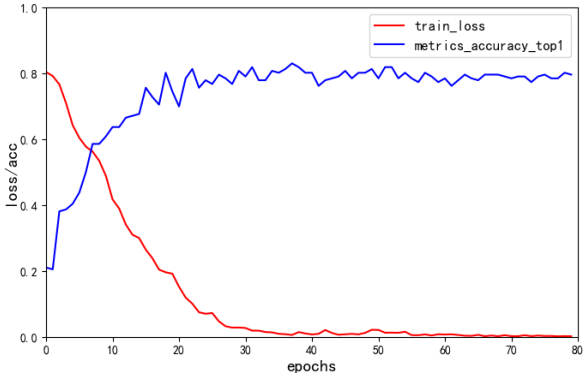

Fig. 4. The training results of the YOLOv5 image classification mode.

TABLE II. THE ACCURACY ON THE TEST

Mode	*Accuracy*
Classify	66.7%
Detect	86.7%

E. Model Actual Performance

The performance of the model for local camera detection in the GeForce MX150 graphics card environment is shown in Fig.5(a), achieving a frame rate of 30~45 FPS. Fig.5(b) shows the results of the same image with different CHS categories, indicating that even if the categories in a single image in the training set are relatively homogeneous, the trained model still performs well in detecting different categories in the same scene.

(a) Real-time camera detection.

(b) Detection results of different CHS in the same image.

Fig. 5. The actual detection effect.

V. Conclusion

In order to effectively identify Chinese herbal medicine slices (CHS), this paper proposes a YOLOv8-based CHS image detection method. YOLOv8 was trained on a CHS dataset consisting of 902 images, and an object detection model was obtained with an mAP@0.5 of 95.8%. The model is the smallest YOLOv8n model, with a size of only 5.94MB, and can achieve 30~45 FPS on a GeForce MX150 graphics card, demonstrating good CHS recognition performance and suitability for low hardware requirements environments. Next, we will consider optimizing the dataset, expanding the number of Chinese herbal medicine slice categories and the number of images per category, and optimizing the YOLOv8 structure to further advance this research.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (62005232); Natural Science Foundation of Fujian Province (2020J01294).

References

[1] T. Li, F. Sun, R. Sun, L. Wang, M. Li and H. Yang, “Chinese Herbal Medicine Classification Using Convolutional Neural Network with Mul- tiscale Images and Data Augmentation,” 2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), Jinan, China, 2018, pp. 109-113.

[2] Sun, Xin, and Huinan Qian. “Chinese Herbal Medicine Image Recog- nition and Retrieval by Convolutional Neural Network.” PloS one vol. 11,6 e0156327. 3 Jun. 2016.

[3] R. Marwaha and B. Fataniya, “Classification of Indian Herbal Plants based on powder microscopic images using Transfer Learning,” 2018 Fifth International Conference on Parallel, Distributed and Grid Com- puting (PDGC), Solan, India, 2018, pp. 500-505.

[4] C. Xing, Y. Huo, X. Huang, C. Lu, Y. Liang and A. Wang, “Research on Image Recognition Technology of Traditional Chinese Medicine Based on Deep Transfer Learning,” 2020 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), Tianjin, China,2020, pp. 140-146.

[5] Wang, Jianqing et al. “Combined Channel Attention and Spatial Atten- tion Module Network for Chinese Herbal Slices Automated Recogni- tion.” Frontiers in neuroscience vol. 16 920820. 13 Jun. 2022.

[6] Lv, Bijun et al. “Traditional Chinese Medicine Recognition Based on Target Detection.” Evidence-based complementary and alternative medicine : eCAM vol. 2022 9220443. 8 Jul. 2022.

[7] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[8] G. Grisetti, C. Stachniss and W. Burgard, “Improved Techniques for Grid Mapping With Rao-Blackwellized Particle Filters,” in IEEE Transactionson Robotics, vol. 23, no. 1, pp. 34-46.