

Lightweight Recognition Method for Korla Pear Based on NanoDet-Plus

Dechuan Qiu, Yongfu Shan *, Yingchao Wang, Lihao Qin and Na Li

School of Information and Engineering, Xinjiang College of Science & Technology, Korla Xinjiang 841000, China

* Corresponding author: Yongfu Shan

Abstract: This study aims to tackle the challenge of identifying Korla pears in real orchard environments by developing a highly accurate and lightweight target detection model. This model is designed to assist with automated harvesting and yield estimation. We improved the feature extraction capabilities of the NanoDet-Plus framework by incorporating MobileNetV3 as the backbone network. Additionally, we used the DBSCAN clustering algorithm to segment and compile a high-quality dataset of pears. The improved model achieved mAP@0.5 and mAP@.5:.95 scores of 97.9% and 85.1% on the test datasets, representing improvements of 0.3% and 0.5% over the original model. After post-quantization compression, the model's size was reduced by 94.2%, allowing it to perform stable detection on devices with limited computing power. This network balances high accuracy with low resource consumption, making it suitable for recognizing pears in complex orchard environments and providing a practical lightweight solution for smart agriculture applications.

Keywords: Object Detection; Korla Pear; NanoDet-Plus.

1. Introduction

The Korla Pear, a flagship specialty fruit of southern Xinjiang, is not only protected under China's Geographical Indication system but also a pillar of the local agricultural economy. Korla, the heartland of pear cultivation with a 40% share of the region's total planting area, has positioned the pear industry as a key driver for boosting agricultural productivity, increasing farmers' incomes, and strengthening urban branding. As agricultural mechanization advances, automated fruit-picking technology has become a critical step in smart agriculture[1]. The accuracy and efficiency of target recognition, which is the first step in automation, are essential for these systems to work well. However, in real orchard settings, identifying pears is quite challenging: they are spread throughout the tree canopy, often hidden by dense foliage and branches; their green skin blends almost perfectly with the background leaves, and local weather conditions, such as frequent sandstorms, have complicated the recognition process.

2. Research Status at Home and Abroad

The main goal of research on Korla pear recognition is to accurately count yield per mu, addressing the inefficiency and errors of current manual methods. Traditional image processing and machine learning techniques often struggle with limited feature representation and robustness, making them unsuitable for real-world use. Deep learning, especially object detection technology, presents a more promising approach. This can be divided into two types of algorithms: two-stage models like Faster R-CNN[2], which are good at detecting small targets and handling complex scenes but are computationally intensive and slow, making them less ideal for urgent agricultural tasks; one-stage models such as YOLO[3,4,5], with their lightweight design, efficient feature extraction, and end-to-end detection, offer much faster performance with little loss in accuracy, significantly

reducing resource use and becoming the preferred choice for automated orchard detection.

Recent advances in this field have confirmed the effectiveness of this approach. For example, Liu[6] used an improved YOLOv5 model for apple detection. Tan[7] tackled the challenges of unstructured orchard environments by enhancing the YOLOv8n architecture, incorporating innovative features such as simplified residual modules, simSPPF, and PConv, which greatly boosted detection accuracy and speed. Shao [8] addressed the complexities of growth environments, including fruit occlusion, lighting variations, and dense arrangements, by developing an improved YOLOv7 model that simultaneously enhances pear and stem detection, achieving fast real-time performance and significantly reducing fruit damage caused by mechanical handling. A comprehensive review of existing research indicates that, due to the variability of orchard environments and diverse task requirements, the key to improving the real-time performance and adaptability of Korla pear recognition systems lies in creating datasets that accurately reflect real-world conditions and employing lightweight detection models suitable for hardware-efficient deployment. This strategy not only addresses the specific needs of different regional orchards but also paves the way for the industrial application of automated Korla pear harvesting technology.

3. Datasets

3.1. Dataset Source and Characteristics

This study constructs a dataset sourced from a Korla fragrant pear orchard, capturing the complex visual characteristics of real-world fruit-growing environments (see Fig.1(a)). The images represent diverse target features, including variations in shape, size, spatial distribution, and maturity, while also presenting challenges such as occlusion from leaves and branches and signs of disease. These conditions introduce significant variability that mirrors real detection scenarios. Collected using standard mobile phones, the high-resolution images (up to 4096×2736 pixels) provide

a solid foundation for developing robust and generalizable object detection models tailored to agricultural applications.

3.2. Input Adaptation and Image Tiling Strategy

The high resolution of the original images creates a compatibility challenge for the NanoDet-Plus model, which is designed for smaller input sizes (320×320 pixels or 416×416 pixels). Simply changing the input layer would not only disrupt pretrained weights but also increase the model’s complexity and slow down inference. To solve this, we use a practical image tiling strategy that allows seamless adaptation without changing the model architecture. This approach ensures efficient feature use and computational performance, making it especially suitable for lightweight deployment in resource-limited environments like mobile or edge devices.

3.3. High-Resolution Annotation Workflow and Sample Generation

To efficiently handle large volumes of high-resolution data, we design a scalable annotation workflow that combines manual precision with automation. As shown in Fig.2, the process starts with initial annotation in LabelMe, followed by automated clustering using the DBSCAN algorithm and the creation of padded bounding boxes to assist with image cropping. Sub-images are then refined through secondary annotation, and all labels are exported in a standard format. This approach balances annotation speed and accuracy, enabling high-throughput dataset generation with minimal human effort. The final result includes 5,591 annotated pear instances, standardized in resolution and label format, providing a high-quality resource for model training and evaluation.



Fig 1. Comparison of Xinjiang Pear Images Before and After Segmentation

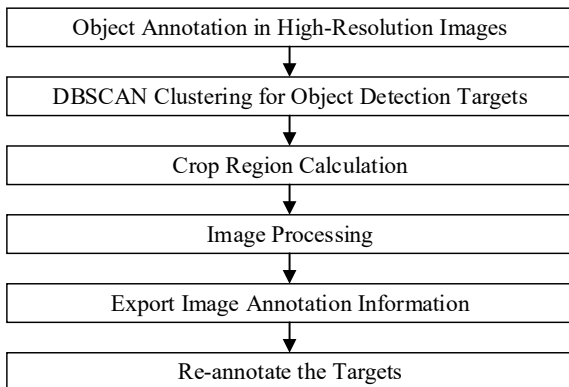


Fig 2. Re-annotation Process of Fragrant Pear Datasets

4. Construction of Object Detection Model for Pears

4.1. Nanodet-Plus

NanoDet-Plus is a lightweight, single-stage, anchor-free detection model designed for real-time object detection on mobile devices. It achieves excellent inference speed and high detection accuracy through innovative features. The model includes an Assignment Guidance Module (AGM) and a Dynamic Soft Label Assigner (DSL), which address the challenge of optimal label assignment common in lightweight architectures. It also features a Ghost-PAN structure that improves multi-scale feature fusion by efficiently reusing features. Supporting multiple input resolutions such as 416×416 pixels and 320×320 pixels, NanoDet-Plus offers flexible configuration options, making it suitable for deployment across various computational environments.

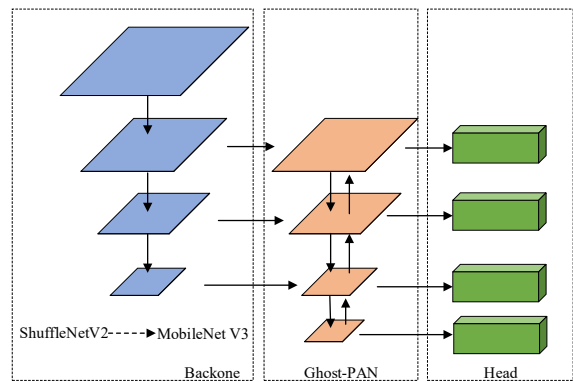


Fig 3. Improved NanoDet-Plus Model Architecture

4.2. Model Improvements

NanoDet-Plus originally uses ShuffleNetV2[9] as its backbone, utilizing depthwise separable convolutions and channel-shuffling mechanisms to greatly improve spatial feature retention while keeping low memory access overhead. This design is especially beneficial for extracting high-resolution features from small objects. Additionally, ShuffleNetV2 avoids excessive group convolutions, which helps reduce feature fragmentation and prevents the loss of detailed small-object information in deeper network layers. Alternatively, MobileNetV3[10], Google’s third-generation lightweight convolutional neural network, offers two versions—Large and Small—to meet various computational resource needs and deployment scenarios. When detecting fragrant pears, the decision between ShuffleNetV2 and MobileNetV3 as the backbone depends on their ability to capture fine details of small objects and their suitability for edge computing. Due to the small size and delicate features of fragrant pears, ShuffleNetV2 often faces limitations, leading to missed or false detections because of insufficient feature extraction. Therefore, this study chooses MobileNetV3-Large as the core architecture because of its better balance between computational efficiency and detection accuracy, as shown in Fig. 3. MobileNetV3-Large, optimized through Neural Architecture Search (NAS), features hardware-aware low-latency design. It uses a dual enhancement strategy: an improved Squeeze-and-Excitation (SE) module and the h-Swish activation function. The SE module dynamically adjusts channel weights to emphasize important features, while the h-Swish function increases non-linear ability without adding computational load. These

components work together to boost the model's sensitivity to small, detailed features like those of fragrant pears. Furthermore, MobileNetV3's flexible network width adjustment allocates computational resources based on how many objects are in the image, optimizing parameter use without losing detection accuracy. This adaptability is especially useful for edge deployments in agriculture, allowing for real-time detection while maintaining low power use and hardware costs during long-term orchard monitoring. Overall, these features create a good balance between computational efficiency and detection accuracy, making it an effective lightweight solution for fragrant pear datasets.

5. Experiment

5.1. Experimental Environment and Datasets Preparation

All models in this study were trained using an NVIDIA GeForce RTX 3060 GPU, supported by CUDA 11.7 and PyTorch 1.13.1. The batch size was set to 16, and training was carried out over 300 epochs. For model optimization, we used the AdamW optimizer with an initial learning rate of 0.001 and a weight decay of 0.05 to effectively manage model complexity. To stabilize training during the early stages, a linear warm-up strategy was employed, gradually increasing the learning rate from 0.0001 to the target rate over the first 500 iterations. The datasets used for systematic evaluation were divided into training, validation, and test sets at a ratio of 6:2:2.

5.2. Model Training and Performance Evaluation

Model evaluation was performed every 10 epochs using two key metrics: mean Average Precision at an Intersection over Union (IoU) threshold of 0.5 (mAP@0.5), and the stricter metric mAP@0.5:0.95, calculated across multiple IoU thresholds (ranging from 0.5 to 0.95 in increments of 0.05). As shown in Fig. 4, both metrics indicated steady

convergence of the model after approximately 100 epochs. Specifically, mAP@0.5 assesses average precision based on predictions overlapping at least 50% with ground-truth bounding boxes, providing a baseline for model accuracy. In contrast, mAP@0.5:0.95 offers a comprehensive evaluation across varying overlap thresholds, more effectively reflecting model performance in demanding real-world detection scenarios.

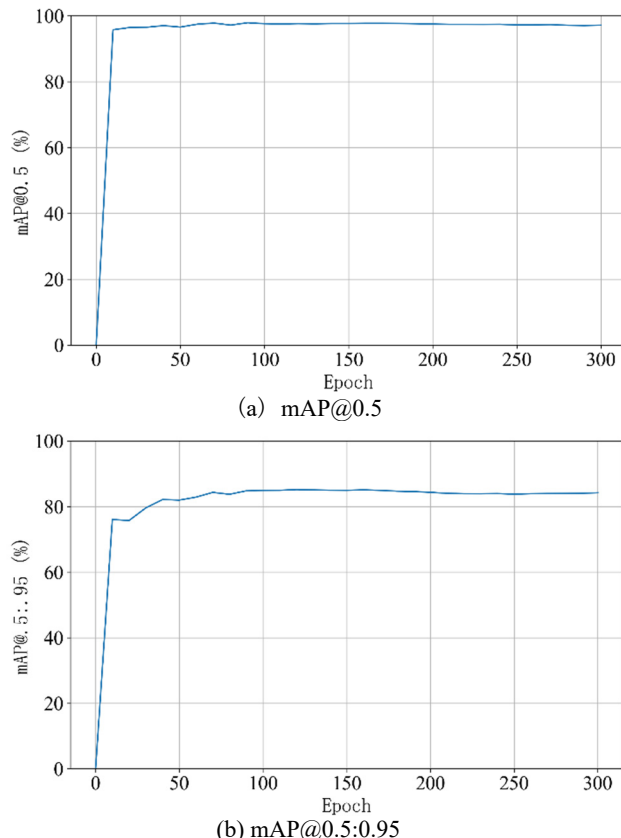


Fig 4. Changes in mAP of the Proposed Model Over Training Iterations

Table 1. mAP Comparison of Various Models

Model	mAP/%	
	mAP@0.5	mAP@.5:.95
NanoDet-Plus-m-1.5x	97.6	84.6
YOLO v5s	96.3	84.6
YOLO v5n	96.3	83.3
Our Model	97.9	85.1

For comparison, we selected the largest variant of NanoDet-Plus (NanoDet-Plus-m-1.5x) and two lightweight YOLO v5 series models (YOLO v5s and YOLO v5n) as benchmarks. Experimental results (Table 1) show that our proposed model achieved a mAP@0.5 score of 97.9%, an improvement of 0.3 percentage points over NanoDet-Plus, and a mAP@0.5:0.95 score of 85.1%, increasing by 0.5 percentage points. This performance improvement mainly comes from replacing ShuffleNetV2 with MobileNetV3-Large, equipped with SE attention modules, which significantly boosts detection accuracy for small, detailed targets without sacrificing real-time inference speed. Despite dataset limitations in scale and target diversity, the proposed method consistently outperformed YOLO v5 models, confirming that optimized lightweight backbones greatly enhance robustness in difficult detection environments.

5.3. Model Compression and Deployment Testing

The best-performing trained model was exported into the ONNX intermediate format, a standard for cross-platform model interchange, ensuring seamless transition across frameworks such as PyTorch and NCNN. Subsequently, the ONNX model was compiled into NCNN-specific formats (param for network architecture and .bin for weights) to meet lightweight inference needs on mobile and embedded devices. Model compression effectively reduced the model size from 118 MB to just 6.79 MB (about 5.8% of its original size). Further optimizations, including operator fusion and graph simplifications, were applied to improve inference efficiency. The compressed model was successfully integrated into an application supporting multi-source data inputs (images,

videos, and live camera feeds). As shown in Fig. 5, stable and efficient real-time inference performance was achieved on resource-constrained hardware, such as the Snapdragon 625 processor. This practical deployment confirms the suitability and robustness of NanoDet-Plus-based solutions in limited-resource scenarios, laying a strong technical foundation for future deployment on various low-power embedded platforms, including Jetson Nano and Raspberry Pi.

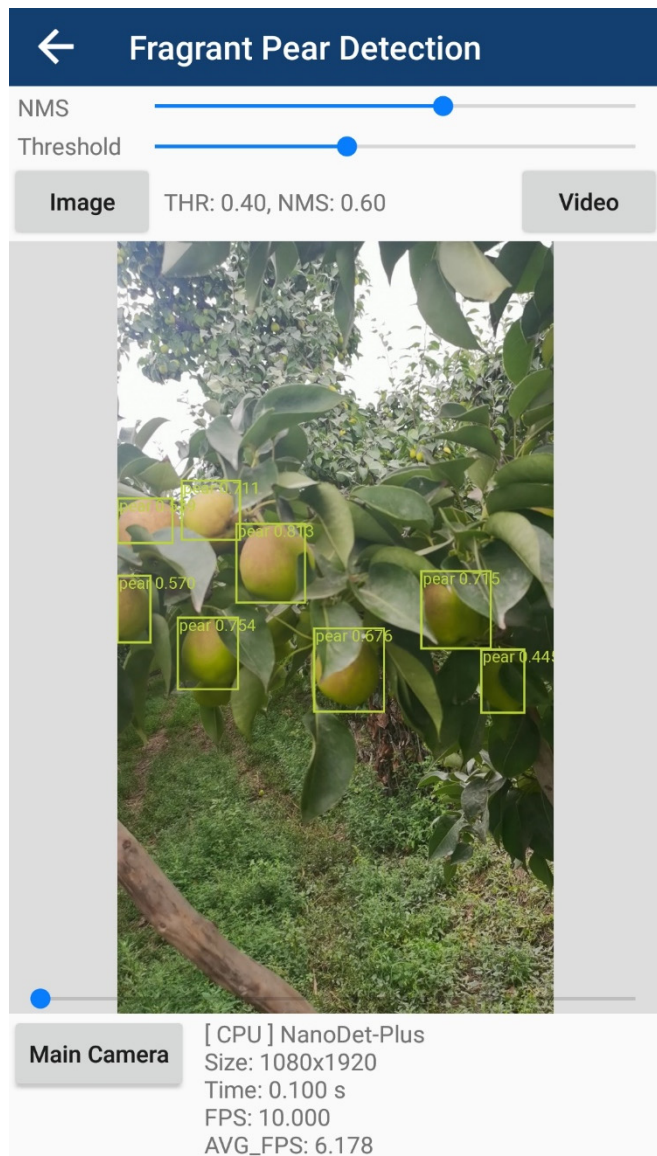


Fig 5. Interface of the Fragrant Pear Detection App

6. Conclusion

Advancing automated fruit harvesting technology is crucial for modernizing and making agriculture smarter. As a unique agricultural product from Xinjiang, the accurate automatic identification of Korla fragrant pears has important implications for local agricultural industrial upgrades. To overcome the limitations of traditional recognition methods, which struggle in complex orchard environments due to low accuracy and limited adaptability, this research introduces an

improved lightweight detection algorithm based on NanoDet-Plus.

By combining MobileNetV3 as the backbone network and applying advanced model compression strategies, this method greatly improves pear detection in difficult orchard conditions. Compared to existing detection methods, the proposed model delivers better detection accuracy on fragrant pear datasets while significantly reducing computational requirements. Upgrades to the backbone network and feature extraction modules specifically enhance the model's suitability for mobile and edge-computing deployment.

Future research will focus on improving adaptability across various orchard environments, advancing coordinated robotic picking systems, creating multimodal datasets (including multispectral imaging, 3D point clouds, etc.) that cover the full growth cycle of fragrant pears to improve generalization, and continuously refining algorithms to boost detection accuracy and real-time performance. These developments will support the practical implementation of automated orchard harvesting solutions.

References

- [1] Feng Zhang, Zhiqin Jiang, Xiaoguang Chen, et al. The Development Factors of *Pyrus sinkiangensis* 'Korla Xiangli' Industry [J]. Chinese Agricultural Science Bulletin.2021,Vol. 37 (No. 34), p.159-164.
- [2] Ren S , He K, Girshick R ,et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
- [3] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Seattle, WA, 2016: 779-788.
- [4] Redmon J , Farhadi A YOLOv3: An Incremental Improvement [J]. arXiv e-prints, 2018.
- [5] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications[EB/OL]. (2022-09-07) [2024-05-07].
- [6] Liu T, Kang H, Chen C. ORB-Livox: A real-time dynamic system for fruit detection and localization[J].Computers and Electronics in Agriculture, 2023:209.
- [7] Housen Tan, Wenhong Ma, Yuan Tian, et al. Improved YOLOv8n object detection of fragrant pears[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, Vol.40(No.11),p. 178-185.
- [8] Mingming Shao, Liping Zhang, Weiqiang Zheng, et al. Picking recognition of fragrant pear and fruit stem based on improved YOLOv7[J]. Journal of Chinese Agricultural Mechanization, 2025, Vol.46 (No.6),p.128-135.
- [9] N N Ma, X Y Zhang, H T Zheng, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design [M] //Anon. Lecture notes in computer science. Cham: Springer International Publishing, 2018:122-138.
- [10] Howard A, Sandler M, Chu G, et al. Searching for MobileNetV3 [J]. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, 1314-1324.