Embedded Selforganizing Systems

Fruit Classification Based on Improved YOLOv7 Algorithm

Shibo Guo
Heilongjiang Province Key Laboratory
of Laser Spectroscopy Technology and
Application
Harbin University of Science and
Technology
Harbin, China
1347613588@QQ.com

Tianyu Ren
Heilongjiang Province Key Laboratory
of Laser Spectroscopy Technology and
Application
Harbin University of Science and
Technology
Harbin, China
rty6662022@qq.com

Qing Wu
Heilongjiang Province Key Laboratory
of Laser Spectroscopy Technology and
Application
Harbin University of Science and
Technology
Harbin, China
wuqing@hrbust.edu.cn

Xiaoyu Yu
College of Electron and Information,
University of Electronic Science and
Technology of China,
Zhongshan Institute,
Zhongshan, China
yuxy@zsc.edu.cn

Aili Wang*
Heilongjiang Province Key Laboratory
of Laser Spectroscopy Technology and
Application
Harbin University of Science and
Technology
Harbin, China
aili925@hrbust.edu.cn

Abstract—**With the rapid development of technology and advancements, unmanned vending machines have emerged as the primary contactless retail method. The efficient and accurate implementation of automated identification technology for agricultural products in their distribution and sales has become an urgent problem that needs to be addressed. This article presents an improved YOLOv7 (You Only Look Once) algorithm for fruit detection in complex environments. By replacing the 3×3 convolutions in the backbone of YOLOv7 with Deformable ConvNet v2(DCNv2), the recognition accuracy and efficiency of fruit classification in YOLOv7 are significantly enhanced. The results indicate that the overall recognition accuracy of this system for ten types of fruits is 98.3%, showcasing its high precision and stability.**

Keywords— fruit classification, YOLOv7, DCNv2

I. INTRODUCTION

At present, the method of manual identification is mostly used in the process of fruit classification, which has problems such as low information collection efficiency, inaccurate classification, and high labor costs. The main occasions for fruit circulation are fruit wholesale markets and various types of supermarkets. Small and medium-sized supermarkets and wholesale markets input the unit price of various fruits into electronic platform scales, electronic scales and other equipment for weighing and billing. In this process, there is a possibility of input errors. When the number of customers increases, the time loss due to input errors greatly reduces the shopping experience of customers. The common defect of traditional sales methods are time-consuming and error-prone.

In order to achieve automation and improve the accuracy of fruit recognition, many researchers have applied image recognition techniques to the field of fruit classification. Arivazhagan et al. [1] used color and texture to identify fruits and validated their effectiveness among 15 different fruits. Rocha et al. [2] combined different features and classifiers to reduce the classification error of 15 classes of fruits and vegetables to 15% with limited training data. Dubey et al. [3] proposed an improved hue-difference histogram texture feature descriptor and verified it in a fruit and vegetable database. Dubey et al. [4] presented a fruit classification method using K-means for image segmentation and Support Vector Machine (SVM) for classification. The experimental results demonstrated the effectiveness of this method in fruit classification. Sachin C et al. [5] applied the YOLO algorithm to vegetable recognition by inputting multiple images of different vegetables into the network, manually drawing bounding boxes around the vegetables using OpenCV, and preprocessing the images before training. The recognition accuracy of this method reached 61.6%.

As a real-time object detector, YOLO has been widely applied in various fields. Rahul Sharma [6] proposed the application of YOLOv3 for detecting anthracnose lesions on apple surfaces. Yu JM et al. [7] applied YOLOv4 to mask detection, achieving a mask recognition mAP of 98.3% with a frame rate of 54.57 FPS, demonstrating good robustness. Hu XL et al. [8] improved YOLOv4 for real-time detection of uneaten feed particles in underwater aquaculture images, achieving an average precision improvement of 27.21% and a reduction of about 30% in computation, obtaining satisfactory recognition results.

This article applies the improved YOLOv7 algorithm to fruit recognition, replacing the 3×3 convolutions in the backbone of YOLOv7 with deformable convolutions

(DCNv2). It achieves high recognition accuracy in testing with various types of fruits.

II. METHODOLOGY

A DCNv2 Network Architecture

The DCN-v2 network is divided into two types: Stacked structure and Parallel structure. The difference between them lies in their architecture. In the Stacked structure, the input vector passes through multiple layers of Cross networks for feature interaction and then goes through a deep neural network (DNN) layer. In the Parallel structure, similar to the DCN structure, the input vector simultaneously goes through both the Cross layer and the DNN layer, and the outputs are then cascaded together.

The Cross network in DCN-v2 assumes that the output vector of the l-th layer is denoted as x_l, and the output vector of the $(l+1)$-th layer x_{x+1}，can be represented as follows:

$$x_{x+1} = x_0 \odot (x_x x_x + x_x) + x_x \tag{1}$$

The weights of the Cross layer transform from an n-dimensional vector to an $x \times x$ matrix. By analyzing the matrices learned by the Cross network in the DCN-v2 model, it is observed that computational costs can be reduced by employing matrix decomposition methods. One approach is to decompose a dense matrix into low-rank matrices. As a result, the computation in the Cross network can be represented by the following formula:

$$x_{l+1} = \sum_{i=1}^{K} G_i(x_l) E_i(x_l) + x_l \tag{2}$$

$$E_i(x_l) = x_l \odot (U_l^i(V_l^i x_l) + b_l) \tag{3}$$

B Establishment of DCNv2-YOLOv7 Network

The YOLOv7 object detection algorithm mainly consists of three components: the Backbone feature extraction structure, the FPN feature fusion structure (in this paper, the SPPCSPC structure is included in FPN), and the head for regression and classification. It can detect and classify objects simultaneously.

Modify_Multi_Concat_Block is a modified version of the Multi_Concat_Block network structure. Their architecture diagram is shown in Figure 1.
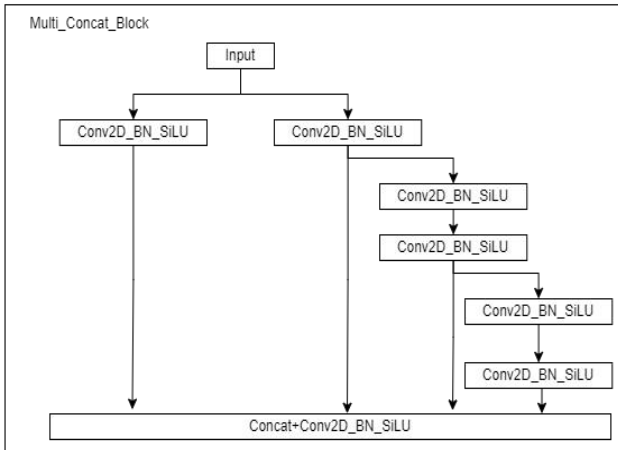
(a) Multi_Concat_Block structure
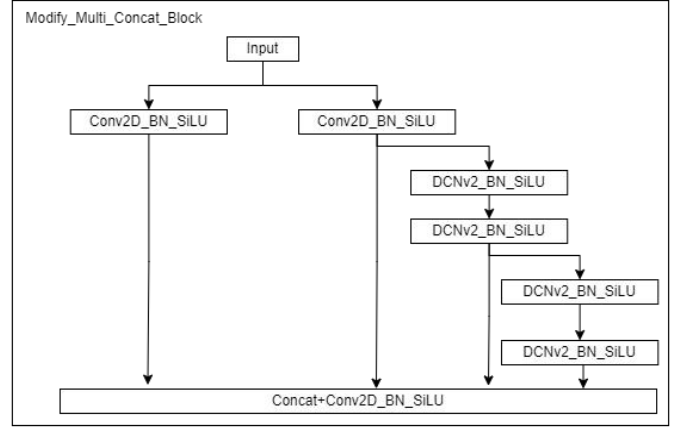
(b)Modifed_Multi_Concat_Block structure diagram
Fig.1. Comparison before and after modification

DCNv2-YOLOv7 network replaces the last convolution layer with DCNv2 in the Backbone, where the convolution kernel size is 3. The structure of DCNv2-YOLOv7 is shown in Figure 2.
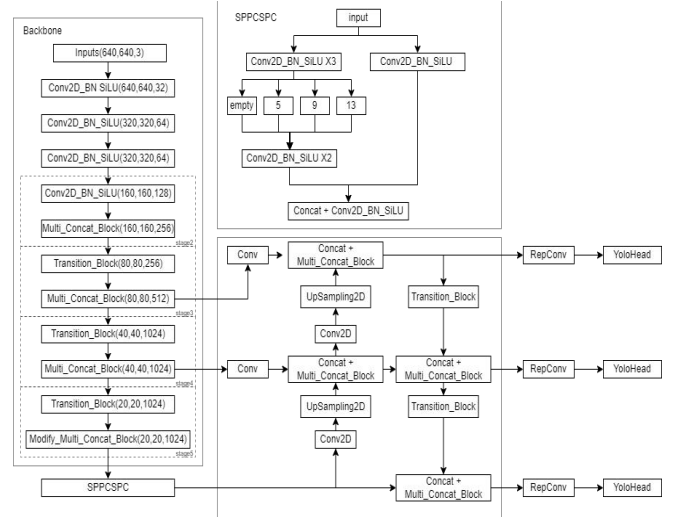

Fig.2. DCNv2-YOLOv7 network structure

In the feature utilization part, YOLOv7 extracts three feature layers located at different levels of the backbone. These feature layers are in the middle, lower, and bottom layers. By utilizing the FPN feature pyramid, we obtain three enhanced feature layers with shapes of (20,20,512), (40,40,256), and (80,80,128). These feature layers are then passed into the YOLO Head to obtain the prediction results. YOLOv7 utilizes the RepConv structure before the YOLO Head. The basic idea is to introduce a special residual structure for training, which can be equivalently simplified to a regular 3x3 convolution during actual prediction. After the above network structure, we obtained the prediction results of the three feature layers, and obtained the position of the prediction frame on the picture after decoding.

III. EXPERIMENTS AND RESULTS ANALYSIS

A. Experimental Environment

In the experiment, we conducte it on a computer with Windows 10 operating system, equipped with a GeForce RTX 2060 GPU. The CUDA version used is 11.7, and the

CuDNN version is 8.8.1. We train the models using the PyTorch framework.

B. Datasets

We use a home-made fruit dataset to evaluate whether model improvements are effective. It includes 2956 pictures of six kinds of fruits including kiwi, rockmelon, strawberry, and mango. A picture may contain a variety of fruits, and the statistical chart of the label is shown in the figure 3.

There is a problem of sample imbalance in the dataset, which can be improved to a certain extent through image weight redistribution and data enhancement such as cropping, rotation, and flipping. Use the ratio of 6:2:2 to divide the data set to training set, verification set and test set.
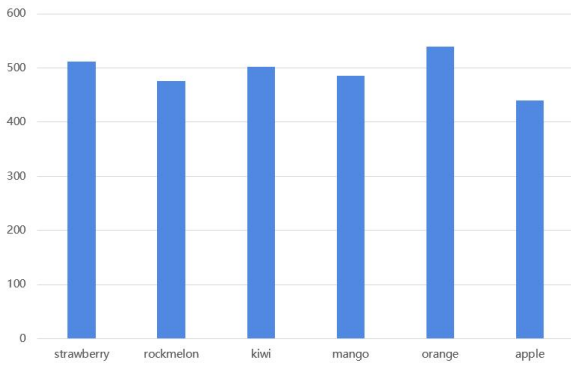

Fig.3. Label statistics for each category

C. Experimental results and analysis.

In order to test our system, we identified 4 fruits: mango, strawberry, orange and cantaloupe. Figure 4 (a), (c), (e) are the DCNv2-YOLOv7 recognition results of cantaloupe, strawberry, and kiwifruit, respectively, and Figure 3 (b), (d), (f) are the recognition results of YOLOv7. For Figure 3(a) and (b), DCNv2-YOLOv7 has a 96% probability of being recognized as a cantaloupe; YOLOv7 has a 90% probability of being recognized as a cantaloupe. For Figure 3(c) and (d), DCNv2-YOLOv7 has 95% probability of being recognized as strawberry, while YOLOv7 has 88% and 93% probability of being recognized as strawberry. For Figure 3(e) and (f), DCNv2-YOLOv7 has a 92% probability of being recognized as kiwi, and YOLOv7 has a 91% probability of being recognized as orange. In addition, the recognition time of the three fruits DCNv2-YOLOv7 is 54ms. The recognition time of these four fruits of YOLOv7 is 58ms.

(a)DCNv2-YOLOv7   (b) YOLOv7
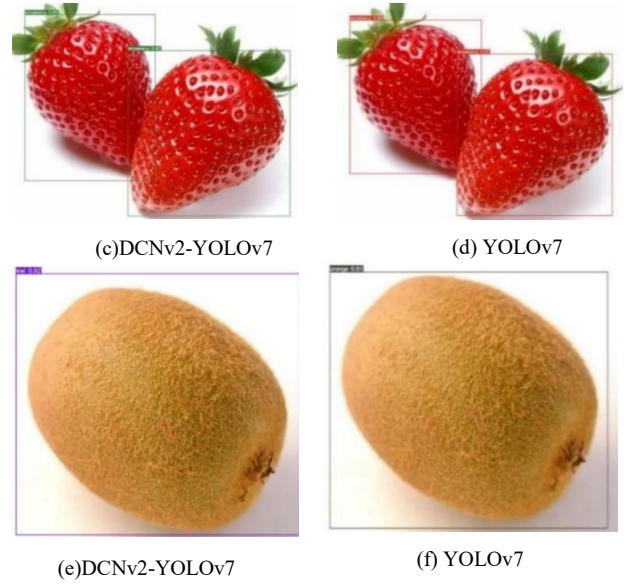
(c)DCNv2-YOLOv7   (d) YOLOv7

(e)DCNv2-YOLOv7   (f) YOLOv7
Fig.4. Comparison of DCNv2-YOLOv7 and YOLOv7.

D. Experimental results.

In practical application scenarios, the brightness of the camera-captured images can be affected by environmental factors such as occlusions. Therefore, we conducted recognition tests on fruits under different brightness conditions. The brightness levels were categorized into dark, normal, and bright, with 20 trials for each group. We used accuracy as the evaluation metric for our model, which can be obtained from the confusion matrix in Table I.

TABLE I.   CONFUSION MATRIX

	Positive	Negative
Predict as positive	TP (True Positive)	FP (False Positive)
Predict as negative	FN (False Negative)	TN (True Negative)

The calculation method for accuracy is as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

TABLE II.   FRUIT RECOGNITION RESULTS OF DCNv2-YOLOv7 UNDER DIFFERENT BRIGHTNESS

Fruit	Brightness	DCN_YOLOv7(%)	YOLOv7(%)
Cantaloupe	Dark	95	90
	Normal	100	95
	Bright	100	100
Strawberry	Dark	100	100
	Normal	100	100
	Bright	100	100
Kiwi	Dark	90	85
	Normal	100	95
	Bright	100	100
Tangerine	Dark	95	95
	Normal	100	100
	Bright	100	100
Total		98.3	96.7

Table II shows the recognition of five fruits under different brightness conditions, and the recognition probability of each fruit gradually increases with the increase of brightness. When the brightness is brighter, all fruits can be successfully identified. When the brightness is darker, the probability of correct recognition of kiwi is lower, and the accuracy rate is 90%. In all 240 sets of tests, the overall

accuracy rate can reach 98.3%. It can be seen from Table II that DCNv2-YOLOv7 can complete the classification task very well, with an accuracy rate of about 98.3%, which meets the expected improvement requirements. In practical applications, it can guarantee high recognition accuracy and efficiency in most cases. How to improve the low recognition rate caused by the detection environment will be the next research direction.

We set the epoch to 100. YOLOv7 and DCNv2-YOLOv7 were trained respectively. We use the test set independent of the training set and the verification set to evaluate the model performance. The obtained model test results are shown in the Table III. YOLOv7-DCNv2 has improved on all four metrics. This proves that our model improvement is effective.

TABLE III. ABLATION STUDY ON TEST-SET OF FRUIT

Methods	Precision (%)	Recall (%)	F1 (%)	mAP@0.5 (%)
YOLOv7	95.1	96.0	61.5	88.0
DCNv2-YOLOv7	96.2 (↑1.2)	97.0 (↑1.0)	62.8 (↑1.3)	89.1 (↑1.1)

E. Ablation Study

For the DCNv2-YOLOv7 model shown in Figure 5, two loss curves and two metric curves of the training set are drawn, as shown in the figure. Each loss tends to be stable and eventually converges to a small value, and the metrics curve eventually stabilizes, which proves that the parameters set by the model are reasonable.

(a) box_loss   (b) precision
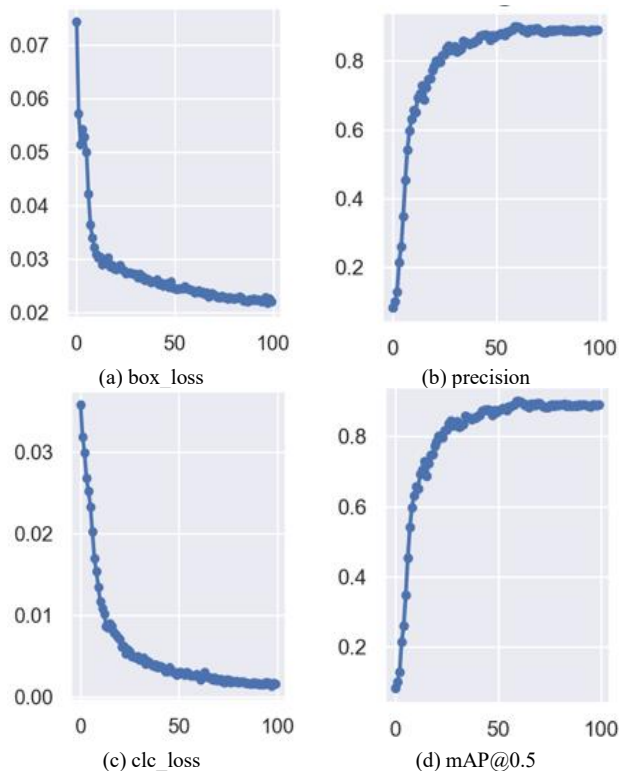
(c) clc_loss   (d) mAP@0.5

Fig.5. Loss and metrics.

As shown in Fig. 6, We also choose some pictures of DCNv2-YOLOv7 detected in the test set as the actual results.

Fig.6. Visualization results from DCNv2-YOLOv7 on test-set.

IV. CONCLUSIONS

This paper improves YOLOv7 and modifies the backbone part of the backbone feature extraction network. The improved DCNv2-YOLOv7 experiment results show that the recognition accuracy of 4 kinds of fruits can reach 98.3% under three lighting conditions: dark, normal and bright. Through the completion of the above work, it can be found that the algorithm research based on deep learning theory has important practical significance for intelligent identification of fruit species.

REFERENCES

[1] Arivazhagan S, Shebiah N, Nidhyanandhan S, Ganesan. "Fruit recognition using color and texture features". in Journal of Emerging Trends in Computing and Information Sciences, 2010, 1(2), pp. 90-94.

[2] Rocha, A. Hauagge, D. C. Wainer, J. Goldenstein, S. "Automatic fruit and vegetable classification from images," in Comput Electron Agric,2010, vol. 70,no.1, pp. 96-104.

[3] I Dubey, S. R. Jalal, A. S. "Robust Approach for Fruit and Vegetable Classification," procedia engineering, 2012, 38: 3449-3453.

[4] Dubey, S. R, A. S. Jalal . "Detection and Classification of Apple Fruit Diseases Using Complete Local Binary Patterns." In Third International Conference on Computer and Communication Technology. IEEE, 2012: 346-351.

[5] S. C, N. Manasa, V. Sharma, N. K. A. A., "Vegetable Classification Using You Only Look Once Algorithm," in ICon-CuTE, 2019, pp. 101-107, doi: 10.1109/ICon-CuTE47290.2019.8991457.

[6] Mr. Rahul Sharma.. "Recognition of Anthracnose Injuries on Apple Surfaces using YOLOV 3-Dense" in International Journal of New Practices in Management and Engineering, 2015, 4(02): 08-14.

[7] Yu J, Zhang W. "Face mask wearing detection algorithm based on improved YOLO-v4".in Sensors, 2021 21(9): 3263.

[8] Hu X, Liu Y, Zhao Z x, Liu Jt, Yang Xt, Sun Ch, Chen Sh, Li B, Zhou C, " Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO-V4 network". in Computers and Electronics in Agriculture, 2021, pp: 185.