Design of Pineapple Recognition System Based on Deep Learning

Wang Runtao*, Li Zelong, Huang Chuqi, Chen Xiaoran, Liang Rongwei, Lin Zixin, Zeng Fanjin, Li Zhitong, Zhang Qingyi
School of Electronic and Electrical Engineering, Lingnan Normal University, Guangdong, China
**Corresponding author*

Abstract: **In view of the current situation that pineapple picking operation is highly dependent on manpower and the level of automation is insufficient, this study designed a set of pineapple fruit automatic recognition system suitable for complex background based on deep learning and embedded technology. The hardware of the system is composed of a target recognition module and a control module. The target recognition module uses k210 as the main control chip to realize the real-time acquisition and processing of pineapple images. By comparing the performance of the algorithm, yolov5 performs best in the pineapple recognition task, and the mean accuracy (map) is 0.987. The benchmark yolov5 model was further improved in two aspects: first, the collaborative attention mechanism (CA) was introduced to improve the map to 0.992; The second is to use the lightweight backbone network shufflenet V2 and optimize the structure, which significantly reduces the amount of model parameters. This research provides an effective technical scheme for the automation and intelligence of pineapple production.**

Keywords: Pineapple Object Detection; Deep Learning; Algorithm Optimization; K210; Attention Mechanism; Model Lightweighting

1. Introduction

Pineapple picking is a typical centralized labor-intensive operation, which is highly dependent on manpower. Realizing mechani-zation and automation to replace most or all manual operations is an important trend to liberate productivity and improve production efficiency. It has become an urgent need and research frontier in the field of agricultural automation to promote the transformation of agricultural production to automation and intelligence at this stage and completely replace manual operation. Since the beginning of the 21st century, with the progress of sensor technology, the combination of automation technology and artificial intelligence has spawned an intelligent robot system that can independently complete fruit picking. However, the complex unstructured orchard environment puts forward higher requirements for the target recognition technology of picking robot. In this context, deep learning technology has developed rapidly after extensive research, especially in the field of target detection, scholars have proposed a variety of effective methods. Researchers have successfully applied the deep convolutional neural network to the fruit image recognition system. Compared with the traditional digital image processing technology, the deep learning model is relatively less affected by light changes, fruit occlusion, and background interference similar to the color of the fruit. It can still maintain high recognition accuracy and strong robustness in complex scenes with large environmental variables.

In view of this, this paper aims to design a pineapple fruit recognition system based on deep learning. Through the experiments of benchmark algorithm comparison test, model structure optimization and performance analysis, an efficient pineapple recognition model was constructed. On this basis, the optimized model is deployed on the embedded hardware platform, and finally a pineapple target real-time recognition system that can run stably under complex background is realized.

2. Hardware Design of Pineapple Recogni-Tion System

The overall hardware architecture of the system adopts hierarchical design, as shown in Figure 1, which is divided into perception decision layer based on machine vision and task execution layer based on real-time control. The perception layer takes kendryte k210, an edge AI computing chip with convolutional neural network acceleration ability, as the core, and its integrated high-power KPU (peak performance of 0.8 tops) is responsible for running the lightweight target detection model; This layer collects visual data through ov2640 digital image sensor, and realizes interactive graphical output with the help of LCD touch screen. The execution layer takes the classic 8-bit microcontroller STC89C52 as the control core, which is responsible for receiving the instructions from the upper layer and generating accurate timing logic; It integrates pca9685 multi-channel PWM servo driver, configures high-resolution PWM waveform with IIC protocol, and then drives multiple tbs2701 digital actuators to form a closed-loop actuator. At the same time, it uses LCD1602 character LCD module to monitor the status. The two levels exchange instructions and data through the asynchronous serial communication (UART) protocol, forming a collaborative closed-loop of "visual perception - intelligent decision-making - precise control". While giving consideration to high-performance AI reasoning and high reliability real-time control, it realizes the decoupling and optimization of perception and execution functions in embedded scenes.The hardware circuit diagram of the overall framework is shown in Figure 1.

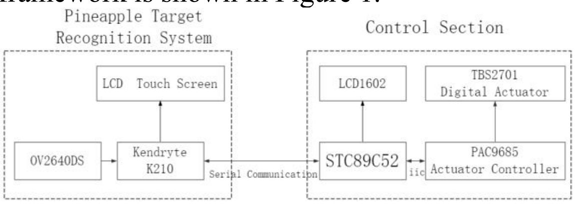

Figure 1. Overall Hardware Block Diagram of the System

3. Data Acquisition and Processing of Pineapple Images

3.1 labeling Of Dataset

In the data set construction work of this study, we integrated the pineapple image data set from kaggle open source platform and other public network resources. In order to adapt to the standard input format required by mains-tream deep learning detection frameworks (such as fast r-cnn and Yolo Series), we have carried out systematic data annotation and format conversion on all images. The specific process is as follows: firstly, labellmg, a general annotation tool, is used to generate the corresponding XML format annotation file for each pineapple image, in which the coordinates and size information of the target boundary box are recorded in detail; Then, by writing a special Python script, the annotation informa-tion in XML format is unified into a standar-dized text format that meets the requirements of target detection network training. This preprocessing process ensures the unity of data format and the standardization of network input. Txt file content is shown in Table 1, and class pineapple indicates pineapple category.

Table 1. Sample Content of TXT Label File

Class	X_center	Y_center	W	H
pineapple	0.5016	0.4962	0.9459	0.9925
pineapple	0.5008	0.5012	0.9476	0.9925
pineapple	0.4434	0.3904	0.5585	0.3012
pineapple	0.4295	0.3695	0.5005	0.3073

3.2 Augmentation of Dataset

In order to improve the robustness of the model to pineapple features and enhance its generalization performance, this study conducted a systematic off-line data enhancement processing on the data set. Due to the influence of illumination conditions and other factors in the acquisition process of the original image, there are significant intra class differences and uneven distribution of samples, which are easy to lead to over fitting of the model. To this end, we implemented a variety of data enhancement operations on image data, including random clipping, rotation, image flipping and noise injection, by writing Python scripts, so as to effectively expand the size of training samples and improve data diversity. Figure 2 shows some sample results after data enhancement processing.

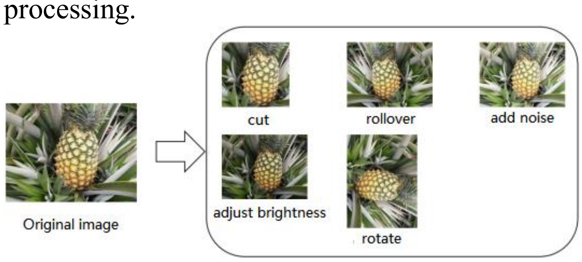

Figure 2. Pineapple Data Enhancement Example

3.3 Establishment of Dataset

The original pineapple image is marked with a bounding box by manual annotation. The specific annotation interface is shown in Figure 3. Subsequently, the scale of the annotated dataset was expanded by using offline data enhancement technology, and a total of 8000 pineapple image samples were finally obtained.

On this basis, the complete data set is randomly divided into training set, verification set and test set by Python script in the proportion of 8:1:1 to support the training and evaluation of subsequent models.

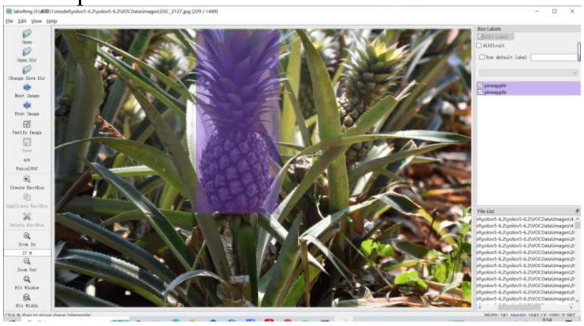

Figure 3. Actual Annotation Page

4. Comparative Analysis of Target Detection Networks

The recognition and detection of pineapple targets by different target detection networks are shown in Table 2.

Table 2. Detection Data of Pineapple Targets by Different Target Detection Networks

Target detection network	Precision	Recall	Detection time (frame/ms)	mAP@0.5
Faster-R-CNN	74.42	97.71	98	0.9292
YOLOv4	81.43	89.77	51	0.8948
YOLOv5	86.8	98	40	0.987

As shown in Table 2, for pineapple target detection task, the performance of three models, namely, faster r-cnn, yolov4 and yolov3, is as follows: the accuracy rate of faster r-cnn is 74.42%, the recall rate is 97.71%, the average detection time is 98ms, and the average accuracy (map) is 92.92%; The accuracy rate of yolov4 was 81.43%, the recall rate was 89.77%, the average detection time was 51ms, and the map was 89.48%; The accuracy rate of yolov5 was 86.8%, the recall rate was 98%, the average detection time was 40ms, and the map was 98.7%. By comprehensive comparison, yolov5 is the best in terms of accuracy, recall, map and detection speed, achieving the best balance between detection accuracy and efficiency. Its performance advantage is mainly due to the introduction of CSP module, mosaic data enhancement, focus structure and fpn+panet in the network structure, which effectively enhances the ability of feature extraction and target classification. In contrast, fast r-cnn is prone to over fitting due to its deep network layer and large number of parameters, although its recall rate is high, resulting in a decline in accuracy and significant detection delay. Therefore, yolov5 can be used as the basic detection network for subsequent improvement research.

5. Analysis of Experimental Results

5.1 Comparison of Collaborative Attention Mechanisms

In order to evaluate the impact of coordinated attention (CA) module on the performance of the model, an improved model integrated with Ca attention mechanism was constructed with yolov5 as the baseline network for comparative experiments. The only difference between the improved model and the baseline model is the introduction of Ca module. See Table 3 for the comparative analysis results of model performance.

Table 3. Comparative Analysis of Collaborative Attention Improvement

model	Parameter quantity (M)	mAP@0.5	Detection time (frame/ms)
YOLOv5s	13.6	0.987	40
YOLOv5s+CA	13.7	0.992	47

After the introduction of coordinate attention (CA) module, the number of model parameters increased moderately, and the mean value of accuracy (map) improved significantly. Compared with the unmodified benchmark network, the reasoning speed of the improved model is only slightly increased. This improvement effectively improves the recognition accuracy on the premise of maintaining the detection efficiency, so as to achieve a better balance between accuracy and speed, and can meet the real-time detection requirements of pineapple targets in dynamic scenes.

5.2 Lightweight Comparison of Network Models

In order to adapt to the limited computing resources of embedded hardware platforms such as picking robots and realize the deployment and real-time reasoning of neural network model, it is very important to improve the lightweight of target detection network. This study uses the lightweight network architecture shufflenetv2 to reconstruct the yolov5 algorithm, aiming to reduce the model complexity and computational overhead, while maintaining high detection accuracy. By replacing the backbone network of yolov5 with shufflenetv2, and using its efficient channel shuffle and group convolution operations, the number of parameters and floating-point operands (flops) are significantly reduced. The performance comparison analysis of the improved lightweight model in pineapple target recognition task is shown in Table 4.

Table 4. Comparison of Lightweight Backbone Network

model	Parameter quantity(M)	mAP@0.5	Detection time (frame/ms)
YOLOv5s	13.6	0.987	40
YOLOv5s shufflenetv2	0.34	0.864	31

The experimental results show that the Yolo algorithm based on shufflenetv2 lightweight backbone network reconstruction has achieved remarkable results in model compression. Specifically, the improvement greatly reduces the parameters of the model and effectively improves the reasoning speed. However, corresponding to this, the mean value of accuracy (map) of the model has declined to a certain extent. This reflects the typical trade-off between accuracy and efficiency in lightweight design.

5.3 Performance Analysis of Model Deploy-Ment

Due to the lack of native support for pytorch, tensorflow and other mainstream deep learning frameworks, k210 edge computing platform cannot directly deploy its training generated model. Therefore, it is necessary to convert the training model into a proprietary format (.Kmodel) supported by the platform through the special tool chain. In view of the performance limitations of embedded hardware and the precision loss and structural changes that may be introduced in the conversion process of each framework, this section will convert the mainstream target detection network models obtained based on the above research, deploy and evaluate the performance on the k210 platform, and systematically analyze the actual reasoning efficiency of different networks at the edge after conversion.

Table 5. Performance Data after Deployment of Different Model Transformation

Original model	Detection time (frame/ms)	Transformation model	Detection time (frame/ms)	Recognition accuracy
FRCNN.pt	98	FRCNN.kmodel	70	55.2%
YOLOv4.pt	51	YOLOv4.kmodel	42	53.9%
YOLOv5.pt	40	YOLOv5.kmodel	38	56.7%

Due to the limited hardware resources and computing power of k210 edge computing platform, it is impossible to directly deploy the mainstream target detection model trained on the standard computing platform. Therefore, the model needs to be transformed and quantified through the special tool chain before it can be adapted and deployed. This study uses maixpy ide to write test code and evaluate the accuracy of the actual model deployed in k210 after conversion. According to the analysis of the results in Table 5, the reasoning time of the model after conversion is generally shortened, which is mainly due to the reduction of parameters and calculation accuracy caused by the optimization operations such as model compression and quantification during the conversion process. At the same time, the recognition accuracy of the model has decreased significantly, which can be attributed to two main reasons: first, there are differences between the image quality actually collected by the camera and the training data; Second, the limited on-chip resources and computational accuracy of k210 restrict the expression ability of complex features of the model.

6. Conclusion

This paper designs a pineapple target detection system based on deep learning, and uses k210 embedded chip as the core processor to realize the real-time recognition of pineapple target. The system provides a feasible technical scheme for promoting agricultural intelligent operation. The research conclusions can be summarized as follows:

(1)Build a dedicated pineapple data set: through field collection, screening, data enhancement and manual annotation, a pineapple data set containing 8000 images is established, and is divided into training set, verification set and test set according to the standard proportion.

(2)Determine the optimal benchmark model: on the server platform, based on the same data set, conduct comparative experiments on three mainstream detection networks: faster r-cnn, yolov4 and yolov5. The results show that yolov5 performs best in terms of detection speed and mean precision (map), so it is selected as the benchmark model for subsequent research.

(3)The performance optimization and lightweight of the model are realized: by introducing the coordinate attention (CA) module in yolov5, the model map is improved to 0.992. Furthermore, the shufflenetv2 lightweight backbone network is used to replace the original structure, reducing the number of parameters by 13.26m, and improving the single frame image

recognition speed to 0.031 seconds, effectively taking into account the accuracy and efficiency.
(4)Completed the embedded deployment and verification: through model transformation, the optimized model was deployed on the k210 embedded platform. The experimental results show that the model can achieve effective recognition on the end-to-side device, but its reasoning performance is lower than that on the server platform.

Acknowledgements
This work was supported by Guangdong Province Higher Vocational Education Reform Research and Practice Project "Construction of the Immersive"2+2" Collaborative Educa-tion Curriculum System for Electrical Engineering and Automation" (GDJG2021-368), 2019 Campus level Talent Special Project of Lingnan normal university “Research on Sensor Angle Compensation and Noise Reduction Methods for Flow Method Production Measurement” [grant numbers ZL2005, ZL2007], College Student Innovation Program of Lingnan normal university “Fruit tree targeted precision variable fertilization robot based on YOLOv5 deep vision algorithm” [grant numbers X202410579025] , “Design of Spray Device Based on Laser SLAM and Adaptive Kalman Filter Contro” [grant numbers S202410579010] and "Zhiguang Youhu Pineapple Field"-Precision Pesticide Spraying System Driven by Laser SLAM and Adaptive Kalman Filtering [grant numbers S202510579004] .

References

[1]Carolina Trentin, Yiannis Ampatzidis, et al. Tree crop yield estimation and prediction using remote sensing and machine learning: A systematic review[J]. Smart Agricultural Technology, 2024.
[2]REN S O，HE K M，GIRSHICK R，t al. Faster R-CNN: towards real- time object detection withregion proposal networksJ. Advances in Neural Information Processing Systems,2015, 28: 91-99.
[3]Hanwen Kang,Chao Chen. Fruit detection, segmentation and 3D visualisation of environments in apple orchards[J]. Computers and Electronics in Agriculture,2020,171(C).
[4]Liu Anwen,Xiang Yang,Li Yajun,Hu Zhengfang, et al. 3D Positioning Method for Pineapple Eyes Based on Multiangle Image Stereo-Matching[J]. Agriculture,2022
[5]LI Xingxu, CHEN Wenbai, WANG Yiqun, et al.Design and experiment of an Automatic cherry tomato harvesting system Based on cascade vision detection [J]. Transac-tions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(1): 136-145.
[6]Girshick R , Donahue J , Darrell T ,et al. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation [J].IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 38(1):142-158.
[7]ZHENG T ,WANG X ,PENG X , et al.Survey of neurocognitive disorder detection methods based on speech, visual, and virtual reality technologies[J].,2024,6(06):421-472.
[8]SUN Jun, QIAN Lei, ZHU Weidong, et al. Apple detection in complex orchard environment based on improved RetinaNet[J]. Transactions of the Chinese Society of Agricultural Engineering(Transactions of the CSAE), 2022, 38(15): 314-322.
[9]GU Baoxing, Liu Qin, Tian Guangzhao, Wang Haiqing, Li He, Xie Shangjie. Tree trunk recognition and localization based on improved YOLOv3 [J]. Transactions of the Chinese Society of Agricultural Engineering, 202,38(06):122-129.

