Eureka (2026), 1(1): 5-25.
https://doi.org/10.63336/Eureka.35

Integrating Soft Gripper and Gripping Agent for Universal Robotic Grasping

Zhuowei Li[1,2]; Miao Zhang[3*]; Jun Yin[3]; Zhiyong Dong[3]; Yuantao Wang[4]; He Zhang[1,2]

[1] University of Nottingham Ningbo China, Ningbo 315100, China
[2] Yongjiang Laboratory, Ningbo 315202, China
[3] Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China
[4] Beijing University of Technology, Beijing, China

Corresponding author: Miao Zhang (zhangmiao@sz.tsinghua.edu.cn)

Submitted: 2025-09-25 / Accepted: 2025-11-27 / Published: 2026/2/28

Abstract: It is crucial for successful operation that robots have the ability to flexibly grasp objects and accurately estimate the point of grasp in the field of general-purpose robotic gripping. To address these challenges, we present an Adaptive Rigid-Soft Gripping Agent (GAgent) that combines the adaptability of soft grippers with the cognitive capabilities of Multimodal Language Models (MMLMs) to effectively perform gripping tasks across diverse objects and environments. Our system features a variable-stiffness soft finger that integrates silicone and a Nitinol spring, supported by a rectangular structure and a double-tendon drive mechanism. This design ensures precise and reliable grasping. Additionally, we introduce task-focused prompts and step-level reasoning to fully leverage the generalization and reasoning capabilities of MMLMs. This enables accurate object texture recognition, categorization of objects into multiple hardness levels, and appropriate stiffness modulation to dynamically adjust the gripper's rigidity, ensuring precise grip point estimation. Moreover, our framework includes Data-Driven Gripper features, integrating a vector database and continuous feedback from past experiences to refine grasping strategies over time. Experimental results validate our system's effectiveness in grasping a wide range of objects under varying lighting conditions. These advancements collectively enhance the gripper's adaptability, reliability, and efficiency in handling a broad spectrum of objects, establishing it as a versatile solution for advanced robotic tasks.

Keywords: Soft gripper, Grasping, Variable Stiffness, MMLM model, Multi Agent, Bionic Robot

1. Introduction

grasping and manipulation are essential capabilities for robotic systems, enabling them to effectively interact with their environment. These abilities are crucial for a wide range of applications, from industrial automation to service robotics. However, traditional rigid grippers often struggle with handling objects of varying shapes, sizes, and fragility.

To address these limitations, soft grippers inspired by biological systems have been developed. These grippers utilize flexible materials to adapt their shape to the objects they handle, thereby reducing the risk of damage.

Various techniques, including fluidic actuation (Z. Li et al., 2024; Sinatra et al., 2019; Zhuang et al., 2023), tendon-driven mechanisms (Y. Chen et al., 2024; Hirose et al., 2019; Lee et al., 2020; Z. Li et al., 2022), and shape memory alloys (Yang et al., 2022), have been explored to enhance their performance. Despite these advancements, achieving an optimal balance between adaptability and robustness remains challenging.

Variable stiffness mechanisms have been introduced to improve the load-bearing capacity of soft grippers while retaining flexibility. Methods such as particle jamming (Y. Li et al., 2017), low melting point alloys (Tonazzini et al., 2016), shape memory alloys (SMA) (Hu et al., 2021; Y.-F. Zhang et al., 2019), electric stimulation (Sheng & Wen, 2012), additional exoskeletons (J. Zhu et al., 2023), and antagonistic actuation mechanisms (Z. Li et al., 2025) enable the gripper to increase its stiffness when needed, providing better performance in diverse scenarios. Among these, antagonistic mechanisms stand out by enhancing stability through balanced force pairs while avoiding significant increases in system complexity or mass. Nevertheless, most other approaches often introduce added weight and mechanical or control complexity.

In parallel, there has been a growing interest in integrating cognitive capabilities into robotic systems, an approach exemplified by initiatives such as Robotgpt (Jin et al., 2024) and Graspgpt (Tang et al., 2023). At the forefront of this integration are Multimodal Models (MMLMs) and Visual-Language Models (VLMs), including MiniGPT-4 (D. Zhu et al., 2023) and LLaVA (Liu, Li, Wu, et al., 2023), which have shown exceptional performance in tasks involving natural instruction-following and visual cognition. These models enhance the interaction between robots and the physical world by providing superior generalization and advanced reasoning abilities. As a result, VLMs are increasingly being integrated into embodied agents to improve their interaction capabilities. Unlike specific-task reinforcement learning (RL) algorithms, VLM-based agents offer significant generalization capabilities (Driess et al., 2023; Liu et al., 2023; J. Wang et al., 2025; Yin et al., 2025; M. Zhang, Yin, et al., 2025), facilitated by sophisticated fine-tuning methodologies such as human demonstrations (G. Chen et al., 2023; Y. Wang et al., 2025), vision-language cross-modal connectors (K. Li et al., 2025; Liu, Li, Li, et al., 2024; M. Zhang, Fang, et al., 2025), and the continuous expansion of skill libraries (G. Wang et al., 2023). However, on-policy RL algorithms can still present challenges, particularly regarding sample efficiency. VLM-based agents, such as those discussed in (Driess et al., 2023), leverage joint training across language, vision, and visual-language domains, enabling effective transfer across diverse tasks and datasets. This capability is crucial as complex tasks often demand cogent reasoning abilities, and VLMs with emergent reasoning capabilities (Liang et al., 2025; Wei et al., 2022) are increasingly employed for complex tasks in few-shot scenarios (Wei et al., 2022; Zeng et al., 2025; M. Zhang, Shen, Yin, et al., 2024; M. Zhang, Shen, Li, et al., 2024a). Despite these advancements, VLM-based agents face environmental challenges, particularly pixel-level noise in complex lighting environments.

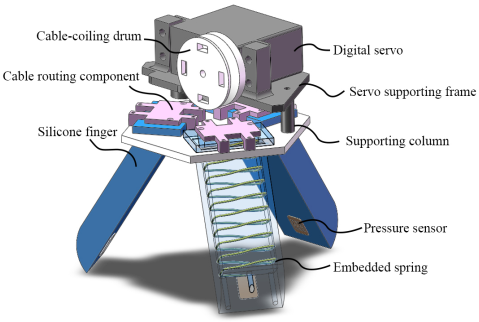


Fig. 1: Three-finger soft-rigid gripper construction including the digital servo, silicone finger, cable-coiling drum, pressure sensor, and embedded spring, designed for precise and responsive motion control.

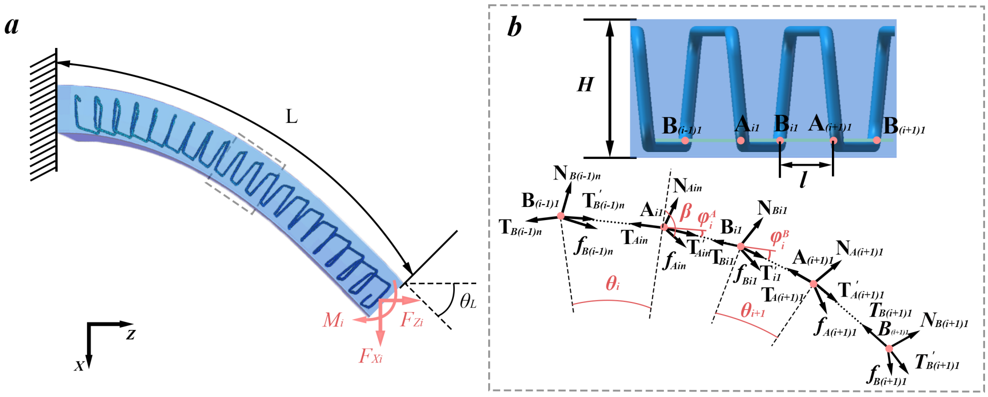


Fig. 2: Static analysis of soft-rigid finger. (a) Schematic diagram of forces on soft-rigid finger. (b) Diagram of tendon force analysis. The light green in the spring is the tendon. From point B to point A is the flexible section of the spring unit of length l and conversely from point A to point B is the rigid section of the spring. The bottom figure analyzes the tendon forces on the finger in bending. T is for tension in the tendon during transmission, N represents the support force of the silicone surface on the spring, f is the friction force, θ denotes the bending angle of the flexible unit, β is the angle between the friction force and the local coordinates, φ is the angle between the front tension and the back tension.

This paper introduces a novel soft-rigid gripper system that combines the adaptability of soft grippers with the cognitive capabilities of intelligent agents based on MMLMs. The gripper features adaptive gripping and a wide gripping range, utilizing envelope gripping and pinching. It is mounted on a continuum robotic arm for versatile object interaction. The gripper's rigidity can self-adjust in real-time based on feedback from an integrated object recognition system, enhancing its adaptability to various grasping scenarios. The key contributions of this article are as follows:

- **Variable Stiffness Soft Finger**: We developed a variable stiffness soft finger that combines silicone with a rectangular nitinol spring, addressing torsional deformation issues via a rectangular structure and a double-tendon drive mechanism, while achieving both variable stiffness and lightweight characteristics.

- **Advanced Object Recognition**: The GAgent framework integrates advanced Multimodal Models (MMLMs) with a monocular camera to accurately recognize object textures, thereby facilitating the effective capture of objects in new and unstructured environments. This framework categorizes objects into five distinct hardness levels, each corresponding to specific stiffness settings of the gripper. This precise classification enables the gripper to modulate its stiffness appropriately, ensuring a firm and secure grasp of rigid objects while applying the necessary delicacy to handle fragile items. Such an approach significantly enhances the gripper's ability to manipulate a diverse array of objects with varying textures and hardness, thereby improving its operational flexibility, adaptability, and reliability in complex and dynamic scenarios.

- **Data Driven Gripper Framework**: The framework establishes a continuous feedback loop by integrating task-focused prompts, step-level reasoning, and a vector database. This approach enhances the gripper's generalization and reasoning capabilities, while also allowing it to learn from past experiences and refine its grasping strategies over time.

The remainder of this article is structured as follows: Section II details the design of the variable rigidity soft gripper. Section III presents the static analysis of the finger and performance of variable stiffness gripper. Section IV introduces novel visual language modeling agents for grippers. The performance of the gripper is evaluated in Section V, and the findings are summarized in Section VI.

2. Bionic soft-rigid Gripper Design Concept

The design concept of the bionic soft-rigid gripper is inspired by the human hand, which has evolved to exhibit remarkable versatility. Whether grasping small items or carrying larger ones, the human hand adapts its technique according to the object's shape and characteristics, employing actions such as pinching, enveloping, hooking, and lifting. Notably, even when confronted with unfamiliar objects, the hand can skillfully grasp them without causing damage, owing to its soft-rigid and responsive structure.

Based on the study of the versatility and complex structure of the human hand (Z. Li et al., 2024), we present a new soft-rigid gripper consisting of silicone, rectangular nitinol springs, pressure sensors, tendons, and support structures, as shown in Fig. 1. The soft silicone mimics the muscle layer of the finger, acting as a cushion and adaptor that absorbs impact when gripping an object, increases friction, and adapts to the object's surface. Rectangular Nitinol springs mimic the skeleton of a human hand, supporting the hollow silicone and enabling the gripper to change stiffness to increase load carrying capacity without losing flexibility.

To achieve precise and biomimetic finger articulation, each finger of the gripper is actuated by a pair of antagonistic cables that function as artificial tendons. This dual-tendon structure design enables stable and controllable bending. These actuation tendons are anchored at the free ends of the gripper's rollers, routed through a rectangular spring, and terminate at the fingertip. This routing strategy effectively shifts the primary friction point from the tendon-silicone interface to the more robust tendon-spring interface, thereby minimizing energy dissipation and significantly amplifying the output force of the finger. The return to the neutral (open) position is facilitated by the synergistic elastic recovery of an embedded Nitinol(NiTi) springs and the inherent elasticity of the silicone matrix, eliminating the need for a dedicated return tendon.

The gripper incorporates an independent, dedicated stiffness-tuning tendon for each finger. This separate tendon is specifically designed to modulate the gripper's rigidity. By applying tension to this tendon, the embedded NiTi spring is further stretched, increasing the overall structural stiffness of the finger. This decoupled design allows for independent control of finger posture (via the antagonistic actuation tendons) and finger stiffness (via the dedicated stiffness-tuning tendon), mimicking the independent control of motion and muscle tone in the human hand. Furthermore, integrated pressure sensors at the fingertips provide real-time tactile feedback, simulating human somatosensation. This sensory input enables the gripper to dynamically adjust its grip force, allowing for the delicate and damage-free handling of fragile objects.

3. Static mechanics and variable stiffness capabilities of gripper

3.1 Static mechanics of a single finger

Since the design is the same for the three fingers, static analysis by Euler-beam model is performed for one finger. The gripper is divided into small units by flexible and rigid modules, as shown in Fig. 2. When a finger grasps an object, the tension transfer relationship of the tendon can be expressed by Eq. (1).

$$\begin{cases} \alpha_i = \frac{T'_{Ain}}{T_{Ain}} \\ \alpha_i = \mu^2\left(1-\gamma_2\right) - \frac{\sqrt{2}}{2}\mu\gamma_1\sqrt{\frac{2\mu^2(1-\gamma_2)+4}{1+\gamma_2}} + 1 \\ \gamma_1 = \sin\varphi_i^A \\ \gamma_2 = \cos\varphi_i^A \end{cases} \tag{1}$$

where α is the transmission relation of tension in a single beam model and μ is the coefficient of friction. The transmission of tendon tension from the proximal to the distal tip of the finger can be expressed by Eq. (2).

$$\begin{cases} T'_{Ain} = \alpha_i^A T_{Ain} \\ T'_{Ain} = T_{Bin} \\ T'_{Bin} = \alpha_i^B T_{Bin} \\ T'_{Bin} = T_{Ai(n+1)} \end{cases} \tag{2}$$

When adding the tensions in the 20 flexible and rigid cells, it is clear that the output force of the motor is equal to the tension acting at the first node:

$$\mathbf{F}_{Ain} = \mathbf{T}_{Ain} \tag{3}$$

The force on the i elastic beam unit can be decomposed as:

$$\left\{\begin{aligned} F_{zi} &= \left|\mathbf{F}_{Ain}\right| \cos\left(\arctan 2\left(x_i, z_i\right)\right) \\ F_{xi} &= \left|\mathbf{F}_{Ain}\right| \sin\left(\arctan 2\left(x_i, z_i\right)\right) \\ M_i &= \tfrac{H}{2}\left(\left|F_{zi}\right| - \left|F_{xi}\right|\right) \end{aligned}\right. \tag{4}$$

Based on the above analysis, the end coordinates of the flexible beam are characterized by a beam constraint model (Awtar & Slocum, 2007; Wu et al., 2024). According to the dimensionless beam formula for a rectangular cross-section and given any set of F_{xi}, F_{zi}, M_i, the beam end coordinates $\hat{z}(1)$, $\hat{x}(1)$, $\overset{\wedge}{\theta}(1)$ can be obtained as:

$$\begin{aligned} \begin{bmatrix} \widehat{x}(1) \\ \hat{\theta}(1) \end{bmatrix} &= \left(\mathbf{A} - F_{zi}\mathbf{B}\right)^{-1} \begin{bmatrix} F_{xi} \\ M_i \end{bmatrix} \\ \hat{z}(1) &= \tfrac{H^2}{12} F_{zi} + \begin{bmatrix} \widehat{x}(1) \\ \hat{\theta}(1) \end{bmatrix}^T \mathbf{C} \begin{bmatrix} \widehat{x}(1) \\ \hat{\theta}(1) \end{bmatrix} \\ & -F_{zi} \begin{bmatrix} \widehat{x}(1) \\ \hat{\theta}(1) \end{bmatrix} \mathbf{D} \begin{bmatrix} \widehat{x}(1) \\ \hat{\theta}(1) \end{bmatrix} + 1 \end{aligned} \tag{5}$$

where $\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}$ are dimensionless characteristic coefficients of the beam constraint model, which can be found in (Awtar, 2003). The deflected shape of the beam can be obtained from the following equation:

$$\begin{aligned} \hat{z}\left(\hat{s}\right) &= \hat{s} - \tfrac{H^2 \hat{s}}{12} F_{zi} - \begin{bmatrix} F_{xi} \\ M_i \end{bmatrix}^T \mathbf{C}^* \begin{bmatrix} F_{xi} \\ M_i \end{bmatrix} \\ \begin{bmatrix} \widehat{x}(s) \\ \hat{\theta}(s) \end{bmatrix} &= \mathbf{K}^* \begin{bmatrix} F_{xi} \\ M_i \end{bmatrix} \end{aligned} \tag{6}$$

where $\hat{s} \in [0, 1]$, matrix $\mathbf{C}^*$ and $\mathbf{K}^*$ can be found in (Ma & Chen, 2016). Quantizing the above dimensionless formula to get:

$$\left\{\begin{aligned} z(s) &= \hat{z}\left(\hat{s}\right) l \\ x(s) &= \hat{x}\left(\hat{s}\right) l \\ \theta(s) &= \overset{\wedge}{\theta}\left(\hat{s}\right) \end{aligned}\right. \tag{7}$$

where $s \in [0, l]$. The coordinates and angles of each unit can be calculated using Eq. (7).

For the curved shape of the soft-rigid gripper, the angles of the 20 flexible beams can be summed up to get the angle of curvature of the whole finger θ_L:

(8)

$$\theta_L = \sum_{i=1}^{20} \theta_i$$

The results of static simulations, Abaqus simulations, and physical experiments show similar outcomes (Fig. 4). In the Abaqus simulation, the fingers are set with a preload so that there is a closer resemblance to reality in the statics simulation. To ensure the fingers achieve an enveloping grasp on an object, the contact points of the gripper can be evaluated using Eq. (9) :

$$r = \frac{L^2 - R^2 + 2RL}{2\left(R + \frac{L}{2}\right)} \tag{9}$$

where r is the bending radius of the three fingers and R is the radius of the object being grasped.

The stiffness of the entire gripper can be expressed as :

$$k_T = \frac{k_s k_t k_m}{k_t k_m + k_s k_m + k_s k_t} \tag{10}$$

k_T is the total stiffness coefficient, k_s is the stiffness coefficient of the spring, k_t is the stiffness coefficient of the tendon and k_m is the stiffness coefficient of the soft material.

The classification of the gripper's variable stiffness into five categories corresponding to the different stiffnesses of the items is intended to cover a wide range of stiffnesses when applying the variable stiffness grippers. This classification maintains accuracy and practicality, corresponding to super-soft, soft, medium-hard, hard, and super-hard items. A summary of these categories is provided in Table 1.

3.2 Variable stiffness analysis of finger

The ability of the human hand to comfortably cope with everything from fragile objects to heavy grips is due to the variable rigidity of the human hand. In order to evaluate the variable stiffness capability of this design, the gripper design was subjected to finite element analysis and compared with a purely flexible design as well as a rigid design, and a stability analysis of the gripper was carried out.

We performed finite element analysis on a soft silicone finger, a hard nitinol finger, and a rigid-soft finger to assess the gripper's ability to change stiffness, as shown in Fig.3(b)-(d). It is clear to see that the soft silicone finger bends at the root of the finger, which reduces the gripper's ability to fit the object, compared to a rigid-soft finger embedded with nitinol, which spreads the stress over the entire finger and bends at a more linear angle. This not only enables more controlled grasping, but also facilitates safer interactions with objects, as shown in Fig. 3(a).

The pure silicone finger stops moving after attaining a vertical position of 45mm. This could be attributed to the fact that the flexible material ($DragonSkin10$) is incompressible, resulting in a significant reverse force on the finger's substructure at this position, rendering it unable to bend. In contrast, the spring-silicone coupled finger achieves a maximum vertical displacement of 62mm, with the spring bending up to 55mm. Upon comparing the vertical and horizontal displacements of all three, it appears that the bending curvature of the spring-silicone coupled finger exhibits greater linearization.

To corroborate the precision of the simulation, the simulated tension is increased and compared with the actual bending. Concurrently, we record the three coordinate positions of the actual finger trajectory corresponding to $0°$, $45°$, and $90°$ of motor rotation to compare with the simulation curve. The actual trajectories of finger displacement basically coincide with those of the theoretical simulation and the finite element simulation, as shown in Fig. 4.

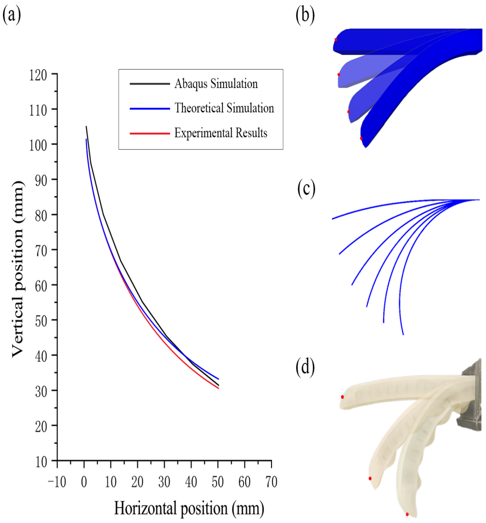


Fig. 3: Displacement motion of coupled fingers. (a) Displacement curves for pure silicone, spring and spring-silicone coupled finger. (b) The movement process of pure silicone fingertips. (c) The process of movement of the end of the spring. (d) Silicone spring coupled finger movement process.

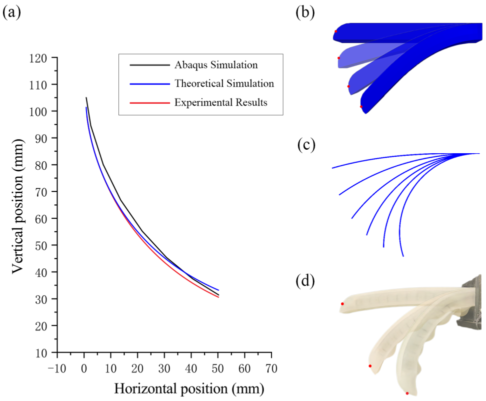


Fig. 4: Comparison of soft-rigid finger behavior: Theoretical, Abaqus Simulation, and Experimental Results. (a) The red curve represents the motion trajectory of the physical soft-rigid finger, the blue curve represents the trajectory curve of the statics simulation, and the black curve is the motion trajectory simulated in Abaqus. (b) soft-rigid finger movement in

Abaqus, red points are trajectory tracking points. (c) Trajectory of the finger movement in the statics simulation. A pre-tension is added at the initial position on the tendon driven. (d) Motion trajectory diagram during physical motion.

3.3 Stabilization of variable stiffness finger

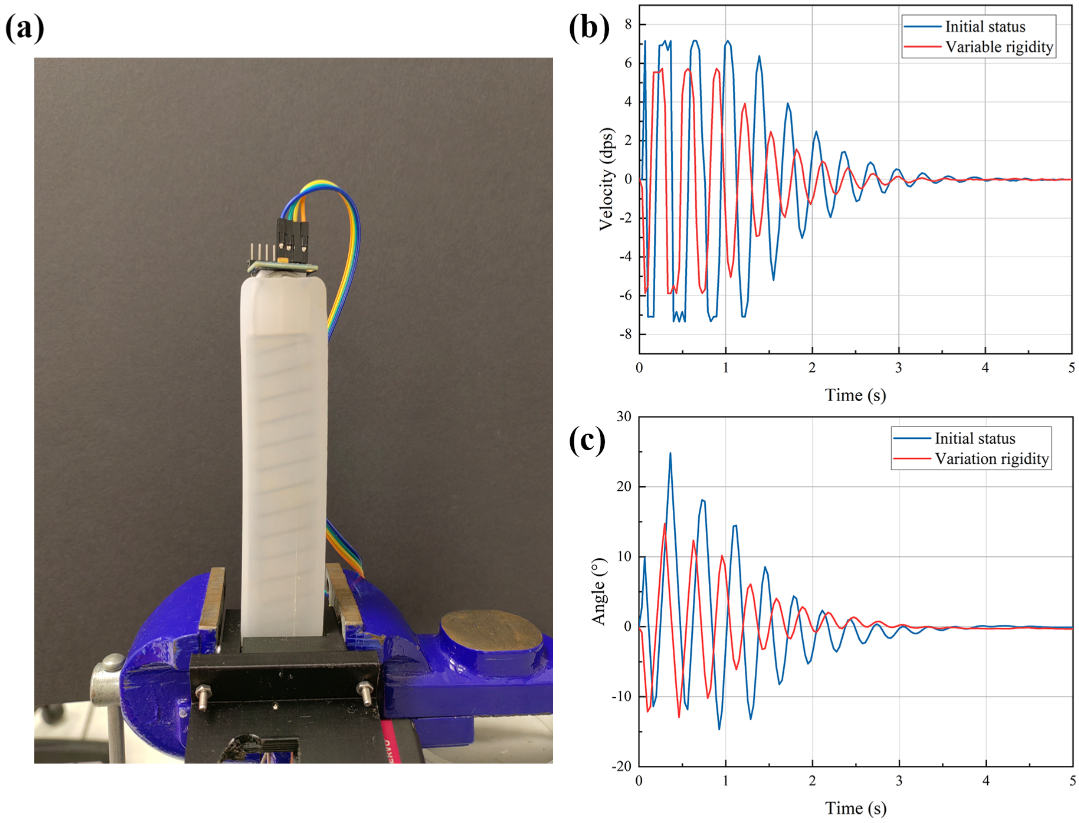


Fig. 5: Rigid-soft finger stability test. (a) The finger is mounted in a bench vise and the IMU at the fingertip collects the angle and velocity signals from the finger after it is struck. (b) Velocity change graph for the initial state and after pre-tensioning. (c) graphs of velocity changes in the initial state and after pre-tensioning.

The stability of a gripper stands as a critical measure of its performance. Given that grippers are used in industrial settings or implemented on robots, it is unavoidable that they will confront collisions that could impact the equipment. For instance, Vibrations formed by collisions in fast-moving or unstructured environments can lead to a decrease in grasping results. Achieving rapid stabilization after an unexpected collision is one of the most important metrics for evaluating grippers. To broaden the applicability of rigid-soft grippers across various spheres, the authors conducted experimentation on the stability of grippers in response to interference, comparing both the time it takes to regain stability before and after spring pre-tensioning and the angular range of change.

In this experiment, we mounted the MPU6050 at the end of the fingertip, and the rigid-soft finger was mounted in a bench vise, as shown in Fig. 5(a). We collected data from IMU by repeatedly tapping fingers for comparison.

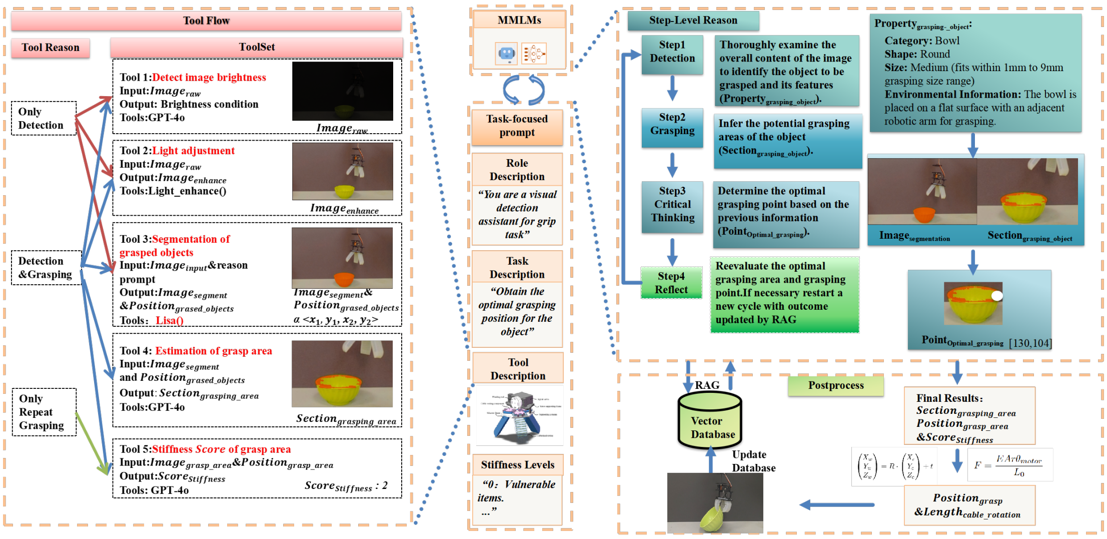
In the experimental crash tests, we perform impact experiments with the same force for the initial state and the variable rigidity state, where the initial and variable rigidity states correspond to levels 0 and 4 in Table 1. As can be clearly seen in Fig. 5(b), under variable stiffness, the initial velocity of the finger after impact is 5.8dps

compared to that of the initial state can be reduced by about 20% of the impact velocity. The initial state can bend up to a maximum angle of 25° after impact and shows an unstable state on both sides, while after the variable stiffness, it is clearly seen that the two sides of the shock are stable after impact, as shown in Fig. 5(c). In the variable stiffness state, the finger was able to return to a stable state in roughly 2.5s, which is a 45% improvement in performance compared to the initial state which stabilized after 4.5s. Consequently, rigid-soft fingers return to a steady position more quickly after tightening, making it easier to design control systems and perform faster operations for mobile robotics, industrial production, and domestic applications.

4. An Adaptive Rigid-Soft Gripping Agent with Vision Language Models

Based on the previously designed soft gripper, we also have developed an intelligent grasping framework, as shown in Fig. 6, which incorporates a Visual-language Model equipped with an adaptive rigidity gripper, optimized for precision grasping. The system inputs include captured images and relevant prior information, while the output consists of the optimal grasping point and stiffness score (see Table 1).

In detail, the framework comprises A Multi-Modal Large Language Model (MMLM), a Task-focused prompt, Step-level reason, Task Flow, and a Vector database.



Tool Flow
Tool Reason
ToolSet
Only Detection
Detection &Grasping
Only Repeat Grasping
Tool 1:Detect image brightness
Input:$Image_{raw}$
Output: Brightness condition
Tools:GPT-4o
$Image_{raw}$
Tool 2:Light adjustment
Input:$Image_{raw}$
Output:$Image_{enhance}$
Tools:Light_enhance()
$Image_{enhance}$
Tool 3:Segmentation of grasped objects
Input:$Image_{input}$&reason prompt
Output:$Image_{segment}$ &$Position_{grased\_objects}$
Tools: Lisa()
$Image_{segment}$& $Position_{grased\_objects}$
$a <x_1, y_1, x_2, y_2>$
Tool 4: Estimation of grasp area
Input:$Image_{segment}$ and $Position_{grased\_objects}$
Output: $Section_{grasping\_area}$
Tools:GPT-4o
$Section_{grasping\_area}$
Tool 5:Stiffness Score of grasp area
Input:$Image_{grasp\_area}$&$Position_{grasp\_area}$
Output:$Score_{Stiffness}$
Tools: GPT-4o
$Score_{Stiffness}: 2$
MMLMs
Task-focused prompt
Role Description
"You are a visual detection assistant for grip task"
Task Description
"Obtain the optimal grasping position for the object"
Tool Description
Stiffness Levels
"0: Vulnerable items. ..."
Step-Level Reason
Step1 Detection
Thoroughly examine the overall content of the image to identify the object to be grasped and its features ($Property_{grasping\_object}$).
Step2 Grasping
Infer the potential grasping areas of the object ($Section_{grasping\_object}$).
Step3 Critical Thinking
Determine the optimal grasping point based on the previous information ($Point_{Optimal\_grasping}$).
Step4 Reflect
Reevaluate the optimal grasping area and grasping point.If necessary restart a new cycle with outcome updated by RAG
$Property_{grasping-\_object}$:
Category: Bowl
Shape: Round
Size: Medium (fits within 1mm to 9mm grasping size range)
Environmental Information: The bowl is placed on a flat surface with an adjacent robotic arm for grasping.
$Image_{segmentation}$
$Section_{grasping\_object}$
$Point_{Optimal\_grasping}$ [130,104]
RAG
Vector Database
Update Database
Postprocess
Final Results:
$Section_{grasping\_area}$
$Position_{grasp\_area}$
&$Score_{Stiffness}$
$\begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} = R \cdot \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} + t$
$F = \frac{EAr\theta_{motor}}{L_0}$
$Position_{grasp}$ &$Length_{cable\_rotation}$

Fig. 6: An Adaptive Rigid-Soft Gripping Agent with Vision Language Models, called GAgent. The framework integrates task-focused prompts and step-level reasoning to determine optimal grasping positions. It begins with a task-focused prompt, detailing the role, task description, and tool description. The step-level reasoning involves four steps: Detection, grasping, critical thinking, and reflection. The tool flow section includes a variety of tools for detecting image brightness, adjusting light, segmenting grasped objects, estimating grasp areas, and scoring the stiffness of grasp areas. The postprocess involves determining the final grasping position, calculating the required cable rotation length for precise manipulation, and updating the vector database with the optimal grasping section, position, and stiffness score.

- MMLM has been utilized by GPT-4o, which is attributed to its powerful reasoning capabilities. It converts complex instructions combined with Task-focused prompts and step-level reason into a position of grasp area and stiffness score, which can guide the soft grasp to complete the grasping task.
- The task-focused prompt compiles task descriptions, role descriptions, tool descriptions, and basic operational guidelines into a coherent prompt that the MMLM can then process.

- The step-level reasoning process consists of four iterative stages: detection, grasping, critical thinking, and reflection. The detection stage examines the overall content of the image to identify the object and its features. During the grasping stage, potential grasping areas of the object are inferred using tools from the ToolSet. The critical thinking stage involves determining the optimal grasping point based on the previously gathered information. Finally, in the reflection stage, the optimal grasping area and point are reevaluated. If necessary, a new cycle can be initiated, incorporating updates from Retrieval Augmented Generation (RAG).

- Tool flow for robotic grasping, integrating various tools to enhance precision and effectiveness. The process is broken down into several steps under three main categories: Only Detection, Detection & Grasping, and Only Grasping.

- Create a continuous feedback loop using a vector database where data retrieval, augmentation, generation, and database updates work together to improve the performance of machine learning models.

In summary, the framework's innovation lies in integrating a large multi-modal model with strong generalization capabilities, step-level reasoning with powerful reasoning ability, and various tools including a low-light enhancement algorithm (M. Zhang, Shen, Li, et al., 2024b), reasoning segmentation algorithm (Lai et al., 2024). This enables efficient operation across a wide range of environments. Simultaneously, the soft gripper ensures precise and flexible object manipulation.

>

Table 1: Hardness classification of soft gripper.

Stiffness Levels	**0-super-soft**	**1-soft**	**2-medium-hard**	**3-hard**	**4-super-hard**
Objects	Vulnerable items such as jelly, potato chips, persimmons, etc.	Easily deformed items such as fruits, plastic packaging, etc.	Tough and easily deformed items such as leather, towels, etc.	Hard plastic or wooden items such as tool box, glass jar, etc.	Heavy objects such as weights, dumbbells, and other metal objects.

5. Experimental Validation of soft-rigid Gripper Performance

5.1 Comprehensive grasping experiments

Capability validation experiments were conducted based on the Amazon Picking Challenge (Hernandez et al., 2017), with 45 items added for grasping experiments. These items were used to verify the gripping range and load capacity of the soft-rigid gripper. Jelly and pudding were the softest items, while potato chips and seaweed were the lightest. The smallest items were small steel balls and coins, while the heaviest and hardest were weights. The largest item was a bucket of water, as well as a spiked durian shell (see Fig. 7).

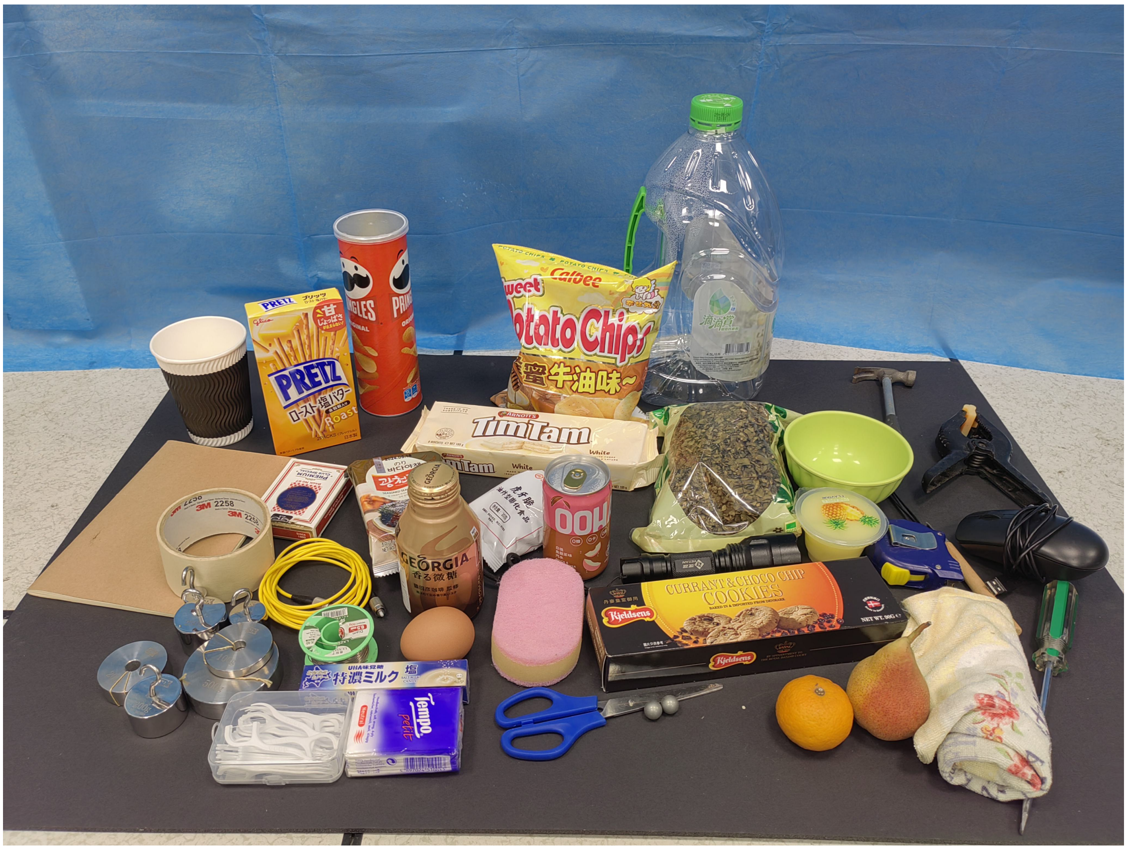


Fig. 7: Real-world grasping experiments with household objects and grasping objects in industrial scenarios.

The soft-rigid gripper was mounted on a continuous robotic arm for comprehensive gripping experiments to evaluate its structural performance. A total of 46 distinct objects, including items of various weights, sizes, and materials (e.g., soft objects, small regular objects, heavy items, long objects, and oversized items), were selected for testing. For each object, 10 repeated grasping trials were conducted, resulting in 460 total attempts. In these experiments, the manipulator operated without a MMLM model, relying on an experience-informed capture strategy while varying the stiffness threshold of the pressure sensor. A grasp was deemed successful if the object remained stable in the air for 5 seconds after lifting and was then smoothly lowered; otherwise, it was counted as a failure.

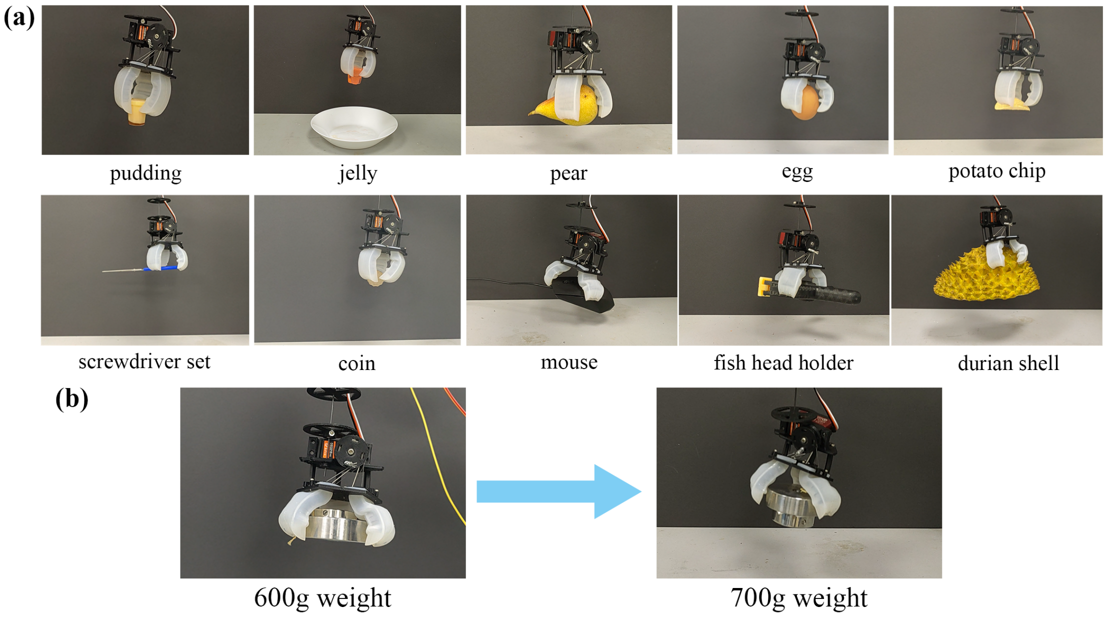


Fig. 8: Grab a display of items that are representative of the items. (a) The gripper can easily cope with objects of different weights and hardnesses, from jelly to durian shells, by changing its rigidity. (b) Improvement in gripping weight capacity from 0 level to 4 levels of variable stiffness.

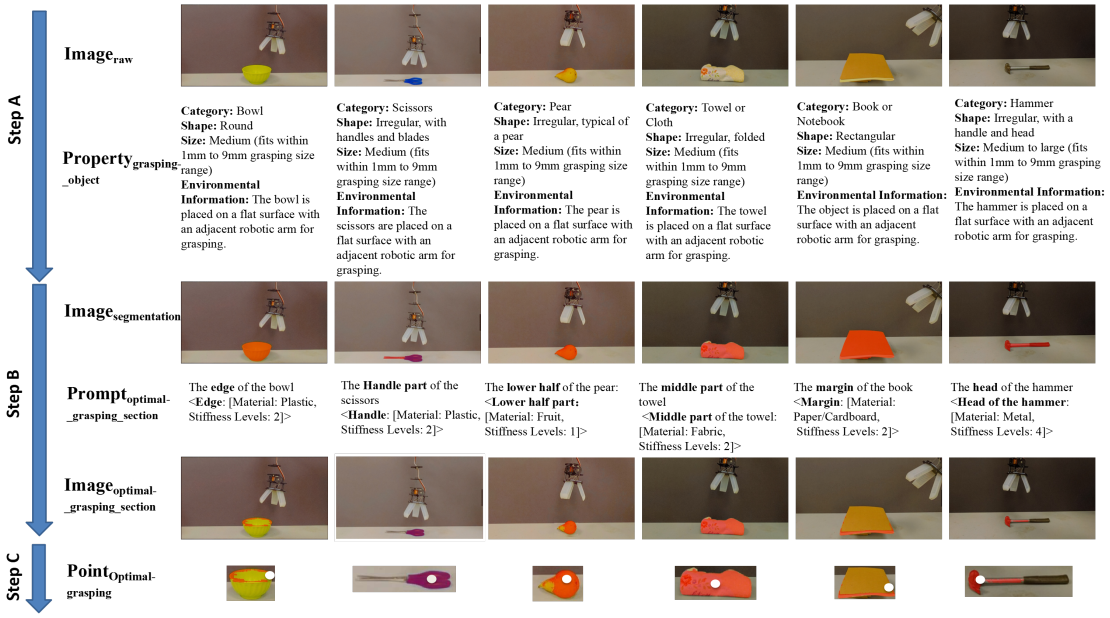
The results demonstrate that the gripper is capable of handling a wide range of objects, with representative successful grasps shown in Fig. 8(a). Overall, the experience-informed approach achieved a high success rate of 88.5 ± 4.1% across all trials. The gripper exhibits efficient performance with soft, small, and regular objects, particularly those smaller than itself. For heavy objects, it maintains a maximum load capacity of 600 g while preserving structural rigidity; by adjusting the spring stiffness, this capacity can be increased to 700 g without risk of detachment, as illustrated in Fig. 8(b).

5.2 Qualitative Analysis of GAgent

This study introduces GAgent to enhance the flexibility and versatility of soft-rigid grippers' grasping capabilities. This section will qualitatively analyze GAgent's overall workflow. As shown in Fig. 9, six objects with diverse shapes were selected for grasping, including a bowl, scissors, books, and others. Step A involves inputting captured images to reason the basic characteristics of objects $\textbf{Property}_{\textit{grasping-object}}$ to be grasped, including category, shape, size, and environmental information. In Step B, we use the results from Step A as the input prompt for the reasoning segmentation algorithm (Lai et al., 2024), which generates a segmentation map $\textbf{Image}_{\textit{segmentation}}$ of the target object, and subsequently produces $\textbf{Prompt}_{\textit{optimal-grasping-section}}$. Afterward, we input $\textbf{Prompt}_{\textit{optimal-grasping-section}}$ back into the reasoning segmentation algorithm to determine the optimal gripping area $\textbf{Image}_{\textit{optimal-grasping-section}}$. For a bowl, the optimal gripping area is located at the edge; for scissors, it is the plastic part; for a pear, it is the lower half; and for a book, it is the edge. Moving to Step 3, we further extract the optimal gripping points $\textbf{Point}_{\textit{optimal-grasping}}$ within these optimal gripping areas, as indicated by the white dots in the last row.

5.3 Quantitative Analysis of GAgent

This section conducts a Quantitative Analysis of GAgent. Table 2 presents success rates for nine objects using different versions of the GAgent system. The table shows success rates for recognizing and handling different objects, including Bowl, Pear, Book, Cookies, Scissors, Hammer, Towel, Mouse, and Coin. Success rates generally increase as more detection steps are employed, with GAgent_All achieving highest success rates across all objects. For example, the success rate for the Bowl object increases from 20% with GAgent_S1 to 90% with GAgent_All. Similarly, for the object Coin, the success rate increases from 0% with GAgent_S1 to 70% with GAgent_All. This indicates the effectiveness of using multiple detection steps in improving the gripper's performance across a variety of objects.


Step A
Step B
Step C
$\text{Image}_{\text{raw}}$
$\text{Property}_{\text{grasping-\_object}}$
$\text{Image}_{\text{segmentation}}$
$\text{Prompt}_{\text{optimal-\_grasping\_section}}$
$\text{Image}_{\text{optimal-\_grasping\_section}}$
$\text{Point}_{\text{Optimal-grasping}}$

Category: Bowl
Shape: Round
Size: Medium (fits within 1mm to 9mm grasping size range)
Environmental Information: The bowl is placed on a flat surface with an adjacent robotic arm for grasping.

Category: Scissors
Shape: Irregular, with handles and blades
Size: Medium (fits within 1mm to 9mm grasping size range)
Environmental Information: The scissors are placed on a flat surface with an adjacent robotic arm for grasping.

Category: Pear
Shape: Irregular, typical of a pear
Size: Medium (fits within 1mm to 9mm grasping size range)
Environmental Information: The pear is placed on a flat surface with an adjacent robotic arm for grasping.

Category: Towel or Cloth
Shape: Irregular, folded
Size: Medium (fits within 1mm to 9mm grasping size range)
Environmental Information: The towel is placed on a flat surface with an adjacent robotic arm for grasping.

Category: Book or Notebook
Shape: Rectangular
Size: Medium (fits within 1mm to 9mm grasping size range)
Environmental Information: The object is placed on a flat surface with an adjacent robotic arm for grasping.

Category: Hammer
Shape: Irregular, with a handle and head
Size: Medium to large (fits within 1mm to 9mm grasping size range)
Environmental Information: The hammer is placed on a flat surface with an adjacent robotic arm for grasping.

The **edge** of the bowl <**Edge**: [Material: Plastic, Stiffness Levels: 2]>

The **Handle part** of the scissors <**Handle**: [Material: Plastic, Stiffness Levels: 2]>

The **lower half** of the pear: <**Lower half part:** [Material: Fruit, Stiffness Levels: 1]>

The **middle part** of the towel <**Middle part** of the towel: [Material: Fabric, Stiffness Levels: 2]>

The **margin** of the book <**Margin**: [Material: Paper/Cardboard, Stiffness Levels: 2]>

The **head** of the hammer <**Head of the hammer**: [Material: Metal, Stiffness Levels: 4]>

Fig. 9: Multi-step process for optimal robotic grasping of various objects.The figure depicts the process of determining optimal grasping points for a three-finger to handle different objects. Each column represents a distinct object, and the text details the stages. The process involves analyzing the object's properties, segmenting the image, identifying and visualizing the optimal grasping section, and determining the specific point of grasp. This systematic approach integrates image processing, object property analysis, and precision in robotic manipulation to achieve optimal grasping for various everyday objects.

5.4 Ablation analysis of GAgent

This gripper is also attached to a section of the continuous robotic arm for gripping experiments. Our goal is to replicate various complex real-world scenarios with spontaneously changing outdoor conditions. Throughout these experiments, our innovative gripper self-adjusts its rigidity. By integrating object recognition feedback from a single-lens camera with the MMLM model, the gripper can adeptly respond to an array of grasping dynamics.

To evaluate the effectiveness of our new model, we conducted an ablation study comparing its performance against GPT-4o and manual operations. The study focused on various metrics, including accuracy and robustness across different tasks.

1. **Accuracy**: Our GAgent framework demonstrated significantly improved accuracy over pure GPT-4o and manual operations. Specifically, it achieved an average accuracy of 93.7% compared to 88.5% for manual operations and different development stages. Fig. 10 shows specific values, with details of each development stage as follows:

    ◦ **Manual Operations**: Average success rate of 88.5 ± 4.1%.

    ◦ **S1 (Step A, Pure GPT-4o for Object Recognition)**: Success rate of 30.0 ± 3.5%, only capable of recognizing objects without providing positional information.

    ◦ **S2 (Step A+B, S1 with Reasoning Segmentation)**: Success rate improved to 53.9 ± 2.7%, accurately identifying the boundaries of objects.

    ◦ **S3 (Step A+B+C, S2 with Refined Grasp Boundaries)**: Success rate increased to 85.4 ± 3.2%, further refining the grasp boundaries.

    ◦ **All (All steps, S3 with Precise Finger Grasp Points, RAG, and Rigidity Adjustment)**: Success rate reached 93.7 ±3.1%, incorporating precise grasp points for each finger along with item and gripper rigidity adjustments, as shown in Fig. 9.

- **Robustness**: The robustness of GAgent was tested across various adverse light conditions. The new model maintained consistent performance with a 2% drop in accuracy under adverse conditions, whereas GPT-4o showed a 5% drop. This indicates that our agent is more resilient to input variations and can handle real-world data more effectively.

- **Error Handling**: In failed gripper tasks, GAgent can record each failed interaction and reflective case to form an independent vector database, conferring our framework with a strong capacity for error correction and handling of gripper problems. GAgent_All has risen to 93.7% from 85.4%.

Table 2: Success rates for various objects using different versions of GAgent. GAgent_S1 refers to the agent utilizing step 1 detection, GAgent_S2 indicates the agent using both steps 1 and 2, GAgent_S3 represents the agent employing all three steps (1, 2, and 3), and GAgent_All denotes the agent utilizing all available steps.

Objects	GAgent_S1	GAgent_S2	GAgent_S3	GAgent_All
Bowl	20	40	80	90
Pear	30	50	90	100
Book	0	10	70	70
Cookies	40	70	100	100

Objects	GAgent_S1	GAgent_S2	GAgent_S3	GAgent_All
Scissor	10	20	70	80
Hammer	20	40	80	100
Towel	40	80	100	100
Mouse	20	40	80	100
Coin	0	60	70	70

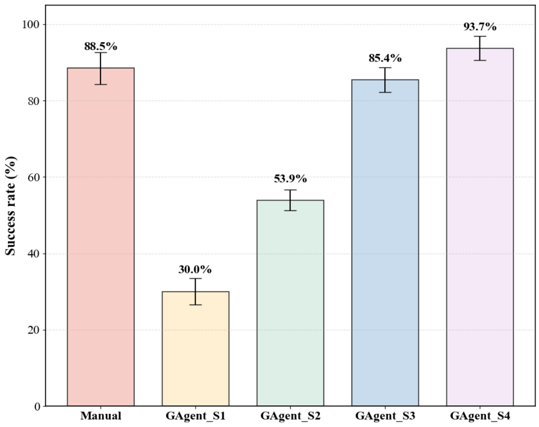


Fig. 10: Total success rate for 45 daily necessities and industrial tools when using different versions of GAgent to grip objects. The graph shows a clear improvement in success rate with each additional step, ending with an 93.7±3.1% success rate for GAgent_All.

6. Conclusion

In this study, we designed a multi-stage adjustable variable stiffness soft gripper composed of a rectangular nitinol spring, silicone, tendons, and motors. The gripper can perform multi-stage stiffness adjustments by stretching the nitinol spring and addresses the torsion problem of soft materials through double tendon balance control. We employed the Euler-beam model theory for static analysis of the gripper and conducted finite element analysis to examine changes in its stiffness. Our data-driven framework incorporates task-centered cues, step-level reasoning, and a vector database, creating a continuous feedback loop that enhances the accuracy of the gripper's operations. Additionally, our GAgent classifies objects into five categories based on their softness and hardness, enabling precise adjustment of the gripper's stiffness for accurate gripping. These advancements collectively enhance the gripper's adaptability, reliability, and efficiency in handling a wide range of objects, making it a versatile solution for advanced robotic tasks. Our system represents a significant step forward in robotic grasping technology, offering improved performance and adaptability in real-world applications.

Acknowledgment

The work described in this paper is partially supported by Shenzhen Fundamental Research (General Program)(WDZC20231129163533001) and Hunan Provincial Transportation Science and Technology Project (No.202403).

References

[1] Awtar, S., & Slocum, A. H. (2007). *Constraint-based design of parallel kinematic XY flexure mechanisms*.

[2] Chen, G., Cui, T., Zhou, T., Peng, Z., Hu, M., Wang, M., Yang, Y., & Yue, Y. (2023). *Human demonstrations are generalizable knowledge for robots*.

[3] Chen, Y., Yao, S., Meng, M. Q.-H., & Liu, L. (2024). Chained spatial beam constraint model: A general kinetostatic model for tendon-driven continuum robots. *IEEE/ASME Transactions on Mechatronics*.

[4] Driess, D., Xia, F., Sajjadi, MehdiS. M., Lynch, C., Chowdhery, A., Ichter, B., Wahid, A., Tompson, J., Vuong, Q., Yu, T., Huang, W., Chebotar, Y., Sermanet, P., Duckworth, D., Levine, S., Vanhoucke, V., Hausman, K., Toussaint, M., Greff, K., … Florence, P. (2023). *PaLM-e: An embodied multimodal language model*.

[5] Hernandez, C., Bharatheesha, M., Ko, W., Gaiser, H., Tan, J., Deurzen, K. van, Vries, M. de, Van Mil, B., Egmond, J. van, Burger, R., et al. (2017). Team delft's robot winner of the amazon picking challenge 2016. *RoboCup 2016: Robot World Cup XX 20*, 613–624.

[6] Hirose, T., Kakiuchi, Y., Okada, K., & Inaba, M. (2019). Design of soft flexible wire-driven finger mechanism for contact pressure distribution. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4699–4705.

[7] Hu, Q., Huang, H., Dong, E., & Sun, D. (2021). A bioinspired composite finger with self-locking joints. *IEEE Robotics and Automation Letters*, *6*(2), 1391–1398.

[8] Jin, Y., Li, D., Yong, A., Shi, J., Hao, P., Sun, F., Zhang, J., & Fang, B. (2024). Robotgpt: Robot manipulation learning from chatgpt. *IEEE Robotics and Automation Letters*.

[9] Lai, X., Tian, Z., Chen, Y., Li, Y., Yuan, Y., Liu, S., & Jia, J. (2024). Lisa: Reasoning segmentation via large language model. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9579–9589.

[10] Lee, K., Wang, Y., & Zheng, C. (2020). Twister hand: Underactuated robotic gripper inspired by origami twisted tower. *IEEE Transactions on Robotics*, *36*(2), 488–500.

[11] Li, K., Wang, J., Zhang, M., & Wang, X. (2025). OMR-diffusion: Optimizing multi-round enhanced training in diffusion models for improved intent understanding. *arXiv Preprint arXiv:2503.17660*.

[12] Li, Y., Chen, Y., Yang, Y., & Wei, Y. (2017). Passive particle jamming and its stiffening of soft robotic grippers. *IEEE Transactions on Robotics*, *33*(2), 446–455.

[13] Li, Z., Lin, X., Zhu, Z., Zhu, Y., Zhou, Y., Li, J., Gerada, C., Zhang, H., & Zhuang, S. (2025). Soft micromanipulation robot for real-time adaptive multimodal operation. *Advanced Science*, e15784.

[14] Li, Z., Yin, M., Huang, B., Cai, Z., Yi, Z., & Wu, X. (2024). A spring-based rigid-soft robotic gripper for conformal grasping and object recognition. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, *46*(4), 249.

[15] Li, Z., Yin, M., Sun, H., Hu, M., Cao, W., & Wu, X. (2022). Master-slave control of the robotic hand driven by tendon-sheath transmission. *International Conference on Intelligent Robotics and Applications*, 747–758.

[16] Liang, X., Tao, M., Xia, Y., Wang, J., Li, K., Wang, Y., He, Y., Yang, J., Shi, T., Wang, Y., et al. (2025). SAGE: Self-evolving agents with reflective and memory-augmented abilities. *Neurocomputing*, 130470.

[17] Liu, H., Li, C., Li, Y., & Lee, Y. (2024). *Improved baselines with visual instruction tuning*.

[18] Liu, H., Li, C., Wu, Q., & Lee, Y. J. (2023). Visual instruction tuning. *arXiv Preprint arXiv:2304.08485*.

[19] Liu, H., Li, C., Wu, Q., Lee, Y., -Madison, -M., & Research, M. (2023). *Visual instruction tuning*.

[20] Ma, F., & Chen, G. (2016). Modeling large planar deflections of flexible beams in compliant mechanisms using chained beam-constraint-model. *Journal of Mechanisms and Robotics*, *8*(2), 021018.

[21] Sheng, P., & Wen, W. (2012). Electrorheological fluids: Mechanisms, dynamics, and microfluidics applications. *Annual Review of Fluid Mechanics*, *44*, 143–174.

[22] Sinatra, N. R., Teeple, C. B., Vogt, D. M., Parker, K. K., Gruber, D. F., & Wood, R. J. (2019). Ultragentle manipulation of delicate structures using a soft robotic gripper. *Science Robotics*, *4*, eaax5425.

[23] Tang, C., Huang, D., Ge, W., Liu, W., & Zhang, H. (2023). Graspgpt: Leveraging semantic knowledge from a large language model for task-oriented grasping. *IEEE Robotics and Automation Letters*.

[24] Tonazzini, A., Mintchev, S., Schubert, B., Mazzolai, B., Shintake, J., & Floreano, D. (2016). Variable stiffness fiber with self-healing capability. *Advanced Materials*, *28*(46), 10142–10148.

[25] Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., Fan, L., Anandkumar, A., Nvidia, N., Caltech, C., Austin, U., & Asu, A. (2023). *VOYAGER: An open-ended embodied agent with large language models*.

[26] Wang, J., He, Y., Li, K., Li, S., Zhao, L., Yin, J., Zhang, M., Shi, T., & Wang, X. (2025). MDANet: A multi-stage domain adaptation framework for generalizable low-light image enhancement. *Neurocomputing*, *627*, 129572.

[27] Wang, Y., He, Y., Wang, J., Li, K., Sun, L., Yin, J., Zhang, M., & Wang, X. (2025). Enhancing intent understanding for ambiguous prompt: A human-machine co-adaption strategy. *Neurocomputing*, 130415.

[28] Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, EdH., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022). *Emergent abilities of large language models*.

[29] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q., & Zhou, D. (2022). *Chain of thought prompting elicits reasoning in large language models*.

[30] Wu, K., Zheng, G., Chen, G., & Awtar, S. (2024). A body-frame beam constraint model. *Mechanism and Machine Theory*, *192*, 105517.

[31] Yang, J., Sun, S., Yang, X., Ma, Y., Yun, G., Chang, R., Tang, S.-Y., Nakano, M., Li, Z., Du, H., et al. (2022). Equipping new sma artificial muscles with controllable mrf exoskeletons for robotic manipulators and grippers. *IEEE/ASME Transactions on Mechatronics*, *27*, 4585–4596.

[32] Yin, J., He, Y., Zhang, M., Zeng, P., Wang, T., Lu, S., & Wang, X. (2025). Promptlnet: Region-adaptive aesthetic enhancement via prompt guidance in low-light enhancement net. *arXiv Preprint arXiv: 2503.08276*.

[33] Zeng, T., Wang, T., Zhang, M., Yin, J., Zeng, Z., Zhang, F., Wang, Y., Jiao, J., Wang, Y., He, Y., et al. (2025). Tcstnet: A text-driven color style transfer network for low-light image enhancement. *Expert Systems with Applications*, 130012.

[34] Zhang, M., Fang, Z., Wang, T., Lu, S., Wang, X., & Shi, T. (2025). CCMA: A framework for cascading cooperative multi-agent in autonomous driving merging using large language models. *Expert Systems with Applications*, 127717.

[35] Zhang, M., Shen, Y., Li, Z., Pan, G., & Lu, S. (2024a). A retinex structure-based low-light enhancement model guided by spatial consistency. *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2154–2161.

[36] Zhang, M., Shen, Y., Li, Z., Pan, G., & Lu, S. (2024b). A retinex structure-based low-light enhancement model guided by spatial consistency. *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2154–2161. https://doi.org/10.1109/ICRA57147.2024.10610021

[37] Zhang, M., Shen, Y., Yin, J., Lu, S., & Wang, X. (2024). ADAGENT: Anomaly detection agent with multimodal large models in adverse environments. *IEEE Access*.

[38] Zhang, M., Yin, J., Zeng, P., Shen, Y., Lu, S., & Wang, X. (2025). TSCnet: A text-driven semantic-level controllable framework for customized low-light image enhancement. *Neurocomputing*, *625*, 129509.

[39] Zhang, Y.-F., Zhang, N., Hingorani, H., Ding, N., Wang, D., Yuan, C., Zhang, B., Gu, G., & Ge, Q. (2019). Fast-response, stiffness-tunable soft actuator by hybrid multimaterial 3D printing. *Advanced Functional Materials*, *29*(15), 1806698.

[40] Zhu, D., Chen, J., Shen, X., Li, X., & Elhoseiny, M. (2023). *MiniGPT-4: Enhancing vision-language understanding with advanced large language models*.

[41] Zhu, J., Chai, Z., Yong, H., Xu, Y., Guo, C., Ding, H., & Wu, Z. (2023). Bioinspired multimodal multipose hybrid fingers for wide-range force, compliant, and stable grasping. *Soft Robotics*, *10*(1), 30–39.

[42] Zhuang, S., Lei, D., Yu, X., Tong, M., Lin, W., Rodriguez-Andina, J. J., Shi, Y., & Gao, H. (2023). Microinjection in biomedical applications: An effortless autonomous omnidirectional microinjection system. *IEEE Industrial Electronics Magazine*.

7. Appendix

Prefix for Gripper Prompt Template

You are a visual detection assistant for grip tasks, the goal is to determine the optimal grasping position for the object. Any combination of the tools listed below may be used to complete this task. Follow the four steps outlined below for implementation.

Tool Description:

- Tool 1: Detect image brightness; Input: $Image_{raw}$; Output: Brightness condition; Tools: GPT-4o

- Tool 2: Light adjustment;Input: $Image_{raw}$; Output: $Image_{enhance}$; Tools: $Light_{enhance()}$

- Tool 3: Segmentation of grasped objects; Input: $Image_{enhance}$ & reason prompt; Output: $Image_{segment}$ & $Position_{grasped\_objects}$; Tools: Lisa()

- Tool 4: Estimation of grasp area; Input: $Image_{segment}$ & $Position_{grasped\_objects}$; Output: $Image_{grasp\_area}$ & $Heatmap_{grasp\_area}$; Tools: GPT-4o

- Tool 5: Stiffness Score of grasp area; Input: $Image_{grasp\_area}$ & $Position_{grasp\_area}$; Output: $Score_{Stiffness}$

Tool Usage Strategy:

- Only Detection (recommended to use Tool 1 & Tool 2)

- Detection and Grasping (all tools can be used)

- Only Repeat Grasping (Tool 3, Tool 4 & Tool 5)

Stiffness Levels:

- 0: Vulnerable items such as jelly, potato chips, persimmons, etc.

- 1: Easily deformed items such as fruits, plastic packaging, etc.

- 2: Tough and easily deformed items such as leather, towels, etc.

- 3: Hard plastic or wooden items such as tool box, glass jar, etc.

- 4: Heavy objects such as weights, dumbbells, and other metal objects.

Gripper Size: Capable of grasping objects from 1mm to 9mm.

Gripper Prompt Template

Step A: Detection, Thoroughly examine the overall content of the image to identify the object to be grasped and its features ($Property_{grasping_{o}bject}$).

A1: First, use Tool 1 to detect image brightness. If no lighting adjustment is needed, state <`Good lighting, no adjustment needed`>. If the image is too dark, use Tool 2 to enhance the brightness and provide the enhanced image as <$Image_{enhance}$>. **A2:** Infer the features of the grasping object from the $Image_{raw}$ or $Image_{enhance}$ image and output them in the following format:

{$Property_{grasping\_object}$: { Category: [ ], Shape: [ ], Size: [ ], "Environmental Information": [ ]}}

Step B: Grasping, Infer the potential grasping areas of the object.

B1: Segment the image of the object. Input $\text{Prompt}_{grasping\_object}$: $\text{Please segment } \langle \text{Prompt}_{grasping\_object}.\text{Category\&Size}\rangle$ into Tool 3 to obtain the segmented image as $Image_{segmentation}$. **B2:** Perform a detailed analysis of $Image_{segmentation}$ to identify the potential grasping areas of the object. Output the results in the following format:

```
{Section_{grasping_object}:
  {section1: [Material, Size, Stiffness Levels],
   section2: ...
  }
}
```

Step C: Critical Thinking, Determine the optimal grasping point based on the previous information.

C1: Using the information from $Property_{grasping\_object}$ and $Section_{grasping\_object}$, infer the prompt for the optimal grasping area $Prompt_{optimal\_grasping\_section}$. For example: $Prompt_{optimal\_grasping\_section}$: `"The grasping pose is the handle of the scissors, Handle: [Material: Plastic, Stiffness Levels: 1]."` **C2:** Use Tool 3 to input the prompt $Prompt_{optimal\_grasping\_section}$ and obtain the optimal grasping area image as $Image_{optimal\_grasping\_section}$. **C3:** From $Image_{optimal\_grasping\_section}$, select a specific point as the optimal grasping point, denoted as $Point_{optimal\_section}$

Step D: Reflect, Reevaluate the optimal grasping area and grasping point.

D1: If necessary, use the information from $Property_{grasping\_object}$ and $Section_{grasping\_object}$ to perform a RAG search, generating supplementary information $Info_{RAG}$.

D2: Repeat steps A, B, and C in conjunction with $Info_{RAG}$ until you are 100% certain that this is the best optimal gripper point.

Awtar, S. (2003). *Synthesis and analysis of parallel kinematic XY flexure mechanisms* [PhD thesis]. Massachusetts Institute of Technology.