Yield estimation of pomegranate based on deep learning YOLO model

haiying wang
Honghe University
xi peng
Honghe University
zhou huoyan
zhou.huoyan@163.com
Yunnan University

Article

Keywords: Agricultural Technology, Image Recognitiong, Precision Agriculture, Data Visualization, Crop Counting

Posted Date: August 30th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4810768/v1

Additional Declarations: No competing interests reported.

Yield estimation of pomegranate based on deep learning YOLO model

Haiying[1] Wang, Xi Peng[2], Huoyan Zhou[3,*]
[1]College of Biological Science and Agronomy, Honghe University, Mengzi 661199, China, 2927298519@qq.com;
[2]College of Biological Science and Agronomy, Honghe University, Mengzi 661199, China, 1302641701@qq.com;
[3]College of Ecology and Environment, Yunnan University, Kunming 650031, China, zhou.huoyan@163.com

ABSTRACT

Pomegranates from Mengzi City, Yunnan Province, are renowned for their nutritional value and flavor, boosting the local economy. Traditional yield prediction relies on manual sampling, which is costly and inefficient. This study proposes an intelligent prediction method using the YOLOv5 deep learning model.Mobile phones captured images of pomegranate trees from four directions to construct a dataset. These images were cropped and labeled for training the YOLOv5 model. The model achieved an 89% recognition accuracy and a 0.94 effectiveness score, accurately identifying and counting pomegranate fruits.Data visualization and model inference estimated the fruit quantity, with predicted values being 91.7% of the actual yield at a confidence level above 0.25. The average yield per tree was estimated at 103.1 kg, closely matching the actual 105.9 kg. This method offers robust data support for rapid and accurate yield estimation, outperforming traditional methods in efficiency and accuracy.

Keywords: Agricultural Technology, Image Recognitiong, Precision Agriculture, Data Visualization, Crop Counting

1. **Introduction**

1.1. Background and Significance of the Topic

Mengzi has a history of over 800 years of pomegranate cultivation, and its pomegranates have long been renowned (Wang Wei et al., 2021). Pomegranates from Mengzi, a gem of Honghe Prefecture, Yunnan Province, are cherished by consumers for their unique taste and have significantly boosted the local economy. Due to its unique geographical location, Mengzi pomegranate has gained international recognition, becoming a hallmark of Mengzi's agricultural culture. Accurate yield prediction before harvest is crucial for farmers, as it aids in the rational allocation and management of labor and orchard resources, reducing operational costs and facilitating subsequent sales. Thus, accurately estimating pomegranate yield is of great significance.

Traditional methods of pomegranate yield estimation primarily rely on manual sampling or farmers' planting experience. Manual prediction methods are costly, time-consuming, and lack accuracy and efficiency. Additionally, yield estimates based on farmers' experience cannot provide reliable data support, highlighting the need for a new, more precise method of yield estimation.

With advancements in intelligent technology, artificial intelligence has revolutionized traditional agriculture and is now an essential trend in modern agricultural development. Using deep learning methods to identify and count pomegranates and estimating yield based on the average weight of local pomegranate fruits in Mengzi offers an accurate, economical, efficient, and timely yield prediction approach. This also aids in estimating crop losses caused by natural disasters (Li Haolu, 2021). Given the proven accuracy and efficiency of deep learning for pomegranate recognition and counting, this study employs deep learning techniques to build a YOLOv5 target detection model. The model identifies pomegranates in pre-processed images, estimates the number of pomegranate fruits in the study area, and accurately predicts yield based on the average weight of local pomegranates. This method provides valuable data support for local pomegranate yield assessment.

1.2. Research Status at Home and Abroad

Currently, yield estimation technology for large-scale cash crops such as soybeans, wheat, and rapeseed has become relatively mature (Xu Haidong et al., 2011). In agriculture, crop yield is a standard measure of agricultural output harvested per unit area of land, making yield estimation a crucial component of agricultural production systems. Among the various crop yield estimation methods, field sampling surveys, spectral index-based estimation, agrometeorological models, and image-based estimation techniques are commonly employed (Xiong X, 2018). Additionally, both domestic and international scholars have applied deep learning methods to the detection and recognition of plant fruits, establishing a foundation for the application of artificial intelligence in yield estimation.

1.2.1. Field Sampling Survey Method

The main steps of the field sampling survey method are as follows: select some representative small plots in the field by group average sampling method or by area, harvest the rice after it is mature, and extract four factors related to yield per unit area, such as effective panicle number, spikelets per ear, seed setting rate, and 1000-grain weight. Then, input the data into the formula to calculate the field yield (Diao Caquan., 1994). Shi Donghong et al. proposed a new sampling survey method combining key sampling and measured yield to make the survey method more reasonable and ensure the objectivity and authenticity of the survey results. Detailed examples of sampling and investigation methods were given to address the defects of the current field sampling survey method in China (Shi Donghong., 2017). Although this method yields high precision data, it is time-consuming, labor-intensive, and inefficient, making it unsuitable for fruit yield estimation.

1.2.2. Crop Yield Estimation Method Based on Spectral Index

The development of remote sensing technology in crop yield estimation has become more stable (Zhao Chunjiang., 2014; Chen Zhongxin et al., 2016). Remote sensing yield estimation uses crop agronomic traits and multi-spectral cameras to obtain reflectance index information of crops in different spectral bands to build yield prediction models comprehensively (Cheng Zhiqiang et al., 2015). Yao et al. integrated remote sensing data and process model theory to estimate corn yield in the Northeast Plain, achieving a Pearson correlation coefficient (R) of 0.827 (Yao F et al., 2015). Wang Pengxin et al. estimated winter wheat yield using a combined model of leaf area index (LAI) and conditional vegetation temperature index (VTCI). They found that the R^2 value of the yield estimation result was 0.531 when assimilating VTCI and LAI simultaneously at dry farming sample points (Wang Pengxin et al., 2016). Although remote sensing technology is effective for large-scale crop yield estimation, it is susceptible to spatial resolution issues, leading to accuracy problems in smaller areas.

1.2.3. Research Status of Deep Learning Methods

YOLOv5 is a deep learning algorithm for object detection based on convolutional neural networks. It can detect various objects in images and classify and locate them. In agricultural crop detection and counting, domestic and international scholars have conducted extensive research on cash crops. In 2019, Gao Yunpeng proposed wheat ear detection methods based on YOLOv3 and Faster-RCNN, achieving recognition accuracies of 87.12% and 97.00%, respectively, with detection speeds of 0.12s and 0.94s per image (Gao Yunpeng, 2019). In 2020, He et al. used UAVs to acquire wheat data, built a wheat ear detection model through an improved YOLOv4 network, adjusted network parameters, and used a new prediction box regression method, achieving a recognition accuracy of 77.68% (He M X et al., 2020).

1.3. Research Purpose and Content

1.3.1. Research Purpose

With the rapid development of artificial intelligence, YOLOv5, as a highly efficient and adaptable detection model, can quickly adapt to different detection backgrounds while ensuring high accuracy. This can make repetitive agricultural work more efficient and convenient. Moreover, with the support of national policies, deep learning has gradually been applied to precision agriculture. To implement the new development concept of "promoting agricultural modernization and building an agricultural power" and realize intelligent online assessment of pomegranate production, this paper studies the identification and counting of pomegranates based on the YOLOv5 deep learning network model, aiming to provide new methods and ideas for pomegranate yield estimation.

1.3.2. Research Content

Image Data Sample Acquisition and Preprocessing: Collect pomegranate fruit images from the study area, use Imagestool and Labelimg to cut and annotate the image data, and divide the annotated image dataset into a training set and a validation set with a ratio of 8:2.

Model Inference and Estimation: Estimate the total number of pomegranate fruits in the study area through model inference. Build an intelligent auxiliary statistical model of the pomegranates in the images, and analyze the feasibility of the model application based on recall, precision, F1-score, MAP50, and other data from the model training and verification results.

Yield Estimation: Estimate the yield in the study area based on the average weight of Mengzi pomegranates.

This structured approach aims to provide accurate, efficient, and timely pomegranate yield estimation, supporting local agricultural practices and economic development.2. Materials and methods.

2 Method

2.1. Study area

Mengzi City, situated in the southeast of Yunnan Province, spans 23°01' to 23°34' north latitude and 103°13' to 103°49' east longitude. The city features a subtropical plateau monsoon climate, characterized by abundant light and heat resources. The Tropic of Cancer passes through the Mingying and Biceshai areas within the city. With a frost-free period of 337 days and an annual average of 2,234 hours of sunshine, Mengzi benefits from an advantageous low-latitude plateau red land and favorable climatic conditions, providing an optimal environment for pomegranate cultivation (Li Bingxiang, 2022). These conditions contribute to the superior quality of Mengzi pomegranates. Figure 1 illustrates the geographical location.

Fig.1

The software used in this study is shown in Table 1.

Table 1

2.2. Data Collection

2.2.1 Image Data Acquisition

The sampling site was located in a pomegranate orchard near Nanshantun, south of Honghe Avenue in Wenlan Town, Mengzi City, Yunnan Province. The pomegranate trees, aged 3 to 5 years, were sour pomegranates situated at latitude 23.34155°N and longitude 103.4097°E. The sampling area covered 300 square meters, with the sampling period spanning from September to November 2022.

A total of 644 images of pomegranate trees were collected using mobile phones, capturing the trees from four different directions. The images were first integrated and sorted, removing those with high overlap. Next, Imagestool was used to crop the images to a spatial resolution of 600x600 pixels. This process took 72 hours and resulted in 4,312 cropped images. These images were then divided into two sets: a training set and a validation set, in a ratio of 8:2. The images were sequentially numbered for identification(Table 2).

Table 2

The dataset is in JPEG format, with the target objects marked using Labelimg. The label for pomegranates is "Shiliu," and the label file format is TXT. The marked file names correspond to the images, and each line in a label file represents a target. The first parameter denotes the target category, the second and third parameters indicate the horizontal and vertical coordinates of the target's center point, and the last two parameters represent the width and height of the target box, respectively. The parameters are separated by spaces.

2.2.2. Fruit fresh weight collection

Randomly select 30 pomegranates, weigh each one individually to determine their fresh weight, and then calculate the average weight.

2.3. Experimental environment

The experiment was conducted using the Windows 11 operating system, the Pytorch framework, and an 11th Gen Intel(R) Core(TM) i5-115577 @ 2.50GHz graphics card for computation. Table 3 provides a detailed description of the configuration environment.

Table 3

2.4. Model building

YOLOv5 is one of the target detection algorithms based on regression, utilizing a multi-scale prediction method to simultaneously detect objects with image features of different sizes. Written in Python, YOLOv5 is easily portable to the Windows platform, facilitating the application of model products. There are four versions of YOLOv5: YOLOv5l, YOLOv5m, YOLOv5s, and YOLOv5x. The simplest of these is YOLOv5s, which was selected for this experiment due to its faster training and inference speed, suitable for the large data volume involved.

The YOLOv5 network structure comprises four main components: Input, Backbone, Neck, and Prediction(Fig.2).

Fig.2

The Input component includes Mosaic data enhancement, adaptive anchor frame calculation, and adaptive image scaling (He et al., 2016). Mosaic data enhancement zooms and crops four images together to add more small targets, making the background image more substantial. Adaptive image scaling adds a small amount of padding to the original image, scales it to 640 × 640 × 3, and then feeds it into the neural network.

The Backbone is mainly divided into the Focus module and CSP (Cross Stage Partial) Bottleneck structure (Wang et al., 2020). The Focus module slices the input image into four parts, splices these fractions to create a new structure with reduced size, and then performs a convolution operation to fuse information and change the feature map's channel number(Fig.3). CSP allows the output image to retain more network gradient information, reducing computation while maintaining network performance. Features from one level can be upsampled or downsampled as the output of the next level and merged with the feature map of the same size as the body part.

Fig.3

The Neck module utilizes the SPP (Spatial Pyramid Pooling) and FPN+PAN (Feature Pyramid Network + Path Aggregation Network) structures (Lin et al., 2017), and also incorporates the CSP2 module to enhance feature fusion capabilities. The SPP module increases the receptive field of backbone features and separates the most crucial features(Fig.4).

Fig.4

The Prediction module is divided into two parts: the Bounding box loss function and Non-Maximum Suppression (NMS) (Neubeck et al., 2006). The YOLOv5 algorithm uses GIOU_loss as its loss function to address the issue of non-coincident boundary frames. During the target detection stage, weighted NMS filters the numerous target boxes to obtain the optimal target box.

2.5. Model training

In the process of model training, to ensure the efficiency and rigor of model training, parameter batch-size is set to 16, epochs to 300 rounds, img-size to 640*640, and workers to 1. Training is the process of making the recognition accuracy gradually approach 1. When the training was iterated to the 123th round, The accuracy had leveled off and the training was terminated after 159 hours and 42 minutes of use.

2.6. Evaluation index

AP (Average Precision) represents the area enclosed by the Precision-Recall (PR) curve, with the recall rate on the horizontal axis and the precision rate on the vertical axis. By calculating the precision value for each recall value, the performance index for a single category is obtained. MAP

(Mean Average Precision) is used to evaluate the overall accuracy of the model across multiple categories (Zhu et al., 2022).

$$\text{Accuracy} = \text{precision} = \frac{TP}{(TP+FP)} \times 100\% \qquad (2\text{-}1)$$

$$\text{Recall rate} = \text{recall ratio} = \frac{TP}{(TP+FN)} \times 100\% \qquad (2\text{-}2)$$

$$AP = \sum_{i-1}^{n-1} (ri + 1 - ri) P_{inter}(r_{i+1}) \qquad (2\text{-}3)$$

$$mAP = \frac{(\sum_{i-1}^{k} APi)}{k} \qquad (2\text{-}4)$$

$$\text{Relative} = \frac{|\text{Measurd value} - \text{True value}|}{\text{True value}} \qquad (2\text{-}5)$$

Precision represents the proportion of correctly predicted positive samples to the total number of samples predicted as positive by the model. Recall represents the proportion of correctly predicted positive samples to the total number of actual positive samples. In this context:

- TP (True Positive) represents the number of positive samples correctly identified by the model.
- FP (False Positive) represents the number of samples incorrectly identified as positive by the model.
- FN (False Negative) represents the number of positive samples incorrectly identified as negative by the model.

$r_1, r_2, \ldots, r_n$ are the recall values corresponding to the first interpolation of the precision interpolation segment arranged in ascending order.

2.7. Estimating yield

Yield estimates are typically based on crop growth and climatic conditions or projections of yield per unit area. For pomegranates, deep learning YOLOv5 image recognition technology was used to estimate the number of fruits produced by each tree, which in turn was used to estimate the yield per tree.

The weight of a single Mengzi pomegranate ranges between 350-400g. The yield of each pomegranate tree is estimated based on the number of fruits identified through target detection..

3. Results

In the results, recall, precision, F1-score, MAP50, and other metrics of the training and validation outcomes were recorded. Additionally, the time taken for image cropping, labeling, and training was also documented.

3.1. Investigate data

3.1.1Fresh fruit weight

The average fresh fruit weight of 30 pomegranates was 367g (Table 4).

Table 4

3.1.2 The measured value of the quantity of pomegranate fruit

Taking 15 samples, the average actual number of pomegranates per tree was 289, and the actual average yield per tree was 105.9 kg (Table 5).

Table 5

The actual average number is 289,The actual average yield is 105.9 kg/ tree.

3.2. Target detection result

As shown in Figures 5 and 6, the training and validation results of the model are highly consistent. The accuracy rate of the training results is 88.6%, while the accuracy rate of the validation results is 88.7%. This close proximity to 90% indicates that the model's training results are feasible(Table 6).

Table 6

3.2.1 Visual Analysis of Training Result Data

To more clearly observe the data changes during the training process and better analyze the experimental results, the data visualization module in the YOLOv5 model was used. This analysis includes the relationship between recognition accuracy and confidence in the pomegranate training model, the relationship between the F1 score and confidence in the training model, the relationship between recall rate (R) and recognition accuracy (P), and the relationship between confidence and recall rate.

Fig.5

The visualization of the training result data for the pomegranate recognition model, shown in Fig. 8, depicts the regression loss data after 123 iterations of training. YOLOv5 uses GIOU loss as the loss function for the bounding box, and "Box" is speculated to be the mean value of the GIOU loss function (Li Fulu et al., 2022). In the figure, box-loss represents the loss function for predicting the location of the anchor frame, with its loss rate mostly between 0 and 0.04, indicating the model's high predictive ability for the anchor frame's location. The mean of classification loss is 0. Obj-loss represents the loss function for target detection, with the mean value ranging from 0 to 0.060. Precision represents the proportion of correctly predicted positive samples among all positive samples. After the fourth iteration, the value stabilizes between 0.84 and 0.88. Recall indicates the probability of the positive class being correctly identified by the model, stabilizing around 0.87 after the fourth iteration. The average precision, represented by mAP, is a comprehensive index to measure the model's training effect. When the model is iterated 25 times, map-0.5 is mostly between 0.89 and 0.94(Fig.6). The results show a good regression effect, with the classification accuracy reaching 0.943, proving the method's effectiveness.

Fig.6

The specific relationship changes are shown in the figure above(Fig.5), illustrating the relationship curves among parameters such as confidence, recall rate, accuracy rate, and F1 score of the training model (Tsung Y L et al., 2008). The identification of pomegranates was studied, and the model parameters were analyzed, resulting in the data curve for Shiliu. In Fig.7 (a), when the confidence is 0.908, the accuracy remains stable after reaching 1. In Fig.7 (b), 0.439 is the key point. F1 scores increase with the rise in confidence up to 0.439, peaking at 0.88, then showing a downward trend. However, the general recall rate remains between 0.2 and 0.6, with the F1 score stabilizing around 0.86, showing minimal fluctuation. The area enclosed by the curve and the horizontal and vertical coordinates in Fig.7(c) reflects the performance of the model classifier. In Fig.7 (d), the recall rate reaches 0.99, indicating excellent performance by the training model. In summary, the training model achieves good results when the confidence is around 0.87, with a recognition effect of 0.943.

Fig.7

3.2.2. Model reasoning

The trained model is deduced from this data set, using a picture of a pomegranate tree in one of the directions as the original. The number of pomegranates identified and estimated under different confidence levels is compared with the true value, as shown in FIG. 9, which respectively represents the inference recognition effect under the confidence levels of 0.25, 0.5, and 0.75, and the corresponding estimated number of pomegranates is 77, 64 and 26. 84 pomegranate fruit trees in this picture are obtained through counting. The predicted number of pomegranates is closest to the true value when the confidence level is above 0.25, so the number of pomegranates in other environmental contexts is estimated based on the confidence level above 0.25.

3.3. Estimated Yield

By taking comprehensive photographs of pomegranate trees in the study area, approximately 608 images of pomegranate fruits were obtained. The average fresh weight of Mengzi pomegranate fruit, determined by weighing, was 367g per fruit. Using this information, the approximate yield of a single tree was estimated. As shown in the table below(Table 7), an average of 281 pomegranate fruits were detected per tree, while the actual average number was 289. The measured value is slightly less than the actual value, leading to an estimated yield of 103.1 kg per tree compared to the actual yield of 105.9 kg. The predicted yield value is close to the actual value, indicating that the model has a certain feasibility in pomegranate yield estimation.

The average estimated number of pomegranates per tree was 281, and the average estimated yield per tree was 103.1 kg. In comparison, the actual average number of pomegranates per tree was 289, with an actual average yield of 105.9 kg. The relative error in the number of pomegranates per tree was 2.8%, and the relative error in yield per tree was 2.6%. The relative error for each tree was less than 3%, demonstrating that the model has high accuracy and a certain level of feasibility.

Table 7

4. Discussion

4.1. Model Performance Evaluation

From the model training results, the recall rate reached 0.877, and the recognition accuracy reached 0.886, indicating that the model has good training outcomes. The verification results also show that the accuracy, recall rate, and mAP50 perform well in the validation set, proving that the regression process of the YOLOv5 network is smooth and that the model training is successful. From the model inference results, when the confidence level is above 0.25, the estimated number of pomegranates is closest to the true value, reaching 86%. When the confidence level is above 0.5, the estimated number of pomegranates accounts for 71% of the true value. This indicates that the pomegranate number estimation method based on the YOLOv5 deep learning network model is feasible to a certain extent. However, when the confidence is 0.75, only 26 pomegranates can be identified in the same picture background. Under this parameter, the pomegranate fruit in the image must be complete before it can be identified. The YOLOv5 confidence threshold should not be too high or too low; more targets can be identified with a smaller confidence threshold, closer to the actual number of pomegranates, but it will also bring more false detections. Finding the right balance of accuracy and detection rate to adjust the confidence threshold is critical.

4.2. Research Contribution and Practical Application

Traditional yield prediction methods rely on manual sampling, which is costly, inefficient, and time-consuming. To solve this problem and enable intelligent and efficient online prediction of pomegranate yield, this study proposed a method to quickly and accurately identify and count ripe pomegranates in their natural growing environment. By using the trained deep learning network model to identify and infer from captured and organized images, the approximate number of pomegranates can be determined, and the yield of each tree can be estimated based on the weight of a single pomegranate. This method lays the foundation for the intelligent automatic picking of pomegranates in the future. In practical applications, farmers only need to use their mobile phones to take pictures for identification and counting, place the images into the "data" - "images" folder in the model directory, and run the "detect.py" weight file in the model to quickly estimate the number of pomegranates in each picture, greatly reducing manpower and material resources. Results can be obtained efficiently and accurately.

4.3. Research Limitations and Prospects

Mengzi pomegranates generally bloom from May to June, with fruit growth from July to August, and the peak harvest time is typically in September. The data set collection time was in October, indicating that farmers had already picked some fruit before we collected data, resulting in the final count being less than the actual quantity. If we advance the data collection time to before the pomegranates are harvested, or further analyze the optimal data collection time, the yield results could be closer to the actual yield.

Deep learning networks have high requirements for software and hardware. In this experiment, the 11th Gen Intel(R) Core(TM) i5-1155G7 @ 2.50GHz processor was used for calculations, which greatly reduced the model training speed. With the support of a graphics processor (GPU), a lot of model training time could be saved, speeding up the experimental process.

To ensure pomegranate yield and regulate water and soil balance, farmers often press the branches, causing some pomegranates to hang close to the ground, making it difficult to distinguish fallen fruit. Moreover, in this study's deep learning network model, only the location of the pomegranates was identified. If fine-grained bird species identification technology (Weng Yuchen, 2018) is applied to pomegranate identification and estimation, distinguishing between good and bad fruit and whether the fruit has fallen can greatly reduce the inference data set sorting time, accelerate the experimental process, and reduce the model's false detection rate. This pomegranate fruit identification and estimation method may also become more applicable.

In this experiment, image acquisition was done manually, consuming a lot of time. The distribution of pomegranates is not simply in the canopy, and the branch-pressing planting method causes pomegranates to layer over each other, leading to certain errors in the estimated yield obtained from current planar recognition. If UAV visual stereo technology (Liu Xiaofei et al., 2024) is introduced, specific volume and more accurate yield calculations could be achieved. Strengthening relevant research and samples later could improve the model's accuracy and better guide production practices. This method aligns with the current robotic arm fruit-picking technology used in smart agriculture, further promoting intelligent agricultural development.

5. Conclusion

To realize efficient and intelligent online assessment of pomegranate yield and provide farmers with more accurate, convenient, and timely yield information before harvest, this study proposes an intelligent online prediction method based on the YOLOv5 deep learning model. The goal is to optimize traditional methods' long time, high cost, and low precision issues. Experiments show that:

The algorithm in this study has a pomegranate recognition accuracy of 89% and an identification effect of 0.94, accurately identifying and counting pomegranates. The number of pomegranates identified in different environmental backgrounds under confidence parameters of 0.25, 0.5, and 0.75 was 77, 64, and 26, respectively, compared to the true count of 84 pomegranates. When the confidence is above 0.25, the predicted number of pomegranates is closest to the true value, meeting actual requirements.

According to the single fruit weight of Mengzi Pomegranates, with an average weight of 367g per fruit and 15 pomegranates selected as samples, the final estimated average number per tree is 281, and the estimated average yield per tree is 103.1 kg. In comparison, the actual average number per tree is 289, with an actual average yield per tree of 105.9 kg. The relative errors of quantity and yield per pomegranate tree were 2.8% and 2.6%, respectively, both lower than 3%, indicating that the model has high accuracy and certain feasibility.

The algorithm used in this study can quickly and accurately identify the number of pomegranates, providing data sources for pomegranate yield estimation and offering new ideas and technical methods for pomegranate yield estimation research.

Author Contributions: Conceptualization, Huoyan Zhou.; Methodology, Haiying Wang.; Validation: Haiying Wang, and Huoyan Zhou; formal analysis, Haiying Wang.; investigation, Xi Peng, Huoyan Zhou, Haiying Wang; writing—original draft preparation, Haiying Wang.; writing—review and editing, Huoyan Zhou; supervision, Huoyan Zhou; project administration, Huoyan Zhou.; funding acquisition, Huoyan Zhou. All authors have read and agreed to the published version of the manuscript.”.

Funding: This research was funded by The Yunnan Postdoctoral Program(ynbh23010)

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Wang Wei, ZHANG Wentao, Li Yao.Study on the development of the Mengzi pomegranate industry from the perspective of geographical indication [J]. Rural Science, Technology and Economy, 2021, 32(13): 200-202.

LI Haolu.Research on Remote sensing recognition and yield prediction of cotton fields in Xinjiang based on Deep learning [D]. Nanjing University of Information Science and Technology, 2021.

XU Haidong, MA Wei, TAN Or et al.Tea bud estimation method based on YOLOv5 deep learning [J]. Journal of China Agricultural University, 2011, 27(12): 213-220. (in Chinese)

XIONG Xiong.Research on ear segmentation and lossless yield prediction of paddy rice based on deep learning [D]. Wuhan: Huazhong University of Science and Technology, 2018.

DIAO Caquan.Various Treatises on crop cultivation: Southern version [M].Beijing: China Agriculture Press, 1994.

Shi Donghong. (in Chinese)Research on sampling survey method of agricultural yield [J]. Journal of Liaoning Normal University: Natural Science Edition, 2017, 19(2): 98-101. (in Chinese)

ZHAO Chunjiang.Advances in agricultural Remote sensing research and application [J]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(12): 277-293.

Chen Zhongxin, REN Jianqiang, TANG Huajun, et al.Progress and prospect of agricultural remote sensing research and application [J]. Journal of Remote Sensing, 2016, 20(5): 748-767. (in Chinese)

Research progress and prospect of crop yield estimation model [J]. Chinese Journal of Eco-Agriculture, 2015, 23(4): 402-415.

Yao F, Tang Y, Wang P, et al. Estimation of maize yield by using a process-based model and remote sensing data in the Northeast China Plain [J]. Physics and Chemistry of the Earth, Parts A/B/C, 2015, 87(1): 142-152.

Wang Pengxin, Sun Huitao, Xie Yi, et al.Yield estimation of winter wheat based on LAI, VTCI, and particle filter assimilation algorithm [J].Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(4): 248-256. (in Chinese with English abstract)

Research on-ear detection method of field wheat based on deep neural network [D]. Beijing Forestry University,2019.

He M X, Hao P, Xin Y Z. A robust method for wheatear detection using UAV in natural scenes[J]. IEEE Access, 2020, 8: 189043-189053.

Li Bingxiang.Research on pest control technology of Mengzi pomegranate [J]. Southern Agriculture, 2022, 16(10): 22-24. (in Chinese)

HE K, ZHANG X, RENS, et al Deep Residual Learning for Image R ecognition [J]. IEEE Conference on Computer Vision Pattern Recognition, 2016: 770-778.

WANG C Y, LIAO H Y M, WU Y H, et al CSP-Net: A New Backbone that Can Enhance Learning Capability of CNN [C]//2020IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 14-19 June 2020, Seattle, WA, USA. 1571-80.

LIN T Y, DOLLáR P, GIRSHICKR, et al Feature Pyramid Networks for Object Detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 21-26 July 2017, Honolulu, HIUSA. 936-44.

NEUBECK A, GOOL L V. Efficient non-maximum suppression[C]// Proceedings of 18th International Conference on Pattern Recognition, Piscataway: IEEE Press, 2006: 850-855.

ZHU Hao, Shunyong Z, Yalang C, et al. Wood surface defect detection algorithm based on improved YOLOv5s[J]. Wood science and technology.(in Chinese)

Tsung Y L, Priya G, Ross G, et al. Focal loss for dense object detection[J]. IEEE Transactions onPattern Analysis ＆Machine Intelligence, 2008, 42(2): 318-327.

LI Fulu, Ji Zhe, Duan Xiusheng.Dynamic target recognition and tracking method based on YOLOv5 Deep Learning model [J].Journal of Shijiazhuang Railway University (Natural Science Edition), 2022, 35(3): 111-117. (in Chinese)

Tan Shilei, Bi Xiongbo, Lu Gonglin, et al.Real-time detection of human mask-wearing based on YOLOv5 network model [J]. Laser Journal, 2021, 42(2): 147-150. (in Chinese)

Weng Y C. Fine-grained bird recognition based on deep learning model [D]. Beijing Forestry University,2018.

LIU Xiaofei, Mingjie, YU Jia.Simulation of ship navigation risk recognition by UAV stereo vision [J]. Computer Simulation, 2024, 41(2): 29-33. (in Chinese)

Figures

Figures 1

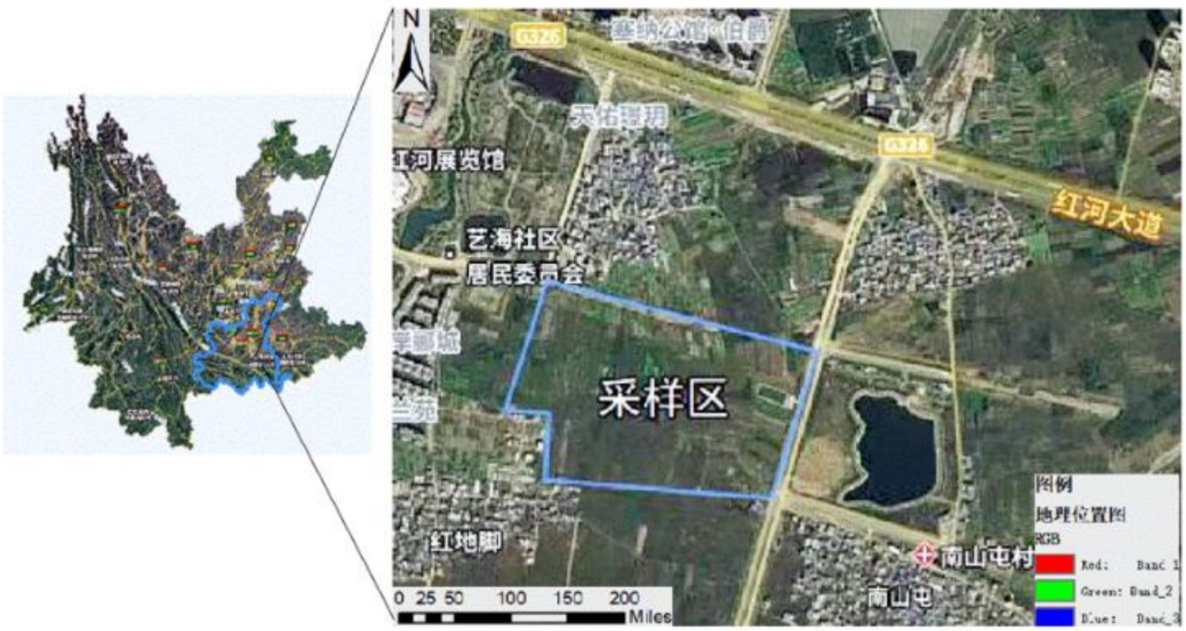


Fig.1. Geographical location map of the study area

Figure 2

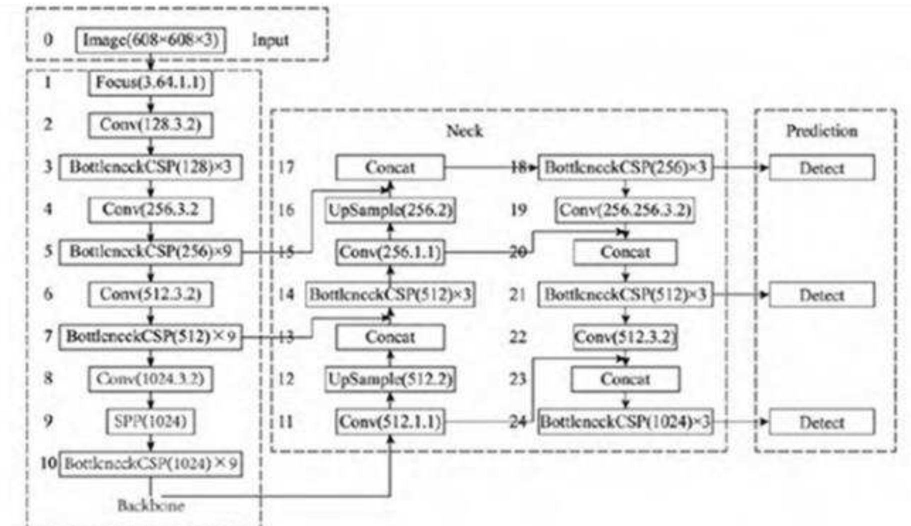


Fig. 2. Network structure of YOLOv5

Figure 3

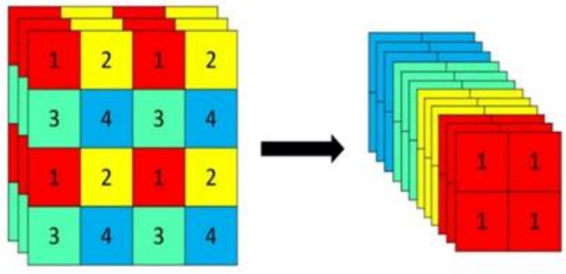


Fig. 3. Focus slicing operation

Figure 4

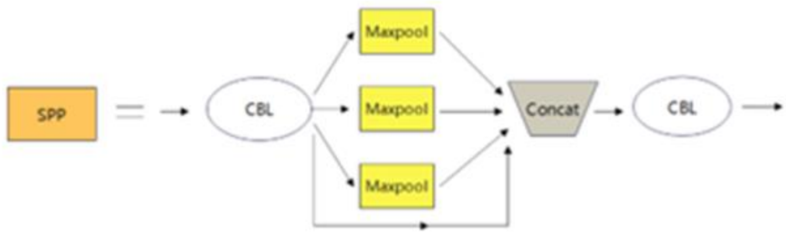
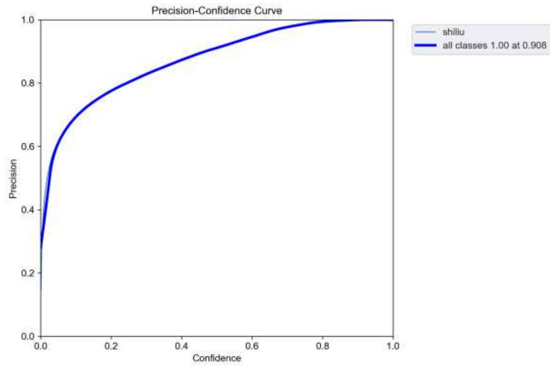
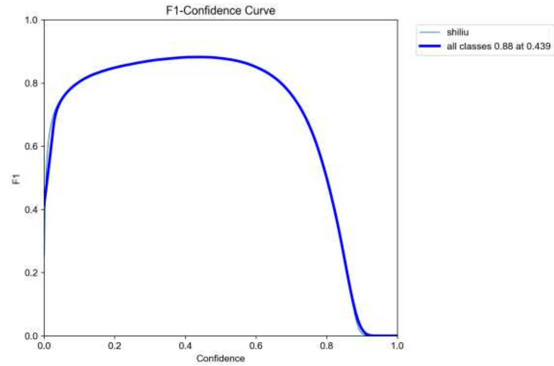


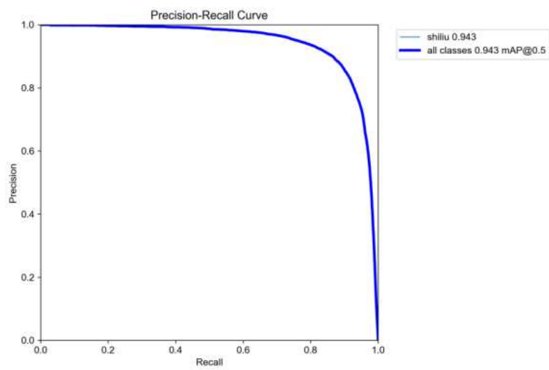
Fig. 4. SPP network structure

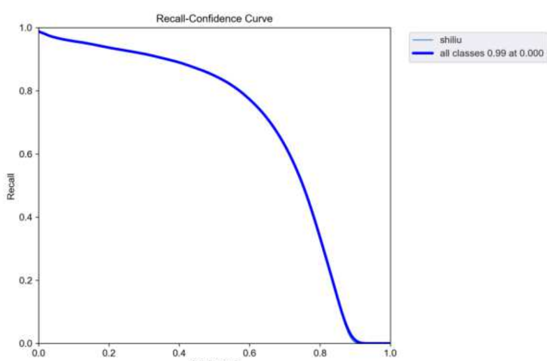
Figure 5





(a) Accuracy P - confidence curve **(b) F1 score - Confidence curve**





(c) Recall rate R - accuracy P curve **(d) Confidence - recall rate**

Fig. 5. Model parameter feature curve

Figure 6

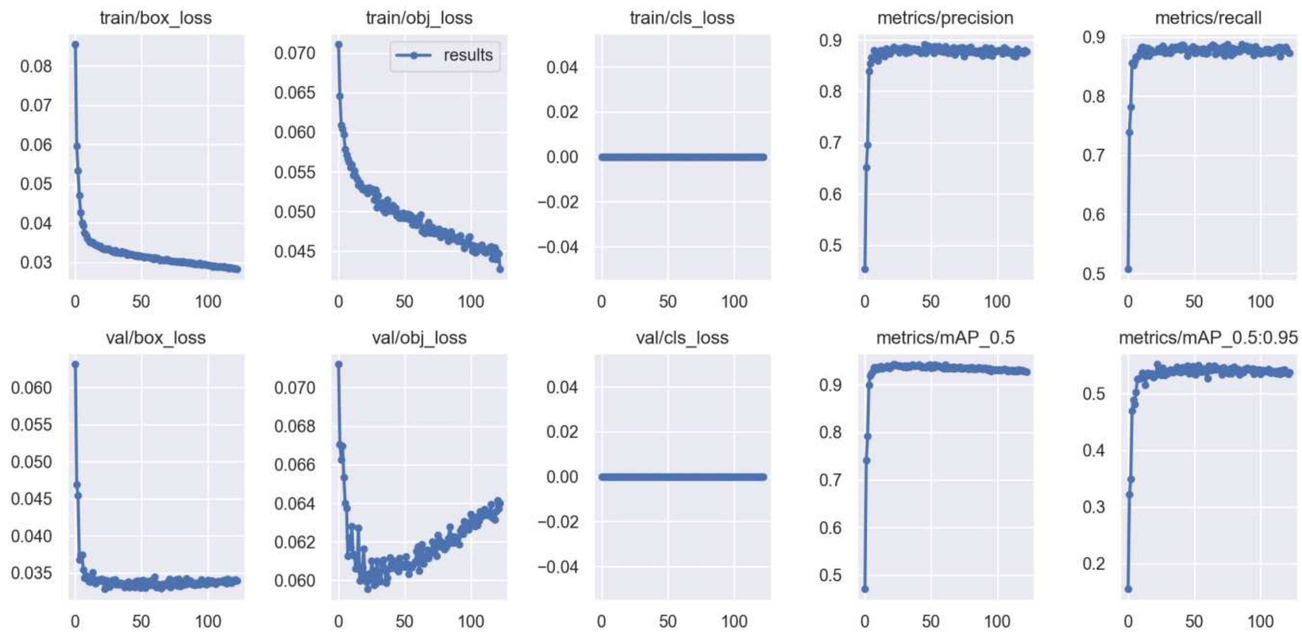
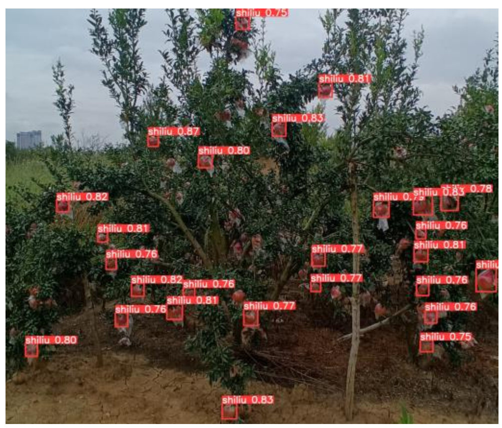


Fig. 6 Shiliu frame regression loss in 123 training iterations

Figure 7

(a)Confidence of more than 0.25 **(b) confidence of more than 0.5**



(c) confidence of more than 0.75

Fig. 7. Comparison of model inference effect

Tables

Table 1

Table 1 Software used and its functions

Configuration	Function
Images tool	Crop the picture
Labelimg	Mark the picture with a pomegranate
Python3.8	Data enhancement and chart drawing through programming
Pycharm	As an environment IDE for python
11th Gen Intel(R) Core(TM) i5-1155G7 @ 2.50Ghz	The CPU was used for model training

Table 2 Division of data sets

argument	Parameter number
Total training set	4312
Training set	3234
Validation set	1078

Table 3 Computing environmental configuration

	Disposition
CPU	11th Gen Intel(R) Core(TM) i5-1155G7 @ 2.50Ghz
System environment	Windows11
Language	Python3.8
Pytorch	1.8.2

Table 4 Statistics of fresh weight of pomegranate

Serial number	Fresh weight /g	Serial number	Fresh weight /g
1	394.5	16	357.9
2	325.0	17	338.1
3	384.3	18	313.3
4	329.8	19	431.9
5	338.7	20	370.9
6	442.9	21	369.5
7	444.2	22	332.3
8	347.9	23	362.5
9	419.8	24	374.2
10	336.0	25	347.5
11	360.4	26	337.8
12	359.4	27	357.8
13	356.8	28	337.9
14	365.8	29	371.4
15	430.8	30	372.3
Average fresh weight /g	367.0		

Table 5 Statistical table of the actual results of some pomegranate trees

Serial number	1	2	3	4	5
One side of the recognition effect					
Actual quantity (PCS)	404	292	422	340	460
Actual yield (kg/ tree)	148.3	107.2	106.0	124.8	168.8
Serial number	6	7	8	9	10
One side of the recognition effect					
Actual quantity (PCS)	307	252	265	351	170
Actual yield (kg/ tree)	112.7	97.5	97.3	128.8	62.4
Serial number	11	12	13	14	15
One side of the recognition effect					
Actual quantity (PCS)	185	149	302	238	191
Actual yield (kg/ tree)	67.9	54.7	110.8	87.3	70.1

Table 6 Results of training and validation of the model

	P	R	MAP-0.50	MAP-0.5:0.95
Training result(%)	88.6	87.7	94.3	55.2
Verification result(%)	88.7	87.7	94.3	55.2

Note: P - accuracy; R - Recall rate; MAP-0.50 - refers to the average accuracy when the intersection ratio (IOU) threshold is 0.5; Map_0.5 :0.95 - Refers to the average accuracy as the intersection ratio (IOU) threshold increases progressively between 0.5 and 0.95.

Table 7 The estimated results of some pomegranate trees

Serial number	1	2	3	4	5
One side of the recognition effect					
Estimated number (PCS)	386	286	424	344	478
Estimated yield (kg/ tree)	141.7	105.0	155.6	126.2	175.4
Serial number	6	7	8	9	10
One side of the recognition effect					
Estimated number (PCS)	308	240	266	351	147
Estimated yield (kg/ tree)	113.0	88.1	97.6	128.8	53.9
Serial number	11	12	13	14	15
One side of the recognition effect					
Estimated number (PCS)	183	141	279	222	160
Estimated yield (kg/ tree)	67.2	51.7	102.4	81.5	58.7