Development of an Automatic Banana Harvesting System in Orchard Environments
果園環境中自動化香蕉採收系統的開發

Phuoc Bao Long Do [1], Duc Tai Nguyen [1], Shih-Hung Huang [2], Jyun-Syong Fan [2], Wei-Chih Lin [1]
杜福寶龍 [1]、阮德泰 [1]、黃世宏 [2]、范俊雄 [2]、林韋至 [1]

[1] Department of Mechanical and Electro-Mechanical Engineering, National Sun Yat-Sen University
[2] Taiwan Banana Research Institute
[1] 國立中山大學機械與機電工程學系
[2] 財團法人台灣香蕉研究所

Abstract

The banana industry faces persistent challenges due to the manual, costly, and labor-intensive nature of harvesting. To address this, we propose an integrated robotic system combining a custom 3-DoF RRP arm with a lightweight vision model, Banana-DL. Built on YOLOv8n, Banana-DL achieves real-time detection of banana bunches and stalks with 96.3% precision and 94.5% mAP50 in a compact 3.7 MB framework. The arm's design, optimized for varying orchard conditions, is controlled via inverse kinematics to execute precise cutting. Field trials validate the system's reliability, highlighting that joint optimization of manipulator design and perception is essential for effective agricultural automation.

Keywords: Banana harvesting, robotic system, deep learning, lightweight detection, agricultural automation

摘　要

香蕉產業長期面臨採收作業仰賴人工、成本高且勞力密集等挑戰。為解決此問題，本研究提出一套整合式機器人系統，結合自製三自由度 RRP 型機械手臂與輕量化影像辨識模型 Banana-DL。該模型基於 YOLOv8n 架構開發，能以僅 3.7 MB 的精簡模型實現即時香蕉果串與花軸偵測，其精確率達 96.3%，在 IoU=0.5 條件下的平均精確率（mAP50）為 94.5%。經實地試驗系統的穩定性與可靠性，結果顯示，機械手臂結構與視覺辨識系統的整合設計，對於提升農業自動化作業的精度與效率具有關鍵作用。

關鍵詞：香蕉採收、機器人系統、深度學習、輕量化辨識、農業自動化

Introduction

Bananas are one of the most widely consumed fruits globally, with production exceeding 127 million tons in 2020 [1]. Despite this scale, banana harvesting is still predominantly performed manually. Farmers typically cut banana stalks with hand tools, a process that is labor-intensive, physically demanding, and increasingly constrained by the shortage of skilled agricultural workers. Furthermore, the average weight of a banana bunch is between 30 and 50 kilograms [2]. Manual handling of this weight can result in mechanical damage to the fruit, which compromises its quality and commercial value. With labor costs rising and rural populations aging, there is a pressing need for automation in banana harvesting [3].

Robotic harvesting systems face two major challenges: (1) accurate detection of banana bunches and stalks under orchard conditions, and (2) reliable execution of the cutting process. The orchard environment presents significant visual complexity, as banana stalks share similar colors with leaves and are frequently occluded by foliage. Additionally, variations in lighting conditions due to sunlight, clouds, and shadows affect detection accuracy. In addition, changing environmental weather conditions such as sunlight, clouds, and shade greatly affect detection results [4]. On the mechanical side, robotic arms must approach and cut stalks precisely to avoid damaging the fruit. Recent advances in deep learning have improved object detection in agriculture, enabling robust recognition of fruits and cutting points. However, many detection models remain computationally heavy, limiting real-time deployment on embedded robotic platforms. Furthermore, multi-stage detection pipelines often increase system complexity and risk of cumulative errors.

To address these challenges, we developed a complete banana harvesting system that combines an optimized lightweight detection model with a mobile robot platform. The system employs a depth camera for image acquisition, a Banana-DL detection model for bunch and stalk localization, and a 3-DoF robotic arm for stalk cutting. This integration provides a streamlined pipeline from perception to action, enabling real-time harvesting in natural orchard conditions.

Materials and Methods

I System overview

Figure 1 presents an overview of the proposed banana harvesting robot and its deployment environment. As shown in Fig. 1(a), the system is built on an autonomous guided vehicle (AGV) platform measuring 3020 mm in length and 1600 mm in width, equipped with an anti-tip mechanism to maintain stability on uneven terrain. A lifting mechanism is installed at the center of the platform to support the Intel RealSense D435i camera and the 3-DoF cutting robot arm. The end-effector of

the arm is positioned above a collecting basket that receives the harvested bunches, while an electronic control box houses the embedded computing unit and power modules. This integrated structure enables the robot to perform perception, planning, and cutting in a compact configuration suitable for orchard operations. Fig. 1(b) illustrates the typical deployment scenario in banana fields, where harvesting routes are narrow (2.5–3 m wide) and surrounded by dense vegetation. The terrain is uneven and often covered with grass, which poses challenges for both locomotion and vision-based detection. The compact dimensions and tracked mobility of the robot allow it to operate reliably under these conditions, ensuring stable navigation and precise harvesting performance.

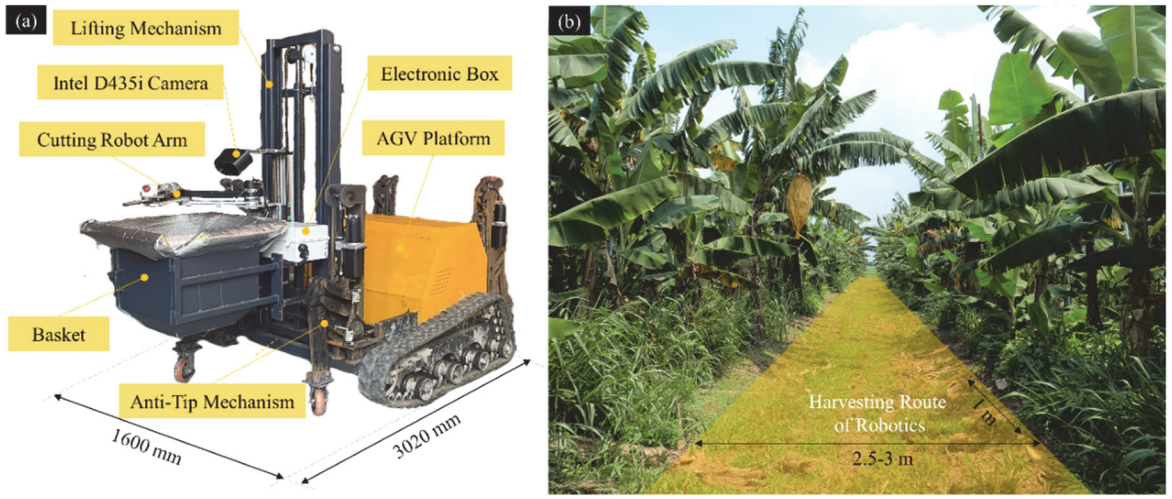


Figure 1 Overview of the proposed banana harvesting system

II Data acquisition and preparation

The dataset for this study was collected in a banana orchard at the Taiwan Banana Research Institute, Pingtung, Taiwan, during the period from May 2024 to July 2024. An Intel RealSense D435i depth camera (Intel Corp., USA) was mounted on the robotic system described in Figure 1, enabling high-resolution image capture of banana bunches and stalks from different perspectives. Image acquisition was performed between 9:00 a.m. and 5:00 p.m. under diverse weather conditions, including sunny, cloudy, and overcast days, in order to enrich dataset diversity. All images were stored in JPG format with a resolution of 1280 × 720 pixels. A total of 980 images were collected. Representative examples of the collected data are shown in Figure 2, where banana bunches are often wrapped in protective bags, partially occluded by leaves, and vary in appearance depending on their growth stage and orchard arrangement.

To generate ground-truth labels, the dataset was annotated using Labelme software version v1.8.6. Bounding boxes were assigned to two classes: banana bunches (‘b’) and banana stalks (‘s’).

For the stalk class, the cutting point was defined as the center of the diagonal intersection of the bounding box. Annotation files were saved in YOLO format. The dataset was divided into training, validation, and testing subsets at an 8:1:1 ratio.

Figure 2　Sample images of banana stalks and bunches collected in orchard environments

III　Detection model: Banana-DL

You Only Look Once version 8 nano (YOLOv8n) represents the strong generation in the YOLO family, optimized for object detection, segmentation, and classification tasks. Compared to YOLOv5, it integrates the C2f module for stronger feature extraction and adopts a decoupled head to separately optimize classification and localization, thereby improving detection accuracy in real time [5]. Among its variants, YOLOv8n is the most suitable for embedded deployment due to its lightweight design, but it still suffers from limitations such as redundant convolutional operations and suboptimal feature fusion in complex orchard scenes. To address these issues, we propose three enhancements tailored for banana detection. First, standard convolutions are replaced with group-shuffle convolutions (GSConv) to reduce parameters while preserving efficiency. Second, the original C2f block is upgraded into a C2f-Fast Efficient Channel Attention module, allowing the network to emphasize informative regions [5]. Finally, the neck structure is redesigned with a BiFPN, which strengthens multi-scale feature aggregation while reducing computational load [6]. The overall architecture of the proposed Banana-DL model is shown in Figure 3, illustrating the integration of these lightweight yet effective modules for robust detection in agricultural environments.

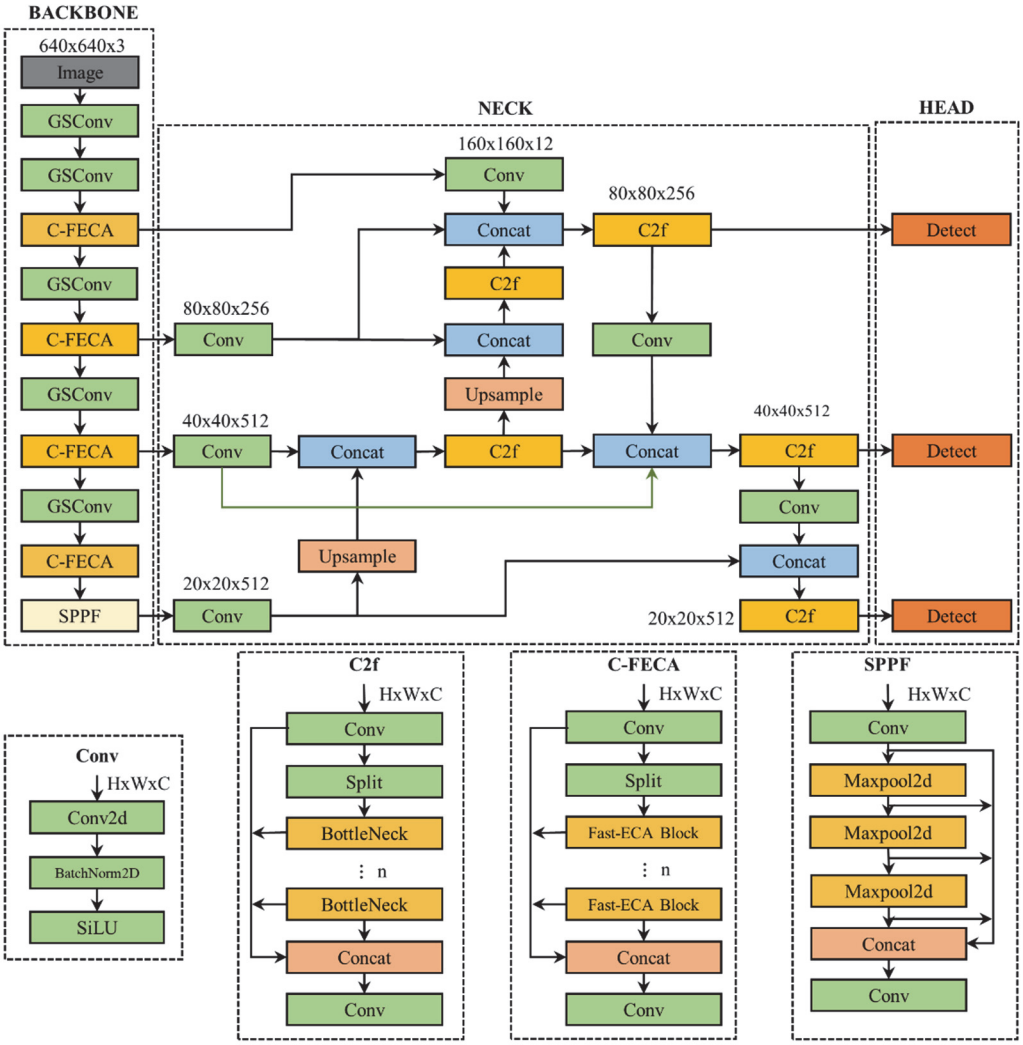
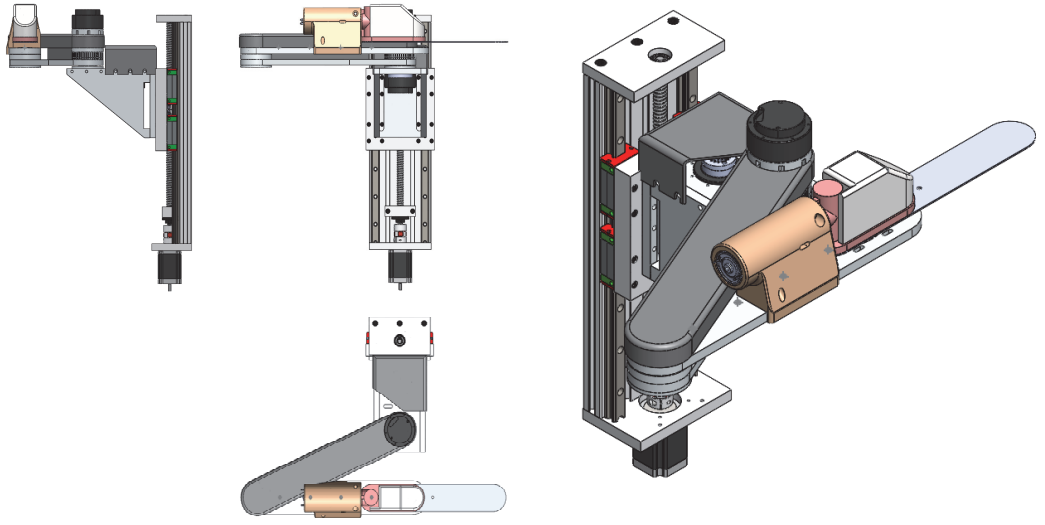


Figure 3 The architecture of the proposed detection model, based on the YOLOv8n baseline

IV Robotic arm

The mechanical foundation of the automated banana harvesting system is a specially designed 3-DoF robotic arm, featuring a hybrid RRP (Revolute-Revolute-Prismatic) kinematic structure aimed at optimizing the workspace within complex agricultural environments, as shown in Figure 4. This structure comprises one prismatic and two revolute joints, with each playing a strategic role. The prismatic joint, driven by a high-power linear actuator, is responsible for all vertical motion. Its core function is to enable the arm to precisely access banana bunches growing at diverse heights along the stem. Meanwhile, the two revolute joints operate in concert on the horizontal plane. They are actuated by high-torque MYACTUATOR RMD-X6-S2 V2 servo actuators. These integrated units were specifically chosen for their high torque output, which is achieved through a 1:36 reduction

gearbox, making them capable of handling the arm's dynamic loads and the weight of the banana bunch. This power ensures flexible reach and precise positioning capabilities even within the narrow spaces between crop rows. This combination provides the necessary dexterity to navigate around obstacles like leaves and to precisely orient the end-effector. The end-effector is a sophisticated, integrated system, comprising a high-speed rotating blade that executes a decisive cut and a mechanical support basket. This basket not only stabilizes the banana bunch but also acts as a gripping mechanism, isolating it from the vibrations of the cutting process and preventing it from free-falling, thereby maximally preserving the quality of the fruit.



(a) (b)

Figure 4 A CAD model of a 3-DOF robotic actuator. (a) Orthographic views. (b) Isometric view

V Control

The system's control architecture is designed as a comprehensive closed-loop model, ensuring high precision and responsiveness in real-world operational environments. The foundation of this architecture is the CAN bus communication network, chosen as the primary protocol to connect the central controller with the servo motors, providing superior electromagnetic interference immunity in agricultural environments.

The operational sequence is initiated when the deep learning model (Banana-DL) identifies the target coordinates. This data is then passed to an Inverse Kinematics (IK) solver to transform the target from Cartesian space to joint space. Subsequently, a Cubic Polynomial Trajectory Planning module calculates the optimal motion path for each joint. The choice of a cubic polynomial allows

the module to define a smooth trajectory between the initial and final positions while satisfying velocity constraints at the start and end points (typically zero velocity). This results in a continuous velocity profile and piecewise constant acceleration, which eliminates abrupt acceleration changes—the primary cause of vibrations and jerk. The result of this method is illustrated in Figure 5. Finally, real-time position feedback from the encoders, transmitted via the CAN bus, allows the system to continuously correct for errors, ensuring the robot accurately and stably follows the planned trajectory.

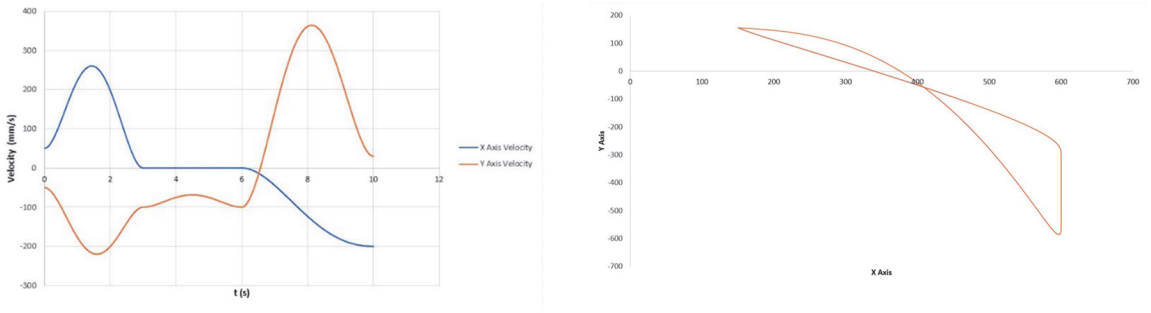


Figure 5　End-effector trajectory and velocity profile generated by the trajectory planner. (a) Velocity along the X and Y axes over time. (b) Motion trajectory in the XY plane.

Results and Discussion

I　Experimental setup

Training was conducted on a workstation with Windows 10, an Intel Core i5-9th Gen CPU @ 2.4 GHz, 16 GB RAM, and an NVIDIA Tesla T4 GPU (15 GB). The software environment included Python 3.10.12, PyTorch 2.3.1, and CUDA 12.1. Images were resized to 640 × 640 pixels, and training used the Adam optimizer with a batch size of 16, weight decay of 0.0005, momentum of 0.937, and 100 epochs.

For deployment, the models were tested on an Jetson Orin NX, which served as the embedded controller of the harvesting robot. The robot was built on an autonomous guided vehicle (3.02 m × 1.6 m) with differential drive and an anti-tip design for stable movement in orchards. An Intel RealSense D435i depth camera was mounted on an adjustable frame to capture images. A 3-DoF robotic arm with one prismatic joint and two revolute joints was used: the prismatic joint moved vertically, while the revolute joints rotated to align the cutting blade. The end-effector combined a rotary blade and a holding basket, ensuring precise and stable cutting of banana stalks.

To evaluate detection performance, several metrics were applied. Model efficiency was

measured using the number of parameters, GFLOPS, and model size, while accuracy was assessed with precision, recall, and mean Average Precision at an IoU threshold of 0.5 (mAP50). Precision indicates the proportion of correct detections among all predicted positives, whereas recall measures the ability to detect all relevant targets. mAP50 reflects the overall balance of precision and recall across classes. In addition, the total processing time per image, including preprocessing, inference, and postprocessing, was recorded to evaluate computational efficiency. For cutting point detection, the Euclidean distance between predicted and ground-truth points was calculated to quantify localization accuracy. These metrics provide a comprehensive evaluation of both detection robustness and suitability for real-time embedded deployment.

II Detection performance

The proposed Banana-DL model demonstrated both high accuracy and computational efficiency in detecting banana bunches and stalks. As shown in Table 1, the model contains only 1.7M parameters with 6.2 GFLOPS,.3%, a recall of 90%, and an mAP50 of 94.5%. The average processing time per image resulting in a compact size of 3.7 MB. Despite its lightweight design, Banana-DL achieved a precision of 96 was 40.79 ms, which corresponds to approximately 24.5 frames per second on the Jetson Orin NX platform, satisfying real-time operation requirements for harvesting robots. In addition to numerical results, the model's feature extraction capability is illustrated in Figure 6. The visualization shows how the network focuses on relevant regions of the image when detecting banana bunches and stalks. The heatmaps indicate that the Banana-DL model consistently attends to the stalk and bunch areas even under varying lighting and occlusion conditions. This confirms that the architectural improvements: GSConv, C2f-Fast Efficient Channel Attention, and BiFPN, enable the model to capture robust spatial and contextual features while maintaining efficiency. Together, these results validate the effectiveness of Banana-DL for practical deployment in orchard environments.

Table 1. Detection performance of the proposed Banana-DL model in terms of efficiency and accuracy metrics.

Model	Parameter (M)	GFLOPS (G)	Size (MB)	Precision (%)	Recall (%)	mAP50 (%)	Processing Time (ms)
Banana-DL	1.7M	6.2	3.7MB	96.3	90	94.5	40.79

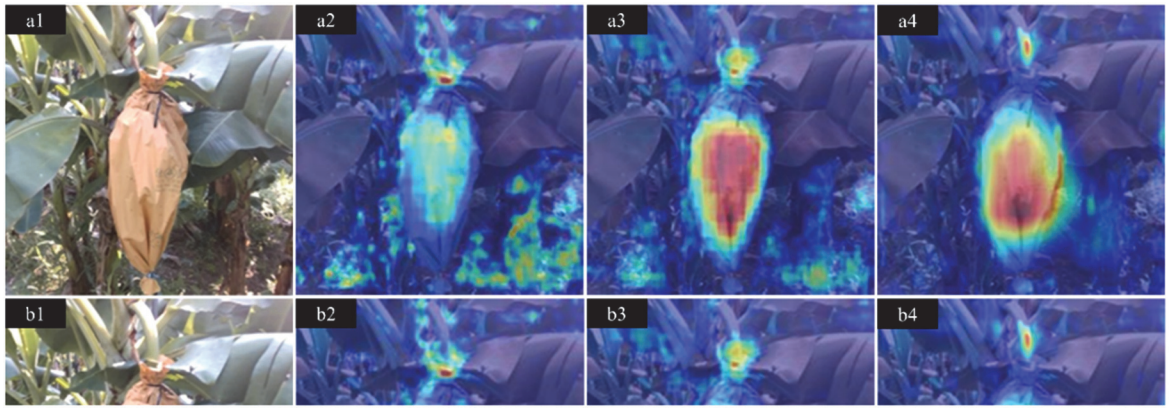


Figure 6 Visualization of feature maps from the Banana-DL model during banana bunch and stalk detection.

III Cutting point detection.

Accurate localization of the cutting point is essential to ensure successful harvesting. As shown in Figure 7, the deviations between predicted and actual cutting points were analyzed along both the X and Y axes. The Banana-DL model achieved the lowest mean error values, measuring 8.02 pixels in the X direction and 8.36 pixels in the Y direction. The overall mean Euclidean distance was 12.95 pixels with a standard deviation of 7.79 pixels. These results indicate that the predicted points were consistently close to the ground truth, confirming the model's reliability in guiding the robotic arm to the correct cutting location. In addition, the relationship between depth values and spatial resolution was evaluated to determine the tolerance threshold for successful cutting. As shown in Figure 8, linear interpolation established that at a camera distance of 40–160 cm with a resolution of 1280 × 720, the maximum spatial resolution was approximately 0.84 mm per pixel. Considering that banana stalk diameters range between 60 and 120 mm, a deviation of less than 30 mm from the stalk center (about 26 pixels in image space) is sufficient for a valid cut [3]. The proposed Banana-DL model satisfied this requirement, with 99.05% of its predictions falling within the acceptable margin. This demonstrates that the system not only achieves high detection accuracy but also provides practical precision for real-world harvesting operations.

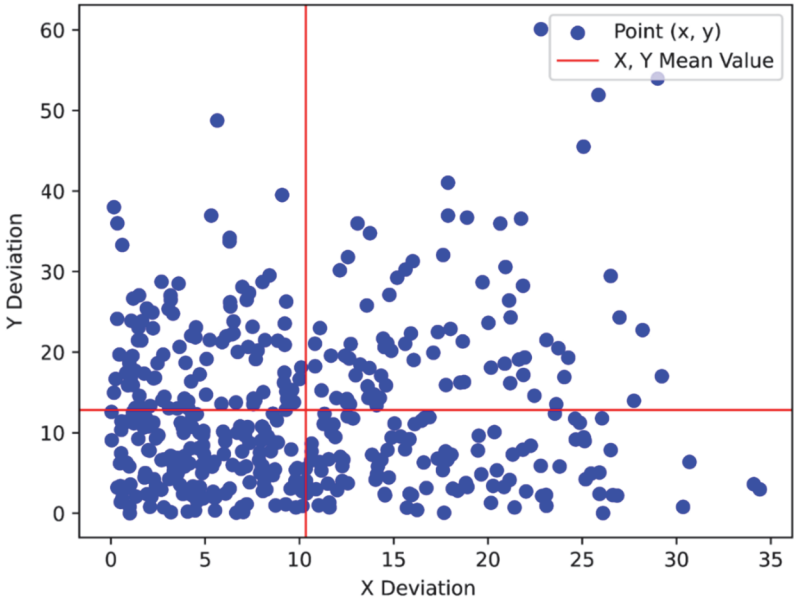


Figure 7　Distance between actual cutting points and predicted points

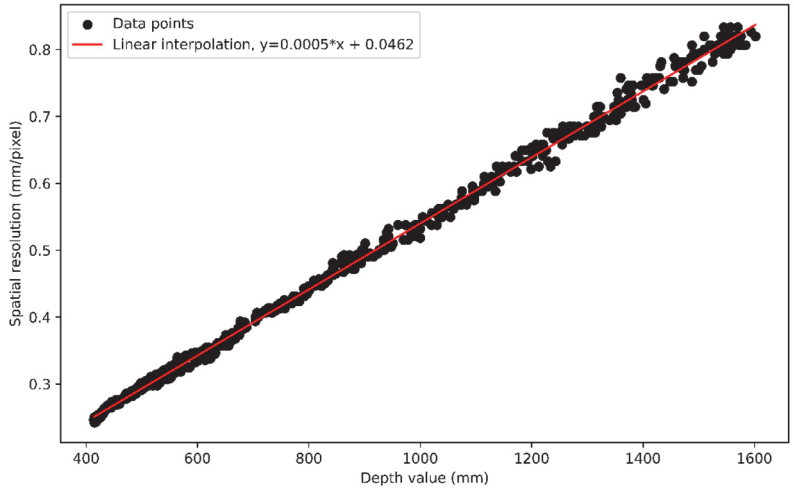


Figure 8　Relationship between depth and spatial resolution.

IV　Harvesting Operational Workflow

Figure 9 presents the operational workflow of the banana harvesting robot. In Figure 9a, the harvesting vehicle approaches and positions itself in front of the target banana bunch. Once aligned, the collecting basket begins to lift, as shown in Figure 9b, and stops automatically when it makes contact with the bunch. This ensures that the bananas are gently supported, reducing the risk of bruising. In the next stage, the robotic arm is guided by the detection model to locate the cutting

point, and then executes a precise cutting motion, as illustrated in Figure 9c and 9d. After the stalk is cut, the bunch falls securely into the basket (Figure 9e), where the added weight activates a limit switch that lowers the basket automatically. Finally, as shown in Figure 9f, the basket door can be opened manually, allowing the farmer to collect the bananas with minimal effort. This sequence represents the integration of perception, manipulation, and handling into one continuous process, ensuring safe, efficient, and practical banana harvesting in orchard environments.

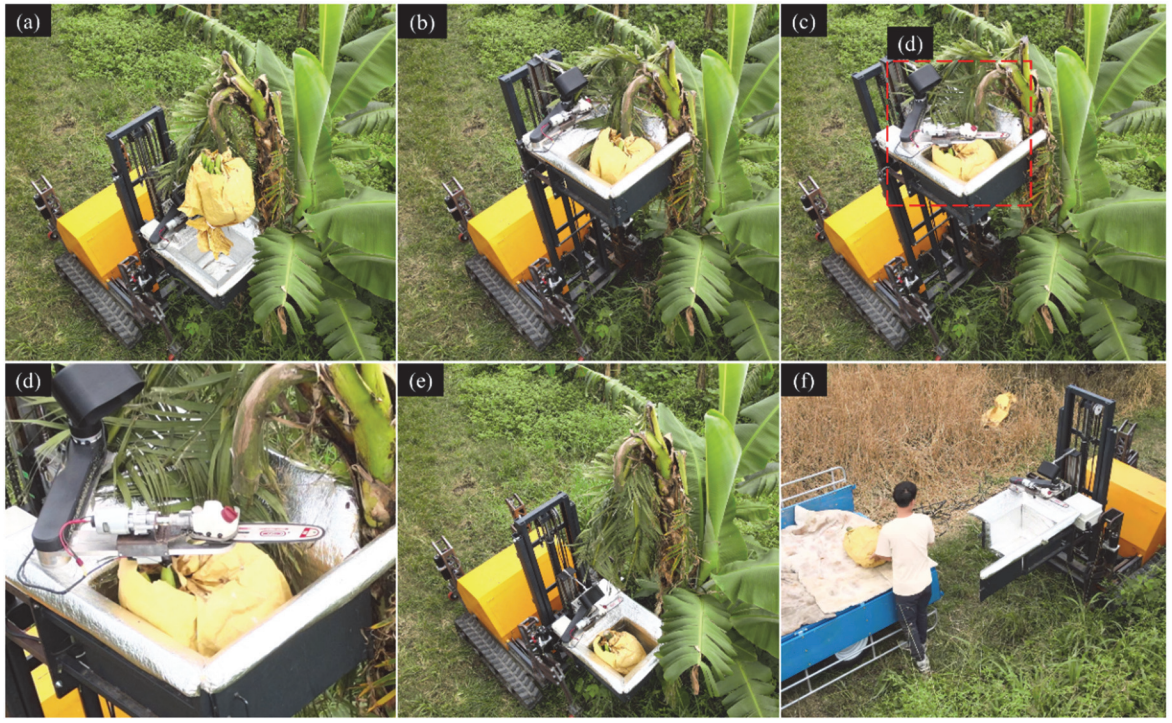


Figure 9　Banana harvesting process in the field.

Conclusion

This study successfully developed and validated an integrated robotic system for autonomous banana harvesting, addressing inherent labor and cost challenges in the industry. The core achievement of this work is the optimal combination of a specialized 3-DoF RRP arm design and a high-performance Banana-DL vision model. The arm, with its optimized kinematics, demonstrated high efficiency and reliability in complex orchard environments. Concurrently, the Banana-DL model, built on YOLOv8n, achieved outstanding accuracy (96.3% precision, 94.5% mAP50) with a compact size of just 3.7 MB, enabling real-time target detection. The seamless integration of these two components—translating visual data into precise cutting movements via an inverse kinematic controller—was the key to the system's success. Field trials have proven the feasibility and reliability

of the solution, confirming that a holistic approach, which combines bespoke mechanical design with efficient AI, can create practical and viable automation solutions for modern agriculture.

Acknowledgements

This research was supported by funding from the Ministry of Agriculture, Taiwan (113AS-12.2.2-AS-02). The authors are grateful for the support and resources provided by the Biomimicking and Engineering Lab (Being2 Lab) at National Sun Yat-sen University, Taiwan and Taiwan Banana Research Institute.

References

1. Panigrahi, N., et al., *Identifying opportunities to improve management of water stress in banana production.* Scientia Horticulturae, 2021. 276: p. 109735.
2. Guo, J., et al., *Research on the Physical Characteristic Parameters of Banana Bunches for the Design and Development of Postharvesting Machinery and Equipment.* Agriculture, 2021. 11(4).
3. Chen, T., et al., *Development, Integration, and Field Experiment Optimization of an Autonomous Banana-Picking Robot.* Agriculture, 2024. 14(8): p. 1389.
4. Wu, F., et al., *Detection and counting of banana bunches by integrating deep learning and classic image-processing algorithms.* Computers and Electronics in Agriculture, 2023. 209: p. 107827.
5. Nguyen, D.T., et al., *A lightweight and optimized deep learning model for detecting banana bunches and stalks in autonomous harvesting vehicles.* Smart Agricultural Technology, 2025. 11: p. 101051.
6. Tan, M., R. Pang, and Q.V. Le. *Efficientdet: Scalable and efficient object detection*. in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.