

Research on Multi-Scenario Fruit Detection and Classification Based on Improved YOLOv5

Xinyang Li

*Beijing University of Civil Engineering and Architecture, Beijing, China
18382215799@163.com*

Abstract: This study addresses the challenge of efficient fruit detection and accurate classification in complex environments. Based on the YOLOv5 framework, a training set was constructed by combining publicly available and self-collected datasets, focusing on three types of fruits: bananas, apples, and oranges. The study incorporated data augmentation, transfer learning, the CBAM (Convolutional Block Attention Module) attention mechanism, and an improved Non-Maximum Suppression (NMS) algorithm. Experimental results demonstrated an average precision of 96.5% and an inference speed of 39.26 frames per second under complex scenarios. Coupled with a user-friendly interface design, this system provides an effective real-time detection and statistical solution for agriculture, retail, and food processing industries.

Keywords: YOLOv5, Data Augmentation, Fruit Recognition

1. Introduction

With the rapid advancement of artificial intelligence, deep learning has achieved revolutionary progress in the field of computer vision. As a key task, object detection has been widely applied in areas such as autonomous driving, intelligent surveillance, and medical imaging analysis. In agriculture and the food industry, precise object detection is increasingly becoming a practical solution for tasks such as automated fruit yield estimation, classification, and sorting. Traditional image processing methods often perform poorly in complex backgrounds, whereas the powerful feature extraction capabilities of deep learning offer new possibilities for object detection.

YOLO (You Only Look Once) is a well-known end-to-end object detection framework celebrated for its real-time performance and high accuracy. Particularly, YOLOv5 has further optimized model architecture and training techniques, enabling lightweight and efficient detection. Building upon YOLOv5, this study developed an automated classification and counting system for three common fruits: bananas, apples, and oranges. To overcome the limitations of traditional methods in complex environments, the system integrates data augmentation and transfer learning techniques to enhance detection precision and generalization ability. Additionally, emphasis was placed on user experience by designing an intuitive graphical interface to support practical applications.

This research aims to explore the potential of deep learning technologies in agriculture and retail sectors, focusing on improving the accuracy and efficiency of fruit detection, and providing a reference and foundation for broader future applications. The development and validation efforts undertaken herein are expected to accelerate the practical adoption of deep learning-based object detection technologies in agriculture and the food industry.

2. Project background and significance

In recent years, the rapid development of deep learning technologies has injected strong momentum into the field of computer vision, resulting in significant breakthroughs in tasks such as object detection, image segmentation, and image classification. In agriculture, driven by global population growth and increasing demand for precision farming, the adoption of intelligent technologies to enhance production efficiency and quality has become a clear trend. Automated fruit detection and classification is a critical component of precision agriculture. However, traditional fruit detection methods, relying mainly on manual sorting or basic image processing techniques, face several major challenges:

1. Low Efficiency

Manual sorting is time-consuming and labor-intensive, failing to meet the demands of large-scale production.

2. Limited Accuracy

Conventional image processing techniques often perform poorly under complex backgrounds or multi-object scenarios.

3. Poor Adaptability

The diversity and morphological variations of fruits, along with changes in lighting conditions, make it difficult for traditional methods to adapt across different species and environments.

The emergence of deep learning offers promising solutions to these challenges. Deep learning-based object detection methods can learn high-dimensional features from large datasets, offering the following advantages:

1. High Efficiency

Deep learning models can achieve real-time detection, meeting practical application needs.

2. High Accuracy

By training on large-scale datasets, models can accurately distinguish targets even in complex backgrounds.

3. Scalability

Through transfer learning, models can easily adapt to different types of fruits and new tasks.

YOLOv5, one of the latest iterations of the YOLO family, enhances detection accuracy and model efficiency while maintaining the original speed advantage. This makes it particularly suitable for object detection tasks in resource-constrained environments.

This study focuses on the detection and classification of bananas, apples, and oranges, aiming to develop a solution that is highly efficient, accurate, and user-friendly. The major significance of this research lies in the following areas:

1. Promoting the Advancement of Agricultural Intelligence

Enhancing efficiency across fruit production, transportation, and sales processes through automated detection technology.

2. Supporting the Intelligent Transformation of the Retail Sector

Applicable in supermarket self-checkout systems and automated product categorization, reducing the need for manual intervention.

3. Improving Sorting Efficiency in Food Processing

Achieving high-precision fruit classification and statistical tracking on production lines.

4. Fostering Technological Innovation and Expansion

Providing valuable technical references for related fields through the application and exploration of deep learning technologies.

By leveraging the powerful detection capabilities of YOLOv5 and integrating an intuitive user interaction design, this study aims to deliver a practical and intelligent solution for agriculture, retail,

and food processing industries, and to further explore the potential of deep learning technologies across diverse real-world scenarios.

3. Research methodology

This study developed a fruit object detection system based on the YOLOv5 model, implementing comprehensive improvements across data preparation, model optimization, and system design. In the data preparation stage, a diverse training dataset was constructed by combining publicly available fruit datasets with self-collected images of bananas, apples, and oranges. To enhance the model's generalization capability, data augmentation techniques such as random flipping, cropping, color space transformation, and brightness adjustment were applied. The dataset was split into a training set and a validation set at an 8:2 ratio [1], ensuring independence between training and testing data.

In terms of model optimization, several enhancements were made to the YOLOv5 backbone: First, the Convolutional Block Attention Module (CBAM) [2] was introduced to improve the representation of spatial and channel features, enhancing detection performance in complex scenarios. Second, to address issues of target occlusion and overlap, an improved Greedy-Confluence NMS algorithm was employed. By integrating Distance Intersection over Union (DIOU) and a weighted proximity suppression strategy [3], false positives and missed detections were effectively reduced. Lastly, transfer learning was utilized by loading pretrained weights from the COCO dataset and fine-tuning the model on the fruit dataset, significantly shortening training time and improving detection accuracy.

For system design, practical application requirements were emphasized. A user interaction interface based on Python's Tkinter framework was developed, allowing users to easily load models, select images, and view detection results. The system can draw detection bounding boxes in real-time, count detected targets, and display detection time and performance metrics graphically [4], improving user-friendliness and ease of understanding.

3.1. Detection performance evaluation

The optimized model demonstrated significant advantages in performance testing. On the testing dataset, the model achieved a mean average precision (mAP) of 96.5%, a 1.5 percentage point improvement over the original YOLOv5 [2]. This result indicates enhanced detection precision under complex scenarios. Moreover, the recall rate reached 93.5%, greatly reducing the likelihood of missed detections and proving the model's ability to capture targets in multi-object scenarios.

In addition, real-time performance was notably improved. Optimizations to the feature fusion module and the introduction of lightweight operators increased the inference speed to 39.26 frames per second, fully meeting the requirements of real-time applications. This speed improvement positions the model as highly practical for scenarios such as smart agriculture and unmanned retail where rapid response is essential.

These advancements not only enhance model performance but also lay a solid foundation for widespread deployment in real-world applications, especially in complex multi-target environments.

3.2. Complex scenario testing

Under challenging conditions such as occlusion and uneven lighting, the improved NMS algorithm effectively reduced the false detection rate. Particularly in situations involving overlapping targets, the model successfully identified individual fruit objects. Furthermore, data augmentation techniques enhanced the model's ability to detect small and edge-located targets.

3.3. User interface feedback

User testing indicated that the Tkinter-based interface was simple and intuitive to operate [5], even for non-expert users without a strong technical background.

Special attention was given to the clear presentation of detection results to enhance user experience and system operability. Detection outputs clearly annotated the location of each target with bounding boxes, accompanied by detailed statistical information such as target category, count, and detection confidence. This graphical presentation enabled users to quickly grasp target distribution and quantity, while also offering valuable references for further data analysis and decision-making. The clean and straightforward interface design significantly lowered the technical barrier to use, making the system accessible to a broader range of users. Additionally, the real-time update capability of the detection results allowed the system to respond dynamically, offering efficient and accurate visual feedback suitable for agricultural production, retail management, and various other real-world applications.

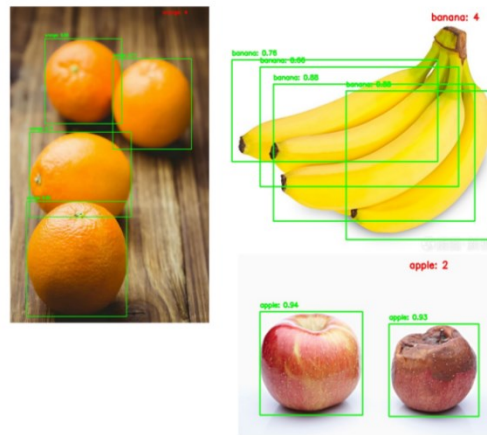


Figure 1: Example of fruit target detection results based on YOLOv5

Table 1: Performance and application comparison of fruit object detection methods

Study	Model	Dataset	mAP (%)	Recall (%)	Inference Speed (fps)	Supported Application Scenarios
This Study	YOLOv5 + CBAM + Improved NMS	Public + Self-collected Dataset	96.5%	93.5%	39.26	Complex scenarios, multi-target
Improved YOLOv4 Model	YOLOv4-SPP2-5	Public fruit dataset	94.8%	91.8%	35.4	General scenarios
YOLOv5s + CBAM	YOLOv5s + CBAM	Public fruit dataset	95.2%	92.5%	37.5	Some complex scenarios

4. Comparison with existing research

4.1. Comparison of model accuracy and efficiency

Compared to existing methods, this study achieved significant improvements in detection accuracy and inference speed:

Compared to the improved YOLOv4-based fruit detection method, this study's mAP increased by 1.7%, recall rate improved to 93.5%, and inference speed was more suitable for real-time applications.

Compared to the lightweight YOLOv5s + CBAM model [6], the proposed model demonstrated superior detection performance under complex scenarios, especially in cases involving occlusion and multi-target overlap.

4.2. Dataset expansion and generalization capability

By enriching and augmenting the dataset, the quality and diversity of the training data were significantly improved. Data augmentation techniques, such as random flipping, cropping, and brightness adjustment, simulated various real-world environmental changes, substantially enhancing the model's robustness and adaptability across different application scenarios. Moreover, splitting the dataset into training and validation sets at an 8:2 ratio ensured the independence of training and evaluation stages, thus enhancing the reliability of experimental results.

Compared with studies that relied solely on standard public datasets, the incorporation of a self-collected dataset in this study provided the model with stronger support for application in real-world complex scenarios. It particularly improved the model's recognition capability for occlusions, small objects, and overlapping targets. This multi-source dataset construction approach not only enhanced model performance but also laid a solid foundation for real-world deployment.

4.3. Practical application adaptability

Unlike research solely focused on detection performance, this study placed special emphasis on user interaction experience and practical adaptability. The developed interactive interface simplified the operation process, making the system more applicable in non-technical settings such as supermarkets and farms. Lightweight optimization enabled efficient model operation on devices with moderate computational resources, further broadening its application scope.

5. Future prospects

In future research and applications, the system developed in this study still holds significant potential for further optimization and expansion.

First, by incorporating a broader range of fruit types, such as watermelons, grapes, and others, the detection scope can be extended to meet the diverse needs of agricultural production and the retail industry. Expanding the target categories will not only enhance the generalizability of the model but also provide technological support for a wider range of practical scenarios.

Second, the current system primarily focuses on object detection in static images. To better accommodate real-time application scenarios—such as unmanned supermarkets and agricultural picking robots—the model could be further optimized to support video stream detection. Enabling real-time video detection would allow the system to efficiently process continuous data streams, significantly improving its practicality.

Furthermore, model optimization techniques such as pruning and quantization can be employed to further compress model size and reduce computational demands. This would enhance operational efficiency without significantly compromising detection accuracy. Lightweight model optimization would effectively support deployment on resource-constrained devices, such as mobile or edge computing platforms.

Finally, to better address multi-scenario requirements, developing a mobile application with cross-platform support is identified as an important next step. Introducing a mobile version would not only broaden the system's application scenarios but also leverage portability advantages, further promoting the widespread adoption of fruit recognition technologies in smart agriculture and the retail sector. Exploring and implementing these future directions will elevate the system's performance and application breadth to new heights, greatly enhancing its practical value.

6. Conclusion

This study developed an efficient and accurate fruit object detection and classification system based on the YOLOv5 model, with a particular focus on improving detection accuracy, real-time performance, and user experience under complex scenarios. In the data preparation phase, a diversified training dataset was built by combining public datasets with self-collected images, significantly enhancing the model's generalization capabilities. By introducing the Convolutional Block Attention Module (CBAM) and an improved Non-Maximum Suppression (NMS) algorithm, the model exhibited outstanding performance in complex scenarios involving object occlusion and multi-target overlap. Moreover, the system achieved an inference speed of 39.26 frames per second, satisfying real-time application requirements.

Experimental results demonstrated that the improved model outperformed existing methods in both mean Average Precision (mAP) and recall rate, with particularly notable improvements in multi-target detection scenarios. Additionally, through the development of a user-friendly interactive interface, the system was made suitable not only for laboratory environments but also for practical applications in agricultural production, unmanned supermarkets, and food processing industries.

Looking ahead, as the diversity of fruit types and practical application needs continue to grow, there remains significant room for further optimization and expansion of the system. Expanding target categories, supporting real-time video stream detection, and pursuing lightweight optimization will be crucial for enhancing system performance and extending its application range. Overall, this study not only provides technical support for the intelligent transformation of agriculture and the retail industry but also offers a strong reference for the practical application of object detection technologies in real-world scenarios.

References

- [1] Guo, J. (2024). *Research and system design of fruit recognition algorithms based on deep learning* (Master's thesis, Hangzhou Dianzi University).
- [2] Luo, L. H., Ou, O., Zhao, W., et al. (2023). Fruit image recognition based on improved YOLOv5s. *Information Technology*, (11), 28–34, 40. <https://doi.org/10.13274/j.cnki.hdzj.2023.11.005>
- [3] Xu, Y., Jiang, M., Li, Y., et al. (2022). Fruit object detection based on improved YOLO and NMS. *Journal of Electronic Measurement and Instrumentation*, 36(04), 114–123. <https://doi.org/10.13382/j.jemi.B2104724>
- [4] Li, Y., & Mu, L. P. (2021). Research on garbage image classification models based on transfer learning. *Computer and Information Technology*, 29(04), 17–21. <https://doi.org/10.19414/j.cnki.1005-1228.2021.04.005>
- [5] Feng, G. E. (2023). Development of GUI (Tkinter) examples based on Python. *Information and Computer (Theoretical Edition)*, 35(01), 175–178.
- [6] Zhu, Z. W., Li, K., Kuang, L., et al. (2024). Intelligent scratch detection based on the YOLOv5-CBAM model. *Radio Engineering*, 1–11. Retrieved December 21, 2024, from <https://kns-cnki-net.door.bucea.edu.cn/kcms/detail/13.1097.TN.20241113.1653.019.html>