

# FruitTouch: A Perceptive Gripper for Gentle and Scalable Fruit Harvesting

Ruohan Zhang<sup>1</sup>, Mohammad Amin Mirzaee<sup>1</sup>, and Wenzhen Yuan<sup>1</sup>

**Abstract**—The automation of fruit harvesting has gained increasing significance in response to rising labor shortages. A sensorized gripper is a key component of this process, which must be compact enough for confined spaces, able to stably grasp diverse fruits, and provide reliable feedback on fruit condition for efficient harvesting. To address this need, we propose FruitTouch, a compact gripper that integrates high-resolution, vision-based tactile sensing through an optimized optical design. This configuration accommodates a wide range of fruit sizes while maintaining low cost and mechanical simplicity. Tactile images captured by an embedded camera provide rich information for real-time force estimation, slip detection, and softness prediction. We validate the gripper in real-world cherry tomato harvesting experiments, demonstrating robust grasp stability, effective damage prevention, and adaptability to challenging agricultural conditions.

## I. INTRODUCTION

The agricultural robotics system for harvesting has been a focus for the community due to the lack of human labor. When designing the automated harvesting system, the choice of the end effector is critical since it is directly responsible for handling delicate fruits, preventing slips, inferring fruit firmness, and preserving the quality of the fruits by avoiding crushing or rubbing during the harvest process [1]. Existing agricultural end-effectors commonly rely on visual feedback for fruit detection [2] and mechanical designs such as suction cups [3], [4], scissor cutters [5], or multi-finger grippers [6], [7] to enhance harvesting success rate. While effective under ideal conditions, these solutions face several limitations. Vision-based feedback can be unreliable when the fruit is occluded by foliage or when lighting conditions vary. Purely mechanical end-effectors, such as suction cups or cutting-based tools (e.g., scissors or blade mechanisms), often lack robustness to environmental variability, including fruit size variation, clustered growth, or surface wetness. While some of these systems later augment vision with low-resolution or binary fingertip contact sensors to signal contact onset or threshold events [8], [9], such signals provide little spatial or directional information. As a result, they cannot recover the rich contact state and fruit material properties for achieving reliable grasp.

On the contrary, human pickers excel at this task by integrating tactile and force cues to detect contact state and assess ripeness during manipulation. Recent advances in high-resolution tactile sensing, such as GelSight [10]–[13] technology, offer a promising path toward overcoming

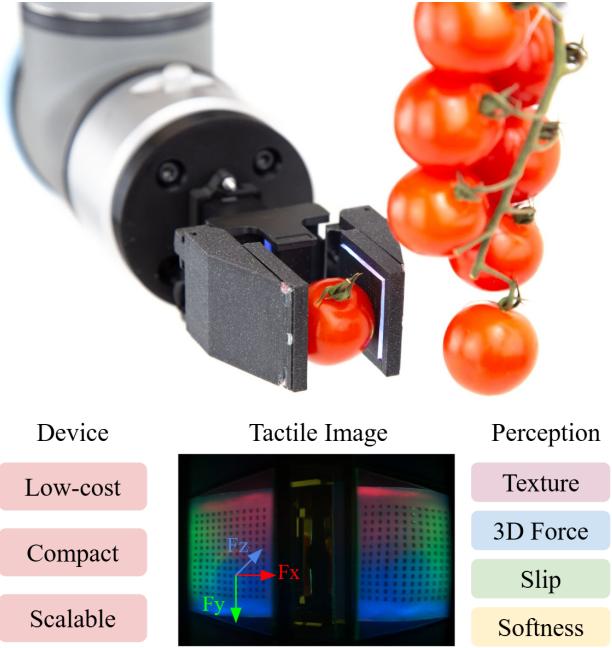


Fig. 1. Demonstration of the FruitTouch gripper harvesting cherry tomatoes. The hardware design is optimized for compactness, low cost, and scalability, while the perception system measures high-resolution contact geometry, 3D force, slip, and object softness. Together, the design integrates mechanical efficiency with rich tactile sensing to enable reliable and efficient fruit harvesting.

these limitations. GelSight sensors capture detailed surface geometry and deformation patterns, enabling accurate estimation of contact forces [14], [15], detection of slip [16], [17], and assessment of surface compliance [18]. These capabilities have driven successful applications of GelSight in various robotic manipulation tasks; however, its integration into agricultural grippers remains limited.

In this work, we present FruitTouch, a compact and lightweight agricultural end-effector that integrates high-resolution tactile sensing to enable robust fruit perception and harvesting in real-world scenarios. The design adopts a parallel-jaw mechanism with a single camera to sensorize both gripping surfaces, reducing cost and complexity for fabrication. Its thin, wedge-shaped fingers allow the gripper to reach into dense foliage and clutter, which is critical for practical field deployment. Besides, the scalable mechanical and optical design opens up the possibility of accommodating fruit sizes ranging from fruits even smaller than cherry tomatoes ( $\sim 28$  mm diameter) to large fruits like apples ( $\sim 75$  mm diameter). To summarize, our proposed FruitTouch

<sup>1</sup>Ruohan Zhang, Mohammad Amin Mirzaee, and Wenzhen Yuan are with University of Illinois at Urbana-Champaign, Champaign, IL, USA  
[{rz21,mirzaee2,yuanwz}@illinois.edu](mailto:{rz21,mirzaee2,yuanwz}@illinois.edu)

gripper can:

- Reconstruct fruit surface texture with high precision,
- Detect and prevent fruit crushing through accurate 3D force measurement,
- Identify and respond to slip events during grasping, and
- Classify fruit softness as a proxy for ripeness.

We use the example of cherry tomatoes as the target fruit —a soft, vine-grown fruit typically forming clusters of around 12 fruits per bunch, with an average fruit diameter of approximately 28.3 mm [7]. This presents significant challenges for common end-effectors due to the need for gentle yet secure handling in cluttered and constrained environments. By integrating high-resolution tactile sensing, a compact and scalable mechanical and optical design, and adaptability for agricultural settings, our system demonstrates strong performance in these demanding conditions, achieving high precision and operational efficiency during harvesting. We believe this work can contribute toward enabling scalable, automated harvesting solutions that reduce harvest losses, increase harvesting efficiency, and lessen reliance on human labor in the foreseeable future.

## II. RELATED WORK

### A. Agricultural End-Effectors for Harvesting

According to the detachment method, agricultural end-effectors can be broadly classified into two categories: cutting-based and grasping-based [19]. Cutting-based methods use a blade or knife to sever the fruit from the stem, making them suitable for fruits with relatively long and accessible stems. For example, [20], [21] employed self-designed cutting end-effectors combined with suction cups to harvest strawberries and tomatoes, while [22] proposed a cut-clip mechanism for detaching grape clusters.

In contrast, grasping-based methods harvest by directly contacting the fruit's surface with the end-effector and applying sufficient wrench for detachment. For example, [23] presented a multi-arm kiwifruit harvesting robot equipped with multiple clamping grippers, achieving an 84% harvesting success rate with an average cycle time of 5.5 s per fruit in field trials. [24] developed a strawberry picking system using a soft pneumatic gripper, which achieved a 78% success rate and 23% damage rate. These grasping-based approaches not only resemble the way humans hold fruit with their fingers but also offer the opportunity to directly sense the fruit through surface contact. This tactile interaction when grasping can provide valuable information about the fruit's physical properties, enabling more informed control strategies to improve harvesting success while reducing damage. However, effectively exploiting this interaction requires suitable sensing technology, for which vision-based tactile sensors offer a promising solution.

### B. Perception Using Vision-based Tactile Sensors

Vision-based tactile sensors, particularly GelSight [25], have shown great potential for capturing detailed contact geometry and providing rich feedback for robotic manipulation tasks. Such sensors can enable accurate force estimation,

slip detection, and object softness detection, which make it suitable for integration into agriculture end-effectors.

For example, [10] demonstrated that the relationship between indentation volume and normal force is approximately linear. Building on this insight, subsequent works [26], [27] employed finite element methods (FEM) for more refined force estimation. Shear force has been estimated by tracking the displacement of surface markers [10], [15], which have also been leveraged for slip detection: average marker displacement has been used for small objects [28], while entropy-based measures have been applied to larger contact areas [16]. Object softness prediction has likewise been an active area of research, with early approaches modeling softness as a linear function of image brightness [29], and more recent studies incorporating analytical models to better interpret tactile signals and improve compliance estimation [30].

Despite their potential for enhancing perception, existing vision-based tactile systems often suffer from bulky form factors, high cost, limited capability, or poor generalizability, making them difficult to apply directly to harvesting. For example, [31] presented a GelSight-equipped 5-DoF gripper for in-hand manipulation, but its complex multi-DOF structure runs counter to the simplicity and scalability required in agricultural applications. Likewise, [18] used GelSight as a standalone post-harvest firmness tester rather than integrating it into the picking process, while [32] developed a strawberry-harvesting end-effector that, although capable of force feedback, remained bulky and lacked both slip detection and ripeness estimation.

In contrast, our proposed gripper is sensorized, compact, and low-cost, providing integrated perception of force, slip, and ripeness within a single harvest-ready design. Unlike prior GelSight-based sensors for agriculture, it employs a single camera shared across both fingers through a common optical path. This approach reduces cost while maintaining sensing effectiveness under varying gripper configurations, enabling practical deployment in dynamic harvesting environments.

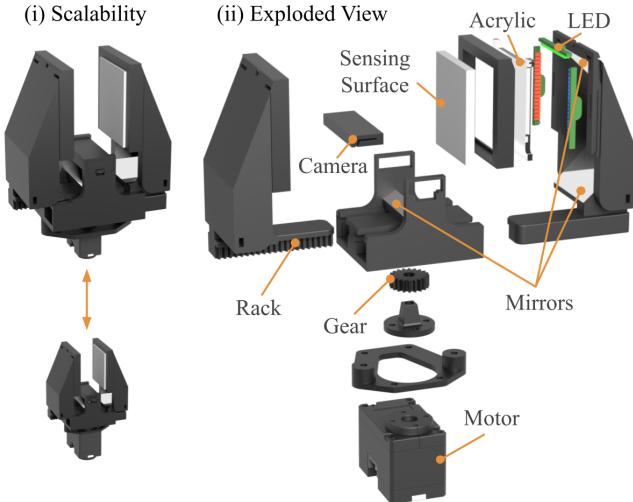
## III. SENSOR-INTEGRATED GRIPPER DESIGN

To achieve a practical balance between simplicity, robustness, and sensing capability, we adopt a parallel-jaw gripper design, which reduces mechanical complexity, minimizes interference with both the optical system and the surrounding environment during harvesting. This also mimics the natural two-finger picking strategy, with integrated tactile sensing on both fingers. The following subsections present the mechanical specifications of the gripper in Section III-A and the optical design in Section III-B.

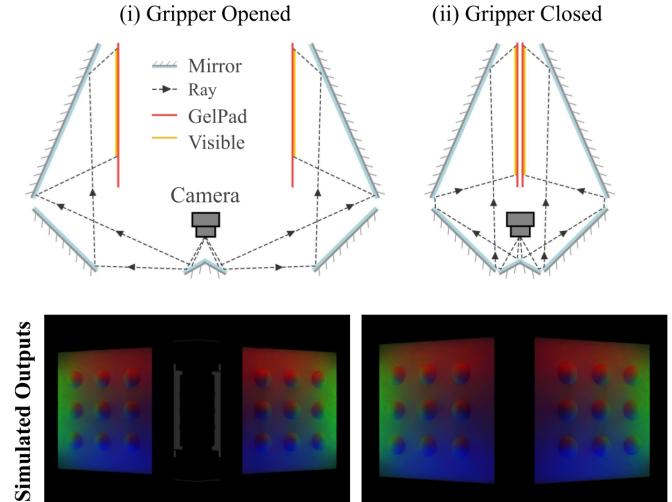
### A. Mechanical Design

We placed the racks and gear at the base of the gripper to actuate the fingers, to minimize interference with the optical subsystem. Racks are attached to the fingers and coupled with a gear connected to a DYNAMIXEL XC330-M288-T motor, as shown in Figure 2A.

### (A) Mechanical Design



### (B) Optical Design



**Fig. 2. Mechanical and Optical Design of FruitTouch Gripper.** (A) Mechanical Design. The gripper components are designed for scalability, enabling harvesting of fruits with varying sizes. We use gear-and-rack mechanism to provide actuation. Each finger consists of a soft silicone sensing surface supported by a transparent acrylic sheet, with three LED strips ensuring uniform illumination. A centrally mounted camera, in combination with mirrors, provides comprehensive coverage of both sensing surfaces. (B) Optical Design for the fingertip tactile sensors. The mirror configuration is optimized to maximize the sensing area while maintaining low distortion across different finger distances. Simulated outputs are shown for both the open and closed states of the gripper.

The smallest version of the FruitTouch gripper was designed to handle small fruits such as cherry tomatoes, which are still typically harvested by humans through grasping and twisting. The sensing surface, based on the mechanical configuration, employed a target gelpad size of 30mm and a stroke of approximately 40mm. Excluding the motor, the closed gripper measures  $35 \times 40 \times 66 \text{ mm}^3$ , and the fingers can open up to 40mm, providing sufficient range to grasp small fruits including grapes, strawberries, and cherry tomatoes. Both the mechanical and optical designs are inherently scalable: while the gripper can be enlarged without specific limitations, further miniaturization is constrained by the physical dimensions of the camera.

#### B. Optical Design

The optical subsystem is defined by the two sensing surfaces at one end and the camera at the other. Intermediate mirrors divide the camera's field of view into two halves, each corresponding to one finger, while LEDs placed near the gelpads enhance the tactile signals.

To determine the optimal configuration of the mirrors and sensing surfaces, we simulated the sensor in two stages. First, using the 2D Ray Optics Simulation on phydemo.app, we coarse-aligned three mirrors relative to the camera for each finger. Our goal was to orient the optical components such that the incident rays strike the sensing surface as orthogonal as possible, since this minimizes perspective distortion in each frame and maintains similar tactile signals in different finger distances. The result of this stage is shown in Figure 2B for the open and closed configuration of the finger. We observed that both cases yield similar perspectives with acceptable levels of distortion. Moreover, the optical

configuration is fully scalable while preserving the same coverage ratio.

We used the geometrical information of this step to run a more detailed optical simulation in Blender to fine-tune the mirror configuration and optimize the light locations. We adopted the method from [33] to optimize the location of the LEDs, obtaining an improved color matching to surface normal (RGB2Norm Metric). This yields a high-quality optical system, as evidenced by the minimally distorted spherical indentations in Figure 2B, which in turn enhances the accuracy of geometry calibration and reconstruction. We also simulated a scaled version of our gripper, where the tactile signals remained consistent and the only variation was in the pixel-to-millimeter ratio.

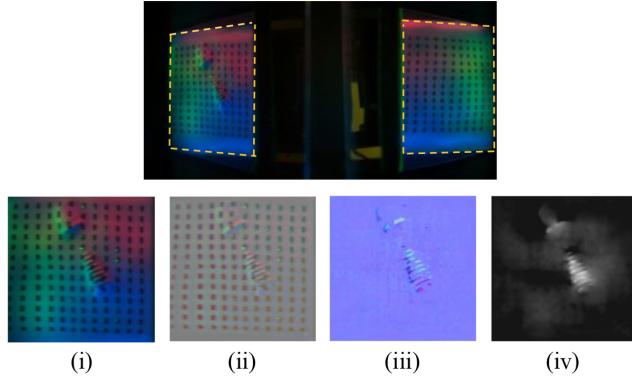
## IV. METHOD OF PERCEPTION

### A. Tactile Geometry Reconstruction

Tactile geometry reconstruction is a core capability of vision-based tactile sensing and it enables downstream tasks such as fruit-type classification, bruise/defect detection, and contact-state reasoning. The idea was first demonstrated by [25], which used colored illumination and an internal camera to map RGB intensities to local surface normals via photometric stereo. From these normals, the contact geometry can be represented directly as a dense normal field or integrated to yield a height map/mesh that captures fine contact textures.

In our work, geometry reconstruction is achieved by first calibrating the sensor through indentations with steel balls of known radius (5 mm). The raw tactile image is rectified into a rectangular frame, as shown in Figure 3A, and the background image is subtracted to enhance the contact

### (A) Reconstruction Pipeline



### (B) Tactile Imprints

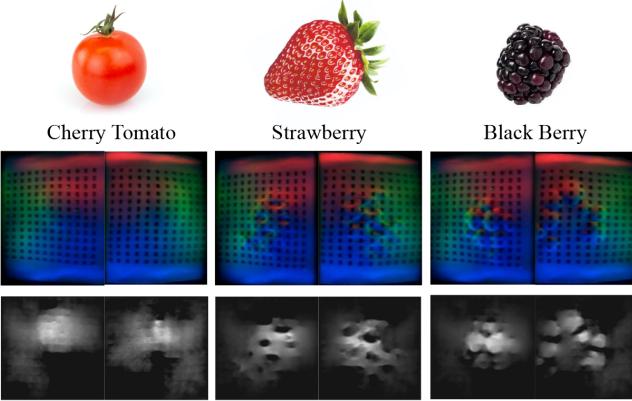


Fig. 3. **Tactile Sensing Pipeline.** (A) Contact-geometry reconstruction pipeline. **Top:** Raw camera output that contains reading from both fingers. (i) Unwarped contact frames on one sensor. (ii) Background-subtracted difference images. (iii) Estimated surface normals. (iv) Reconstructed 3D shape (shown in the form of a heightmap). (B) Example tactile images and the reconstructed shape for a cherry tomato, a strawberry, and a blackberry.

signal. The resulting contact images are paired with manually labeled surface normals to build a calibration dataset. Using this dataset, we train a MLP to learn the mapping from RGB pixel intensities to surface normals. At inference time, the predicted normals are numerically integrated to recover the height map of the indentation, thereby reconstructing the local contact geometry. The complete pipeline is illustrated in Figure 3A.

### B. Three-dimensional Force Estimation

Three-dimensional force estimation, including both normal and shear components, is essential for closed-loop control in fruit harvesting. For example, normal force estimation helps maintain appropriate grasp strength, while shear force prediction assists in preventing slip. To achieve this, we leverage FruitTouch’s physical properties with a lightweight, data-driven model that fuses actuator signals and visual cues from the gel surface for real-time inference.

In our parallel-jaw gripper setting, normal forces at the fingertips generate opposing contact loads that are transmitted through the racks and gear to the motor. This increases the

torque demand at the actuator, and under fixed gearing, the motor current is approximately proportional to the applied torque. We therefore fit a simple linear model,

$$F_n = aI + b, \quad (1)$$

that maps the current  $I$  to the normal force  $F_n$ .

For shear force estimation, we infer both magnitude and direction by visually tracking the displacement of etched surface markers on the gel. Following the magnetic-field analogy and Helmholtz–Hodge decomposition described in [15], the displacement field is separated into curl-free and divergence-free components, which correspond to translational and rotational shear patterns, respectively. A feature vector is then constructed by applying low-order polynomial expansions to these components, enabling a linear readout of  $(F_x, F_y)$ .

To be specific, the 2D marker displacement field  $\mathbf{V}$  is decomposed into three orthogonal components:

$$\mathbf{V} = \mathbf{P} + \mathbf{S} + \mathbf{H} \quad (2)$$

Where  $\mathbf{P}$  denotes the curl-free (irrotational) component with  $\nabla \times \mathbf{P} = 0$ ,  $\mathbf{S}$  is the divergence-free (solenoidal) component with  $\nabla \cdot \mathbf{S} = 0$ , and  $\mathbf{H}$  is the harmonic component satisfying both  $\nabla \cdot \mathbf{H} = 0$  and  $\nabla \times \mathbf{H} = 0$ . The tactile feature vector is then constructed by applying polynomial expansions to the x- and y-components of the decomposed fields. Denoting the interpolated marker components as  $\mathbf{p}, \mathbf{s}, \mathbf{h}$  and their respective scalar components using subscripts (e.g.,  $p_x, p_y$ ), the shear deformation feature is defined as:

$$\begin{aligned} \mathbf{x}_{\text{shear}} &= [\mathbf{v}_x, \mathbf{v}_y, \mathbf{p}_x^*, \mathbf{p}_y^*, \mathbf{s}_x^*, \mathbf{s}_y^*]^T, \\ \mathbf{p}_x^* &= [\mathbf{p}_x, \mathbf{p}_x^2]^T & \mathbf{p}_y^* &= [\mathbf{p}_y, \mathbf{p}_y^2]^T \\ \mathbf{s}_x^* &= [\mathbf{s}_x, \mathbf{s}_x^2]^T & \mathbf{s}_y^* &= [\mathbf{s}_y, \mathbf{s}_y^2]^T \end{aligned} \quad (3)$$

and the total shear force prediction can be written as

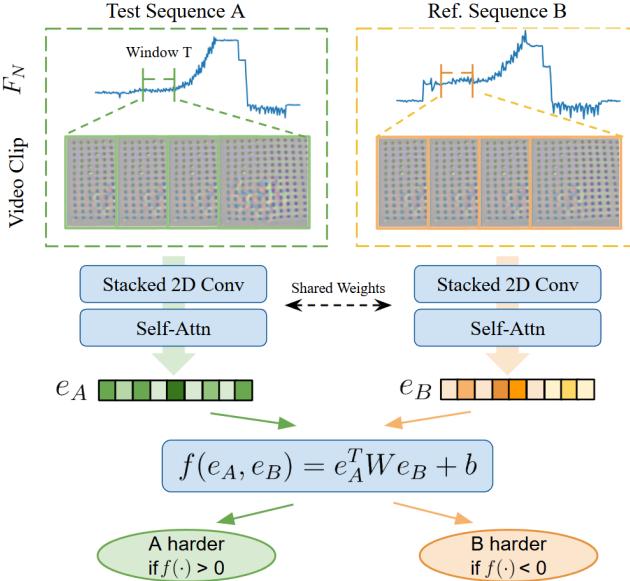
$$\mathbf{F}_{\text{shear}} = \begin{bmatrix} F_x \\ F_y \end{bmatrix} = \begin{bmatrix} \mathbf{w}_x^T \\ \mathbf{w}_y^T \end{bmatrix} [\mathbf{x}_{\text{shear}}] + \begin{bmatrix} b_x \\ b_y \end{bmatrix} \quad (4)$$

### C. Real-Time Slip Detection

In agricultural harvesting, slip refers to unintended relative motion between the end effector and the fruit, often leading to damage, grasp failure, or reduced efficiency. Achieving stable agricultural harvesting, therefore, requires critical capabilities in slip detection. Inspired by [34], our approach detects slip by directly comparing the motion of a grasped object with the displacement of surface markers, a method made possible by high-resolution contact sensing.

To estimate the object’s motion, we first segment the contact region by thresholding the reconstructed height map. The average velocity of this region’s center point serves as a proxy for the object’s velocity. We then compute the marker velocities within the same contact region and evaluate the difference between the object and marker velocities. Slip is reported when this difference exceeds a predefined threshold.

### (A) Softness Perception Pipeline



### (B) Softness Perception Accuracy

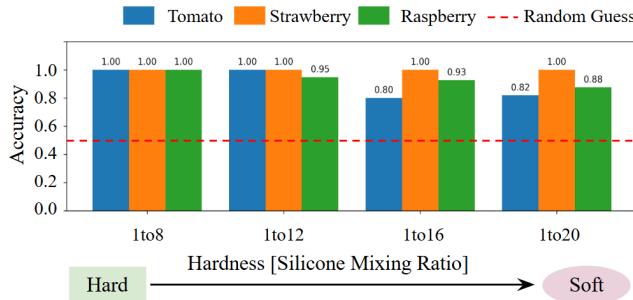


Fig. 4. (A) Softness Prediction Pipeline. We randomly segment tactile image sequences and the corresponding normal force signals from the pressing process of the same fruit, and extract embeddings using a shared backbone. A lightweight ranking head then predicts whether the test object is harder or softer than the reference. (B) Softness perception results. Each bar shows classification accuracy for a specific hardness within a fruit type (comparisons within the same fruit). The red dotted line marks random guess (50%), and the overall average accuracy is 94.7%.

#### D. Object Softness Measurement

Determining whether a fruit is ready to be picked requires estimating its firmness or softness as a proxy for ripeness. Our gripper enables this capability in-hand: during a gentle squeeze, it combines real-time normal force from motor current with tactile imprints from the gels to infer softness online, allowing pre-pick screening without additional labor. Moreover, the method is fruit-agnostic, with a single model generalizing across varieties without retraining.

Previous methods for estimating softness generally fall into two classes: classification-based [35] and regression-based [36], [37]. Classification-based methods assign the object to one of several predefined firmness/ripeness categories (e.g., unripe/ready/overripe). In contrast, regression-based methods predict a continuous-valued mechanical property,

such as Shore hardness, elastic modulus, or compliance, providing a calibrated measure of softness. Classification simplifies downstream decisions but depends on well-chosen class boundaries and cannot generalize without retraining for each fruit; regression yields finer resolution but requires accurate ground-truth measurements, which is almost impossible in agriculture.

Inspired by the human picking process, where we simply set a threshold in mind for the right picking ripeness and compare all samples with it among the same kind of fruit, we frame softness prediction problem in agriculture as a pairwise comparison. Given two short compression sequences of the same fruit, the model predicts whether the first is harder than the second. This formulation avoids the need for absolute calibration, supports cross-variety generalization, and can be extended into an absolute softness scale through aggregation if desired.

We employ a lightweight ranker that encodes each compression clip with a short stack of 2D frame encoders followed by a self-attention to capture temporal features. A real-time estimate of normal force is projected and concatenated with the visual embedding. Given embeddings  $e_A, e_B \in \mathbb{R}^D$ , a asymmetric bilinear comparator  $f(e_A, e_B) = e_A^T W e_B + b$  with  $W = -W^\top$  produces a logit  $f$ . We train with binary cross-entropy on pair labels, forming within-fruit pairs. At inference, we predict “ $A > B$ ” when  $f \geq 0$ . We visualize the model structure in Figure 4A.

## V. EXPERIMENTS

### A. Geometry Reconstruction

Figure 3B shows the output of both fingers and the reconstructed height map of the contact shape. Although the gripper is opened to different widths, our calibration model remains agnostic to the opening size and accurately estimates the corresponding normal maps.

To quantitatively assess sensor performance, we 3D-printed small objects with known geometries and compared the reconstructed height maps to ground truth. Specifically, using a six-sided pyramid (10 mm diameter, 2 mm height) to indent the surface at multiple locations, we achieved an MSE of  $0.201 \text{ mm}^2$ , demonstrating that our sensor reconstructs contact features with high accuracy.

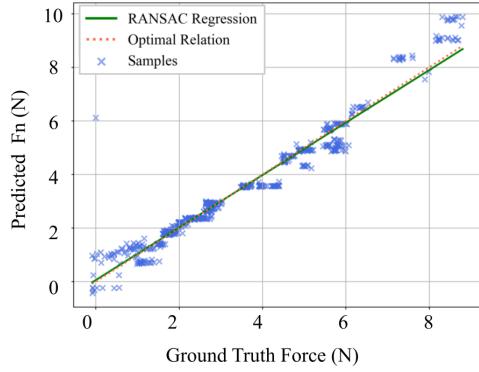
### B. Gasping Force Estimation

For force estimation experiments, we mounted an ATI Nano17 force/torque sensor and grasped it with the proposed gripper at various locations, as shown in Figure 5A. We gradually closed the gripper while recording the force measurements and the motor current simultaneously. The procedure was repeated 20 times with different indenter shapes attached to the FT sensor to assess generalization across contact geometries. In total the model’s normal-force predictions from motor current closely matched ground truth with a coefficient of determination  $R^2 = 0.951$ . The shear-force model, based on marker-displacement features, also showed strong agreement with  $R^2 = 0.903$ . Figure 5

(A) Setup for Force Prediction



(B) Normal Force Prediction Result



(C) Shear Force Prediction Result

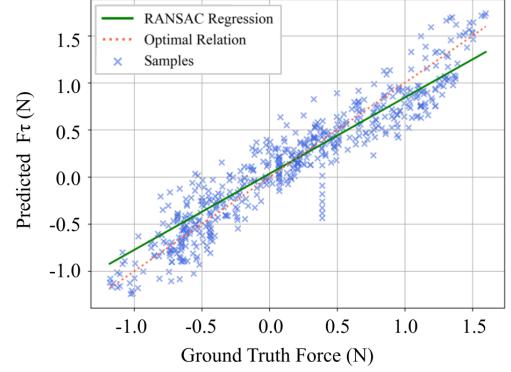


Fig. 5. **Setup and Results for Force Prediction.** (A) Experimental setup with the gripper mounted on an ATI Nano Force/Torque sensor, which measures the ground-truth force. (B, C) Results of force estimation. The **middle** panel compares normal force ( $F_n$ ), while the **right** panel compares shear force ( $F_r$ ). For clarity, the magnitude of the shear force is used for visualization. The solid green line indicates the RANSAC fit, and the dotted red line denotes the identity line ( $y=x$ ).

demonstrate the relationship between the measured forces and prediction for both directions.

### C. Slip Detection

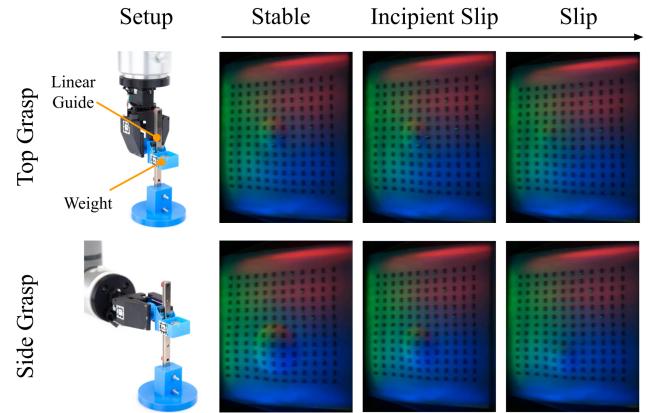
We evaluated the slip detection system on a low-friction linear rail with adjustable loading, as shown in top row in Figure 6. The gripper was tested in two grasping poses (top and side) to simulate diverse harvesting configurations, where the end-effector's approach direction may be constrained by the environment. To modulate sliding speed, we used three different loads: 10, 20, and 50 grams. Each experimental combination was repeated twice to ensure the reliability of the results. We use the AprilTag's displacement from its starting position as the ground-truth indicator of slip.

The results are summarized in Figure 6, which shows sliding demonstrations and frame-by-frame prediction outputs for different trials. We adopt this frame-by-frame approach to provide early warnings for possible slip events, and it can be integrated directly with the controller for applications such as slip recovery. Quantitatively, the classifier attains a precision of 0.725, recall of 0.661, and F1 of 0.692. On correctly detected trials, the predicted slip typically precedes the first observable relative motion by 0.11 s on average, indicating preferential early reporting and enabling pre-emptive intervention. The primary failure mode occurs when the gripper fully loses contact with the fruit (post-detachment), at which point the signals collapse and the model cannot issue a warning.

### D. Softness Measurement

To systematically collect grasping data across different textures and hardness levels, we fabricated silicone replicas of three types of fruit: strawberry, raspberry, and cherry tomato using custom molds. For each fruit, we prepared four hardness variants by mixing two-part silicones (XP-565, Silicone Inc.) in different ratios (1: 8, 1:12, 1:16, and 1:20 for parts A and B). This results in hardness values of 68.4, 64.8, 51.4, and 42.2 on the Shore 00 scale. For each object,

(A) Experimental Setup and Slip Stages



(B) Ground-truth vs. Predicted Slip

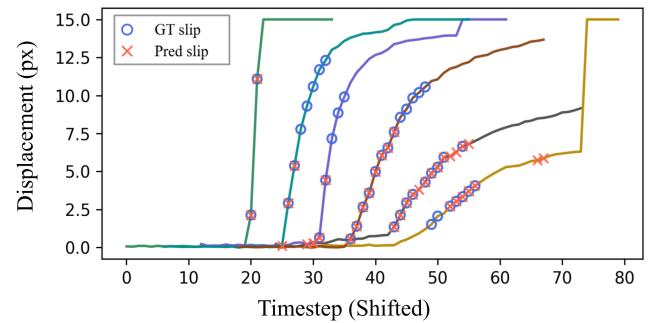


Fig. 6. **Setting and results of slip detection.** (A) Two grasp positions from the experiment, illustrating different slip stages (stable → incipient slip → full slip). (B) Time-series comparison of ground-truth and predicted slip events. The y-axis shows the AprilTag displacement from the initial pose, which we use as the ground-truth slip signal. Predictions are overlaid for better visualization.

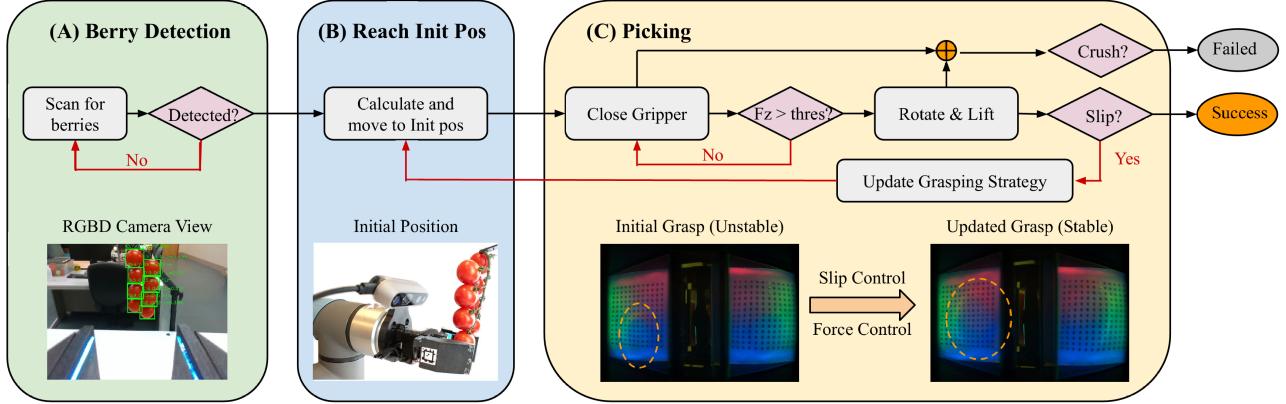


Fig. 7. Flow diagram of harvest experiment that includes berry detection, initial pose calculation, and picking phases. (A) Use the RGB-D camera to scan the area and identifying berries. (B) Align and move the manipulator to initial pose. (C) The gripper executes the picking action according to the chosen control strategy.

we conducted ten grasping trials at varied positions while recording video data of the interactions. Normal force was estimated using the proposed linear model, and the resulting signals were used to train the pairwise classification network.

Figure 4B illustrates the model’s performance on the test sequences. Each bar shows classification accuracy for a specific hardness level within a fruit type, with comparisons restricted to fruits of the same kind. In general, fruits with complex surface textures (e.g., strawberry, raspberry) yield higher accuracy than textureless fruits (e.g., cherry tomato), as the richer contact patterns provide more informative tactile signals. This suggests that surface geometry plays an important role in facilitating reliable softness estimation. Aggregated across all fruits, the model achieves  $\sim 94.7\%$  validation accuracy. These results support the use of the pairwise ranker for in-hand, fruit-agnostic pre-pick screening.

#### E. Robot System For Harvesting

We integrate the proposed FruitTouch gripper into a fully autonomous robotic system to validate its performance in cherry tomato harvesting. As illustrated in Figure 7, the system uses a state machine for motion planning, while real-time force estimation and slip detection enhance grasp stability and responsiveness during the picking process. The gripper is mounted on a UR5e robotic arm, and cherry tomatoes are suspended on a vine within the robot’s reachable workspace to simulate a realistic harvesting scenario. Using a RealSense D435 camera, a YOLOv5 detector first localizes the fruits in the world frame, after which the robot end effector moves to an appropriate grasping pose.

To highlight the importance of enhanced sensing at the end effector, we conducted ablation experiments to compare three control strategies: open-loop, slip control, and slip+force control. In the open-loop setting, once the fruit’s position and diameter are detected, the gripper closes to a fixed width (fruit diameter minus 2 mm) to attempt the pick. In the slip control strategy, slip events are continuously monitored. Upon detecting slip, the gripper closes an additional 2 mm

TABLE I  
RESULTS ON REAL TOMATO VINE GRASPING. SUCCESS IS DEFINED AS DETACHING THE TOMATO WITHOUT BREAKING; METRICS INCLUDE AVERAGE ATTEMPTS AND NORMAL GRASPING FORCE.

Strategy	Succ. (%)	Attempts	Mean $F_z$ (N)	Var $F_z$
Open-loop	58.3	—	—	—
Slip control	100	1.167	1.571	0.506
Slip+Force control	100	1.417	1.557	0.062

before retrying. In the slip+force control strategy, retries are instead guided by force: the initial pressing force is set to 1.2 N, and subsequent attempts increase the force in 0.3 N increments. Each strategy was evaluated on 12 fruits, with a grasp considered successful if the tomato was detached without breaking. As summarized in Table I, the combined slip+force controller achieves best picking success with the lowest variation in peak force at the cost of requiring slightly more retries. This demonstrates the gripper’s ability to achieve reliable, gentle fruit picking, with reduced force variability ensuring consistent contact control and minimal damage.

#### VI. CONCLUSION

We presented FruitTocuh, a compact, low-cost, customizable, sensorized gripper that brings high-resolution tactile perception, including geometry reconstruction, force estimation, slip detection, and in-hand softness assessment into a practical harvesting system. In controlled evaluations, the gripper achieved accurate force estimation (normal  $R^2=0.951$ , shear  $R^2=0.903$ ), timely slip warnings (precision = 0.725, recall = 0.661, F1=0.692; 0.11s early), and fruit-agnostic pairwise softness ranking ( $\sim 94.7\%$  accuracy). Harvest ablations confirmed the benefits of tactile feedback: both slip control and slip+force control reached nearly 100% success, with the latter yielding the lowest peak-force variance at the cost of slightly more retries, indicating reliable yet

gentle picking. Together, these results demonstrate progress toward three key challenges in automated harvesting: enabling damage-less grasping of fragile fruits through gentle and adaptive control, achieving a cost-efficiency balance via a compact and scalable design, and integrating in-hand sensing for fruit quality assessment and grading.

## REFERENCES

- [1] Z. Hussein, O. A. Fawole, and U. L. Opara, "Harvest and postharvest factors affecting bruise damage of fresh fruits," *Horticultural Plant Journal*, vol. 6, no. 1, pp. 1–13, 2020.
- [2] K. Koe, P. K. Shah, B. Walt, J. Westphal, S. Marri, S. Kamtikar, J. S. Nam, N. K. Uppalapati, G. Krishnan, and G. Chowdhary, "Precision harvesting in cluttered environments: Integrating end effector design with dual camera perception," 2025.
- [3] C. Tawk, A. Gillett, M. in het Panhuis, G. M. Spinks, and G. Alici, "A 3d-printed omni-purpose soft gripper," *IEEE Transactions on Robotics*, vol. 35, no. 5, pp. 1268–1275, 2019.
- [4] P. V. P. Reddy and V. Suresh, "A review on importance of universal gripper in industrial robot applications," *Int. J. Mech. Eng. Robot. Res.*, vol. 2, no. 2, pp. 255–264, 2013.
- [5] F. Qingchun, W. Zou, P. Fan, C. Zhang, and X. Wang, "Design and test of robotic harvesting system for cherry tomato," *International Journal of Agricultural and Biological Engineering*, vol. 11, pp. 96–100, 01 2018.
- [6] A. Silwal, J. R. Davidson, M. Karkee, C. Mo, Q. Zhang, and K. Lewis, "Design, integration, and field evaluation of a robotic apple harvester," *Journal of Field Robotics*, vol. 34, no. 6, pp. 1140–1159, 2017.
- [7] J. Gao, F. Zhang, J. Zhang, T. Yuan, J. Yin, H. Guo, and C. Yang, "Development and evaluation of a pneumatic finger-like end-effector for cherry tomato harvesting robot in greenhouse," *Computers and Electronics in Agriculture*, vol. 197, p. 106879, 2022.
- [8] H. Zhou, J. Xiao, H. Kang, X. Wang, W. Au, and C. Chen, "Learning-based slip detection for robotic fruit grasping and manipulation under leaf interference," *Sensors*, vol. 22, no. 15, 2022.
- [9] G. Tian, J. Zhou, and B. Gu, "Slipping detection and control in gripping fruits and vegetables for agricultural robot," *International Journal of Agricultural and Biological Engineering*, vol. 11, pp. 29–34, 08 2018.
- [10] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.
- [11] M. A. Mirzaee, H.-J. Huang, and W. Yuan, "Gelbelt: A vision-based tactile sensor for continuous sensing of large surfaces," *IEEE Robotics and Automation Letters*, vol. 10, no. 2, pp. 2016–2023, 2025.
- [12] S. Wang, Y. She, B. Romero, and E. Adelson, "Gelsight wedge: Measuring high-resolution 3d contact geometry with a compact robot finger," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. Xi'an, China: IEEE, 2021, pp. 6468–6475.
- [13] R. Zhang, U. Yoo, Y. Li, A. Argawal, and W. Yuan, "Pneugelsight: Soft robotic vision-based proprioception and tactile sensing," 2025.
- [14] G. Zhang, Y. Du, H. Yu, and M. Y. Wang, "Deltact: A vision-based tactile sensor using dense color pattern," 2022.
- [15] Y. Zhang, Z. Kan, Y. Yang, Y. A. Tse, and M. Y. Wang, "Effective estimation of contact force and torque for vision-based tactile sensors with helmholtz-hodge decomposition," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4094–4101, 2019.
- [16] X. Hu, A. Venkatesh, Y. Wan, G. Zheng, N. Jawale, N. Kaur, X. Chen, and P. Birkmeyer, "Learning to detect slip through tactile estimation of the contact force field and its entropy," 2024.
- [17] I. H. Taylor, S. Dong, and A. Rodriguez, "Gelslim 3.0: High-resolution measurement of shape, force and slip in a compact tactile-sensing finger," in *2022 International Conference on Robotics and Automation (ICRA)*. Philadelphia, PA, USA: IEEE, 2022, pp. 10781–10787.
- [18] L. He, L. Tao, Z. Ma, X. Du, and W. Wan, "Cherry tomato firmness detection and prediction using a vision-based tactile sensor," *Journal of Food Measurement and Characterization*, vol. 18, pp. 1–12, 11 2023.
- [19] Q. Wang, Y. Tu, W. Xu, J. Zhang, A. Knoll, M. Zhou, and Y. Ying, "Towards damage-less robotic fragile fruit grasping: A systematic review on system design, end effector, and visual and tactile feedback," *Journal of Field Robotics*, vol. n/a, no. n/a.
- [20] S. Hayashi, K. Shigematsu, S. Yamamoto, K. Kobayashi, Y. Kohno, J. Kamata, and M. Kurita, "Evaluation of a strawberry-harvesting robot in a field test," *Biosystems Engineering*, vol. 105, no. 2, pp. 160–171, 2010.
- [21] T. Fujinaga, S. Yasukawa, and K. Ishii, "Development and evaluation of a tomato fruit suction cutting device," in *2021 IEEE/SICE International Symposium on System Integration (SII)*, 2021, pp. 628–633.
- [22] Y. Jiang, J. Liu, J. Wang, W. Li, Y. Peng, and H. Shan, "Development of a dual-arm rapid grape-harvesting robot for horizontal trellis cultivation," *Frontiers in Plant Science*, vol. Volume 13 - 2022, 2022.
- [23] H. A. Williams, M. H. Jones, M. Nejati, M. J. Seabright, J. Bell, N. D. Penhall, J. J. Barnett, M. D. Duke, A. J. Scarfe, H. S. Ahn, J. Lim, and B. A. MacDonald, "Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms," *Biosystems Engineering*, vol. 181, pp. 140–156, 2019.
- [24] G. Ren, T. Wu, T. Lin, L. Yang, G. Chowdhary, K. C. Ting, and Y. Ying, "Mobile robotics platform for strawberry sensing and harvesting within precision indoor farming systems," *Journal of Field Robotics*, vol. 41, no. 7, pp. 2047–2065, 2024.
- [25] M. K. Johnson and E. H. Adelson, "Retrographic sensing for the measurement of surface texture and shape," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1070–1077.
- [26] D. Ma, E. Donlon, S. Dong, and A. Rodriguez, "Dense tactile force estimation using gelslim and inverse fem," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5418–5424.
- [27] C. Zhao, J. Liu, and D. Ma, "ifem2.0: Dense 3-d contact force field reconstruction and assessment for vision-based tactile sensors," *IEEE Transactions on Robotics*, vol. 41, pp. 289–305, 2025.
- [28] W. Yuan, R. Li, M. A. Srinivasan, and E. H. Adelson, "Measurement of shear and slip with a gelsight tactile sensor," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 304–311.
- [29] W. Yuan, M. A. Srinivasan, and E. H. Adelson, "Estimating object hardness with a gelsight touch sensor," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 208–215.
- [30] M. Burgess, J. Zhao, and L. Willemet, "Learning object compliance via young's modulus from single grasps using camera-based tactile sensors," 2025.
- [31] Y. Zhou, P. Zhou, S. Wang, and Y. She, "In-hand singulation, scooping, and cable untangling with a 5-dof tactile-reactive gripper," *Advanced Robotics Research*, vol. 1, no. 2, p. 202500020, 2025.
- [32] F. Visentin, F. Castellini, and R. Muradore, "A soft, sensorized gripper for delicate harvesting of small fruits," *Computers and Electronics in Agriculture*, vol. 213, p. 108202, 2023.
- [33] A. Agarwal, M. A. Mirzaee, X. Sun, and W. Yuan, "A modularized design approach for gelsight family of vision-based tactile sensors," *The International Journal of Robotics Research*, May 2025.
- [34] S. Dong, W. Yuan, and E. H. Adelson, "Improved gelsight tactile sensor for measuring geometry and slip," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 137–144.
- [35] Y. Qiu, F. Wang, Z. Zhang, K. Shi, Y. Song, J. Lu, M. Xu, M. Qian, W. Zhang, J. Wu, Z. Zhang, H. Chai, A. Liu, H. Jiang, and H. Wu, "Quantitative softness and texture bimodal haptic sensors for robotic clinical feature identification and intelligent picking," *Science Advances*, vol. 10, no. 30, p. eadp0348, 2024.
- [36] S. Nam, T. Jack, L. Y. Lee, and N. F. Lepora, "Softness prediction with a soft biomimetic optical tactile sensor," in *2024 IEEE 7th International Conference on Soft Robotics (RoboSoft)*, 2024, pp. 121–126.
- [37] W. Yuan, C. Zhu, A. Owens, M. A. Srinivasan, and E. H. Adelson, "Shape-independent hardness estimation using deep learning and a gelsight tactile sensor," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2017, p. 951–958.