Tactile-integrated FlexiRay: Breaking Planar Limits by Harnessing Large Deformations for Flexible, Full-Coverage, Human-like Multimodal Sensing

Author Information

Affiliations

Yanzhe Wang[†], Haotian Guo[†], Hao Wu[†], Huixu Dong*

Grasp Lab, School of Mechanical Engineering, Zhejiang University, Hangzhou, China

Contributions

Y.W., H.G. and H.W. conceived the idea. Y.W., H.G. and H.W. designed and fabricated the proposed FlexiRay hardware. Y.W. formulated the optimization algorithm for the camera and multi-mirror configuration. Y.W., H.G. and H.W. designed and conducted the experiments. Y.W., H.G. and H.W. collected and analysed the data. Y.W., H.G. and H.W. drafted and proofread the manuscript. Y.W., H.G. and H.W. contributed equally to this work[†]. The study was supervised by H.D..

Corresponding author

*Correspondence to: Huixu Dong (e-mail: huixudong@zju.edu.cn)

Abstract

Integrating tactile sensing into soft grippers holds great promise for safer robotic grasping and enhanced human-robot interactions. However, achieving multimodal, high-resolution sensing remains a significant challenge. While existing visual-tactile sensors offer unparalleled spatial resolution at an affordable cost, they rely on rigid structures to stabilize optical paths, hindering non-planar contact perception and violating the inherent adaptability of soft grippers. To relieve all relevant research gaps, we introduce FlexiRay, a novel soft gripper combining visual-tactile sensing with the bio-inspired Finray Effect, characterised with low cost, high compliance, dynamic sensory coverage. Combining a flexible substrate, adaptive illumination, and temperature-sensitive materials, FlexiRay replicates five core human tactile modalities of seven. A novel multi-mirror optical system, optimized for high coverage despite arbitrary dynamic deformations, enables consistent perception with just a single camera. Furthermore, employing a human-like multimodal deep learning framework to decouple contact forces, position, texture, temperature, and proprioception, FlexiRay achieves a force sensing accuracy of 0.14 N, a proprioception accuracy of 0.17 mm, and retains 90% effective coverage across dynamic interactions. Flexiray's structural compliance and multimodal sensing capabilities promote exceptional recognition of non-planar objects interactions and autonomous human-robot interaction, showcasing significant potential for safer and more intelligent service robotics.

Introduction

The human tactile system possesses exceptionally complex perceptual mechanisms, consisting of three principal sensory systems: the cutaneous, kinesthetic, and haptic systems[1]. These systems enable the human hand to perceive seven key modalities, namely pressure, contact localization, texture, temperature, vibration, proprioception, and pain[2]. Together, they allow humans to perform various intricate and precise tasks[3]. Translating this exceptional sensory ability to robots brings about significant benefits, yet significant challenges[1,4]. Over the past three decades, researchers have explored nearly all forms of sensing, such as resistive[5,6], magnetic[7], pressure-sensitive[8,9], capacitive[6], waveguide-based[10,11], acoustic[12], and thermal sensing[13] et al. A special focus has been placed on the development of tactile sensing in soft robotic grippers[10,12], aiming to enable more dexterous and safer environmental interactions via the combination of structural adaptivity and tactile perception. Despite substantial advancements, achieving large-area, high-resolution, and multimodal tactile sensing remains a tremendous challenge. This can be primarily ascribed to two factors: the high production cost of taxel-based measurement methods[5] and the limitations in spatial resolution of data-driven computational sensors[11–14]. Therefore, existing designs oftern struggle in the dilemma among resolution, coverage, cost-effectiveness, and multimodal sensing capablities. A comprehensive, human-like tactile perception solution remains an elusive goal.

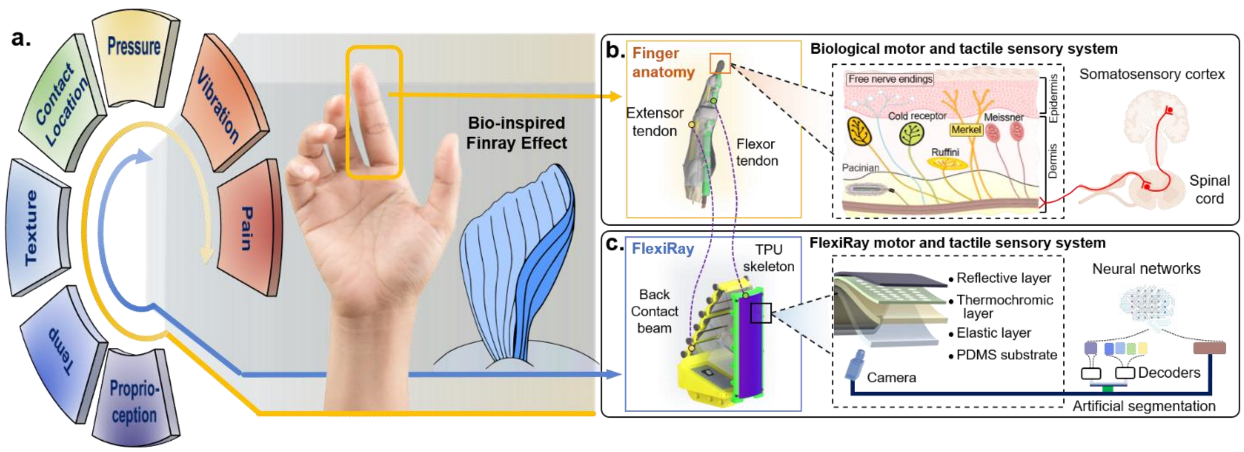
Visual-tactile sensors (VTS) have emerged as a promising solution to address these challenges by leveraging metal-oxide-semiconductor (CMOS) optical arrays to convert multimodal tactile information into high-resolution, pixel-level images[15–19], thereby enabling insights into interactions such as pressure[16,20] and texture[21,22]. Most existing VTS systems require stable optical paths to avoid occlusion and perception disturbances, which results in bulky designs and a heavy reliance on rigid structures[21,23–28]. This rigidity leads to a fixed sensing coverage, presenting significant challenges for integrating VTS with flexible grippers, as it conflicts with the compliance and flexibility inherent to soft robotic systems. Although some studies have achieved integrated robotic finger designs through optical path optimization[28,29] or camera arrangement adjustments[24,26,29], these grippers lack the structural compliance necessary for safe interaction. In particular, their reliance on planar contact information significantly limits their perceptual capabilities on non-planar surfaces, reducing their versatility and adaptability in complex environments.

Bio-inspired Finray Effect (FRE) soft grippers provide an elegant solution for grasping objects of various shapes through adaptive enveloping, leveraging passive structural deformation to ensure safe interactions[30]. Integrating VTS with compliant FRE grippers partially mitigates the inherent compliance limitations of existing VTS[31,32]. However, current VTS systems prove inadequate for meeting the demands of broader and safer soft robotic interactions, particularly under more significant structural deformations. Some studies focus on enhancing the perceptual coverage and providing moderate flexibility, such as segmenting multiple cameras to adapt to structural deformations[4], using a single rigid curved mirror[29], or replacing rigid acrylic with thin, flexible mylar at the front contact surface to enhance surface deformation[31,32]. However, to avoid visual occlusion and maintain the optical path under large deformations, the back structures of these designs remain somewhat rigid, which undermines the key advantage of FRE's natural compliance and fails to address the inherent limitations of current VTS designs. The challenge persists in systematically integrating VTS with flexible robotic grippers to ensure high-resolution, large-area multimodal tactile perception during compliant interactions.

In this paper, we report FlexiRay, a tactile-integrated soft robotic gripper inspired by human multimodal touch, capable of simultaneously perceiving contact pressure, localization, texture, temperature, and proprioception (Fig. 1). By strategically integrating compliant FRE grippers with VTS, FlexiRay achieves a sizeable sensory area of 1560 mm^2, maintaining an average effective coverage of 87.2% under arbitrary dynamic deformations. It also achieves an overall force accuracy of 0.14 N and a spatial proprioception accuracy of 0.17 mm. FlexiRay demonstrates unparalleled structural compliance, with a deformation capacity over 400% greater than existing compliant VTS of the same type [22] under the identical load. Table 1 provides a detailed comparison of FlexiRay with representative state-of-the-art VTS. We emphasize the relevant differences and recommend that readers refer to and examine them more thoroughly.

The following contributions drive these advancements: First, we propose a novel integration design of VTS and the FRE soft gripper, realizing a compact, human-inspired tactile perception modality with a multi-layered structure that balances compliance and sensory performance. Second, a new flexible VTS substrate architecture is developed, combining integrated manufacturing processes with the FRE base structure, a polydimethylsiloxane(PDMS)-based contact substrate, and a flexible silicone tactile material. Temperature-sensitive materials are also incorporated to further enhance multimodal sensing capabilities. Third, a layout optimization method for the inside optical sensing system, based on Covariance Matrix Adaptation Evolution Strategy (CMA-ES), is proposed. This method optimizes the layout of the single camera and multiple mirrors via leveraging structural deformation collected during physical interactions, aiming to maximize dynamic perceptual coverage during deformation. Notably, it cleverly transforms optical interference from a limiting factor into a functional design element, with discrete mirrors ensuring stable and continuous sensing during dynamic deformation without compromising flexibility. Finally, FlexiRay showcases its excellent ability to classify complex, non-flat objects, such as textured perception balls, through human-inspired multimodal sensing and deep learning models. In a handover task, it adeptly distinguishes cups of varying temperatures, secures a stable grasp with minimal force that prevents slippery or crushing, and intuitively releases objects upon detecting human interaction. These findings highlight the potential of the proposed FlexiRay, paving the way for intelligent robotic systems in dynamic, real-world applications such as housekeeping.

Fig. 1: FlexiRay: A flexible Finray Effect gripper enables human-like multi-modal tactile and proprioception perception.



a FlexiRay incorporates five of the seven sensory modalities found in the human hand. Leveraging the bio-inspired Finray Effect design, it features exceptional structural adaptability. **b** The kinesthetic system in the human hand relays real-time hand postures and motion information via muscles, tendons, and joints, supporting precise tasks like delicate grasping. The cutaneous system, with widely distributed receptors such as Pacinian corpuscles (vibration), Merkel's discs (texture), and Ruffini endings (pressure), provides essential information about object properties and interaction dynamics (e.g., slippage). The haptic system integrates spatial and temporal data from both sensory streams to interpret complex contact characteristics, enabling adaptive and dexterous actions. **c** FlexiRay innovatively integrates bio-inspired skin with a tendon-like skeleton, seamlessly integrating proprioception with multimodal tactile sensing.

Table 1 Comparison between the state-of-the-art VTS systems and FlexiRay.

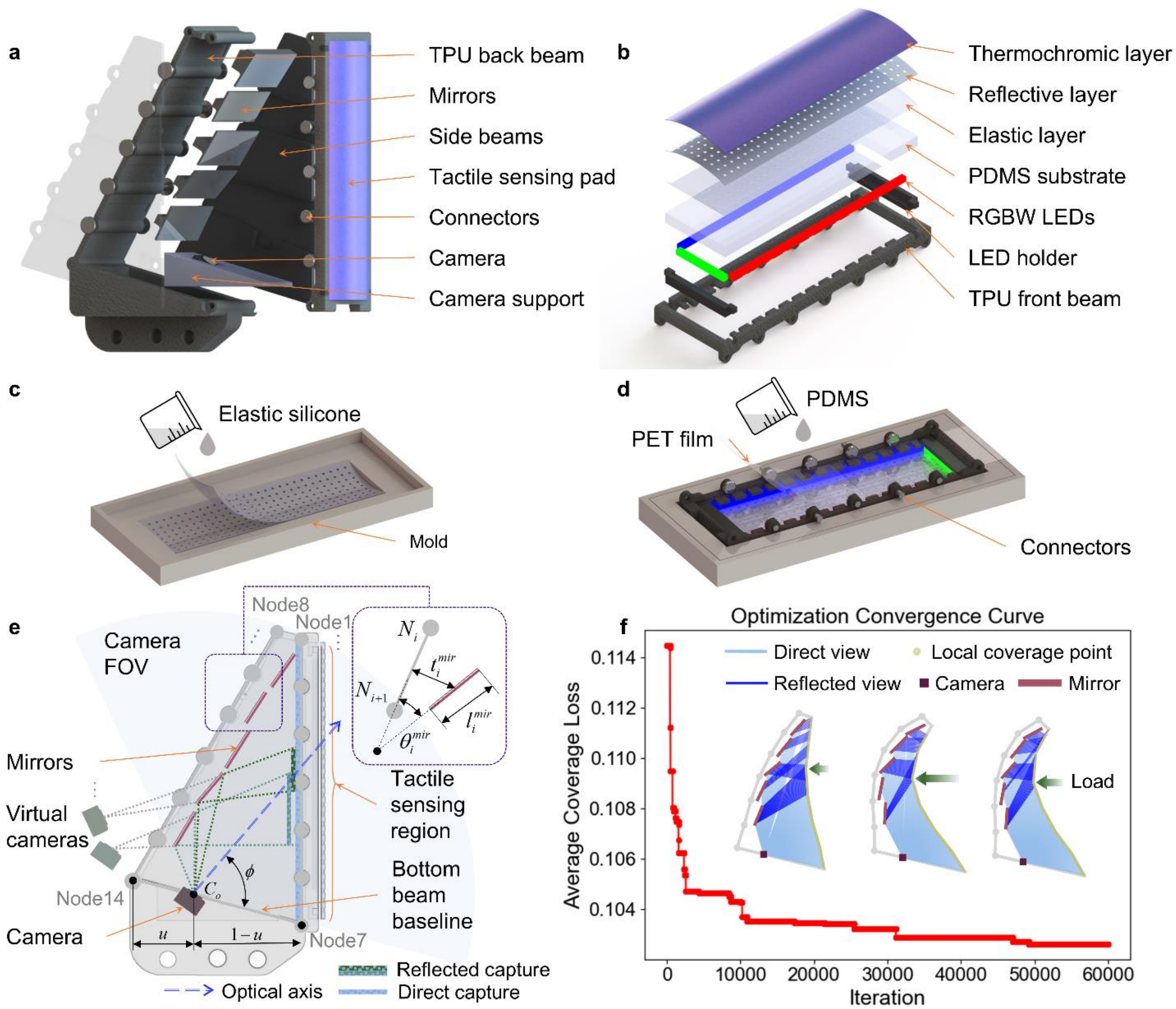
<table>
<tr><th>Sensor</th><th>Principle</th><th>System Rigidity</th><th>Area (mm^2)</th><th>Model</th><th>Modality</th><th>Force Error (N)</th><th>Proprioception Error (mm)</th><th>Non-planar Adaptability</th></tr>
<tr><td>GelSight[21]</td><td>Camera</td><td>Rigid</td><td>250</td><td>CNN</td><td>Force, Texture</td><td>F_N: ~0.67</td><td>—</td><td rowspan="7">Limited contact area</td></tr>
<tr><td>GelSight360[26]</td><td>Camera</td><td>Rigid</td><td>—</td><td>MLP</td><td>Texture</td><td>—</td><td>—</td></tr>
<tr><td>Digit[27]</td><td>Camera</td><td>Rigid</td><td>304</td><td>ResNet</td><td>Texture</td><td>—</td><td>—</td></tr>
<tr><td>Gelslim[28]</td><td>Camera</td><td>Rigid</td><td>1200</td><td>iFEM</td><td>Force, Texture</td><td>F_N: ~0.32</td><td>—</td></tr>
<tr><td>GelSight Wedge[33]</td><td>Camera + Single-mirror</td><td>Rigid</td><td>768</td><td>MLP</td><td>Texture</td><td>—</td><td>—</td></tr>
<tr><td>Insight[34]</td><td>Camera</td><td>Semi-compliant (Hollow skeleton)</td><td>4,800</td><td>ResNet</td><td>Force, Texture</td><td>0.03</td><td>—</td></tr>
<tr><td>GelSight Svelte[29]</td><td>Camera + Single-mirror</td><td>Semi-compliant (Soft front beam)</td><td>~1895</td><td>CNN</td><td>Bend/ Twist, Texture</td><td>T_B: ~9.4 Nmm/T_T: ~7.6 Nmm</td><td>—</td></tr>
<tr><td>GelSight Baby Fin Ray[22]</td><td>Camera + Single-mirror</td><td>Compliant (FRE with rigid connections)</td><td>990</td><td>ResNet</td><td>Texture</td><td>—</td><td>—</td><td>Sacrifice structural compliance to maintain stable optical paths</td></tr>
<tr><td>GelFlex[4]</td><td>Multi-Cameras</td><td>Highly flexible (Serial Joints)</td><td>—</td><td>LeNet</td><td>Proprioception, Texture</td><td>—</td><td>~0.77</td><td>For each phalange, limited contact area</td></tr>
<tr><td>Liu et al.[35]</td><td>Camera</td><td>Highly flexible (Compliant Spatial Truss)</td><td>—</td><td>MLP</td><td>Force, Proprioception</td><td>F_N: ~0.25</td><td>~1.18</td><td>Only perceive contact at the beams</td></tr>
<tr><td>FlexiRay (Ours)</td><td>Camera + Multi-mirrors</td><td>Highly flexible (FRE with flexible connections)</td><td>1560</td><td>PP-LiteSeg+ ResNet</td><td>Force, Location, Proprioception, Texture, Temperature</td><td>F_N: ~0.14</td><td>~0.17</td><td>Average coverage of 87.2% under different deformations</td></tr>
</table>

Results

Working principles of FlexiRay

The design of proposed Tactile-integrated FlexiRay is illustrated in Fig. 2a. Inspired by the hierarchical structure of the human hand, this design integrates both the perceptual and structural elements to address the gap in visual-tactile sensing and soft gripper integration. The system is primarily composed of a compliant finger framework, an optical system, and a tactile sensing pad. The compliant finger framework includes a back beam, a front beam, and side beams. The back and front beams, made of TPU material, provide elasticity similar to the tendons in the human hand, maintaining structural stability while transmitting forces and deformations. The rigid side beams act as the bones of the hand, being hinged to the ends of the back and front beams to provide variable joints. The optical system includes a camera for image capture, multi-segment reflective mirrors for enhancing the camera's field of view (FOV), and flexible LED light strips for illumination. The tactile sensing pad consists of a PDMS-based substrate, an elastic silicone layer, a reflective layer, and a temperature-sensing layer, as shown in Fig. 2b. The PDMS substrate provides flexible support without compromising compliance (Fig. 2d), while the low-hardness transparent silicone elastic layer is used to enable texture mapping (Fig. 2c). The external silicone reflective layer, infused with aluminum, enhances the capability to map of contact textures. The outermost layer is a silicone temperature-sensing layer, incorporating thermochromic materials, which can be captured by the camera through marker holes in the reflective layer. Thus, FlexiRay exhibits sensitivity to external physical contacts and temperature stimuli, similar to the mechanoreceptors and thermoreceptors in human skin.

Fig. 2: Illustration of the integrated design and layout optimization of FlexiRay.



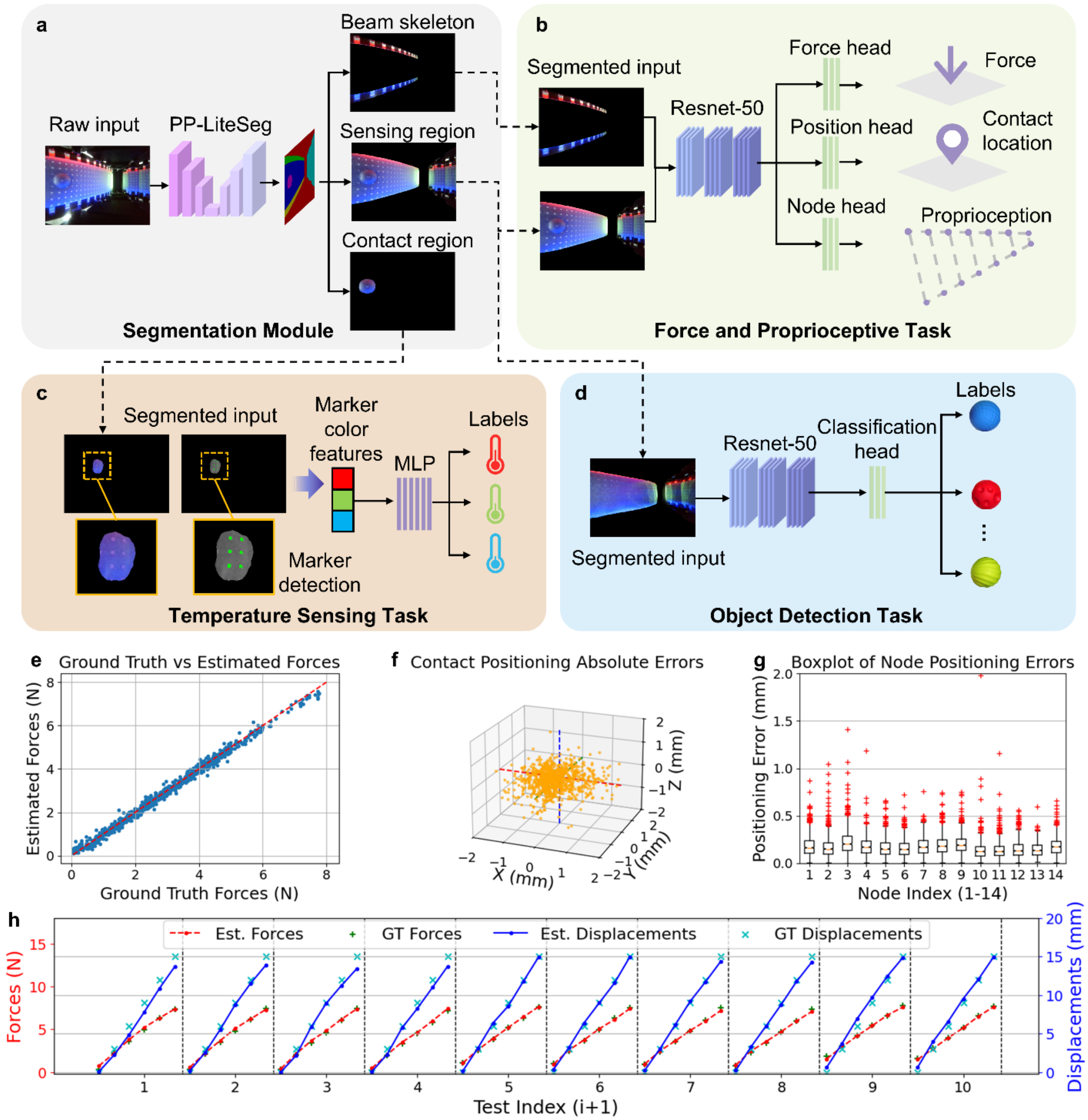
a Exploded view of FlexiRay. **b** Exploded view of the tactile sensing pad. **c** Casting of the elastic layer. **d** Casting of the PDMS substrate. **e** Schematic diagram of the inside optical sensing system layout optimization method. **f** The optimization convergence curve of average coverage loss.

The key challenge in compatibility between VTS and flexible structures lies in the significant deformation of the soft robotic hand during the interaction, which restricts the camera's FOV. Rather than limiting the flexibility of the hand, we address this issue through systematic optimization, incorporating a multi-mirror configuration to transform the unwanted deformation into a functionaln advantage. As shown in Fig. 2e, each mirror is independently attached to the back beam, passively altering the direction of the camera light path. This allows the discrete capture of perception regions that the camera's FOV does not cover. Combining the views from multiple mirrors achieves continuous coverage of a large perception area under dynamic loading conditions. To ensure optimal compliance adaptation, the optical system layout is modeled as a 2D geometric parameter optimization problem in the lateral cross-section. The optimization objective is maximizing the coverage of the tactile sensing regions captured by direct camera views and mirror reflections. The optimization parameters include the camera's position along the bottom beam baseline, its shooting angle, the length of each mirror, as well as the distance and angle of each mirror relative to the back beam. CMA-ES is used as the solving tool. Deformation data for the back and front beam joint nodes under different loads are collected to serve as the prior information. This optimization process essentially aims to find the layout parameters that maximize the camera's FOV coverage across various deformations, leveraging the passive FOV reconstruction capability of multiple mirrors to enhance the camera's perception of blind regions in deformed views. The convergence curve of the coverage loss during the optimization process is shown in Fig. 2f.

To achieve multi-modal perception in flexible 3D space, we developed a series of deep learning models that decouple perception tasks and allocate them to specialized models. This approach effectively addresses the challenging spatial deformation coupling between the flexible base and the contact area during interactions. Additionally, combining these sub-models enables the solution of more complex real-world tasks. First, we developed an image region extraction model based on the PP-LiteSeg model (Fig. 3a). This lightweight semantic segmentation model helps distinguish contact information into cutaneous and kinesthetic systems, segmenting the front beam skeleton, the front beam interaction region (direct perception), the mirror region (reflective perception), and the local contact region, which is a preprocessing step to enhance the quality of subsequent tasks. Next, a proprioception model based on the ResNet50 backbone is implemented (Fig. 3b), with the segmented images of the front beam skeleton and the sensing region as input. This model estimates the normal contact force, the 3D position of the contact point relative to its reference frame, and the side beam node positions, using dedicated heads for force, position, and node estimation. Furthermore, a texture recognition model is constructed using the ResNet50 backbone (Fig. 3d), which takes the sensing region image as input and classifies the contact texture through a classification head. Additionally, a temperature sensing model (Fig. 3c) is developed, which utilizes the color features in the markers of the local contact region image to provide contact temperature information.

Figure 3e illustrates the estimation of normal contact force for 1000 different contact positions and loading conditions. The x-axis represents the ground truth forces, while the y-axis represents the estimated forces. The red dashed line indicates the ideal perfect match line. The root mean square error (RMSE) of the predicted forces is 0.135 N, with a correlation coefficient of 0.997. These results indicate that the model's estimated forces are highly consistent with the actual measured values, demonstrating high accuracy in force perception. Figure 3f shows the distribution of absolute errors for the estimated contact positions in 3D space under the same 1000 loading conditions. Scatter points closer to the origin represent more minor errors. Statistical results reveal an average error of 0.81 mm and a standard deviation of 0.38 mm, validating the model's good accuracy and stability in estimating contact positions. For the node positions that characterize the hand configuration, a box plot of the positioning errors for the 14 nodes of the FlexiRay under the same 1000 loading conditions is presented in Fig. 3g. The average positioning error for all nodes is approximately 0.17 mm, with an average standard deviation of 0.10 mm, indicating the model's good positioning accuracy and robustness. To evaluate the proprioception accuracy during continuous interaction, tests are conducted in which the contact depth is gradually increased from a randomly initialized position to the target load. Figure 3h demonstrates a comparison between the actual measurement data and estimated results for normal contact force and depth across ten repeated random trials. The results reflect that the model maintains high accuracy and stability in proprioception under dynamic continuous prediction.

191 **Fig. 3: Learning-based multi-modal perception pipeline and proprioception accuracy analysis.**



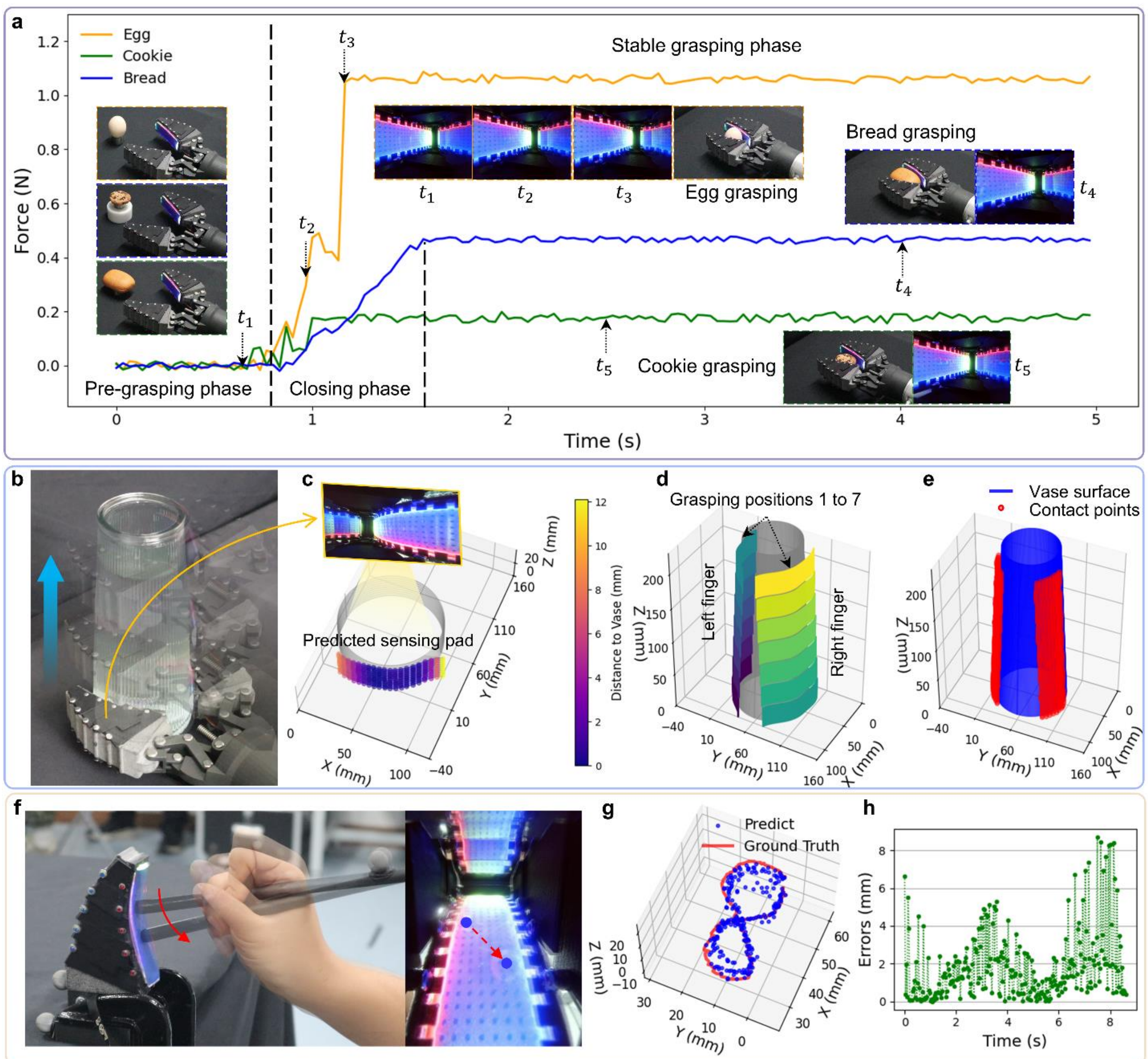
192

193 **a** Semantic segmentation model is employed to segment the front beam skeleton, perception region, and
194 contact region. **b** Sub-model for sensing normal contact force, position, and proprioception deformation. **c**
195 Color mapping model for temperature sensing. **d** Texture recognition model for tactile-based object
196 classification. **e** Accuracy analysis of normal contact force estimation under 1000 varying loading trials. **f**

Absolute error distribution of contact position prediction under 1000 varying loading trials. **g** Box plot of positioning errors for joint nodes under 1000 varying loading trials. **h** Continuous estimation of normal contact force and depth under 10 random tests. "Est." refers to the estimated data, and "GT" refers to the ground truth data.

To further validatethe multi-modal perception performance of Flexiray in practical applications, three experiments are conducted, respectively. Firstly, we demonstrate the proposed FlexiRay is able to perform compliant force-close-loop grasping on objects, especially soft and fragile objects, such as eggs, cookies, and bread. Figure 4a presents the estimated gripping force collected during the grasping processes. In particular, the force thresholds are roughly set according to their weights, which are approximately 1 N for the egg (44g), 0.2 N for the cookie (~10g), and 0.48 N for the bread (20g). The results indicate that Flexiray can sense subtle contact forces provide stable force estimization regardless of structural deformation facilitating more precise and safer interactions. Secondly, improved compliance, large coverage and precise proprioception of FlexiRay enable it to efficiently and accurate reconstruct the surface shape of a grasped object with fewer attempts, exemplified with a cone-shaped vase. Specifically, the Flexiray gripper performs adaptive gripping at equal height intervals from the bottom to the top of the vase, capturing tactile images during stable gripping (Fig. 4b). Using the proprioception model, the positions of the nodes on both sides of Flexiray are estimated and mapped into three-dimensional space, based on which continuous curves of the beam skeleton are generated through spline interpolation. Further surface interpolation enables the reconstruction of the sensing pad (Fig. 4c). The contact regions extracted from the reconstructed surface reveal local geometric features of the grasped object. A total of seven different gripping positions are recorded during this experiment, as shown in Fig. 4d. By integrating local shape information from various sensing areas, Flexiray demonstrates comprehensive capabilities for object shape analysis, as depicted in Fig. 4e. We designed the last validation experiment of FlexiRay's excellent contact localization capability using the pen-nib's position collected by visual motion capture system (VICON) as ground truth measurement. As shown in Fig. 4f, an operator holding the pen-nib draws a trajectory of “8” on the FlexiRay’s soft sensing pad and the contact localization model provide estimated contact information throughout dynamic interactions. Figure 4g compares the predicted and actual trajectories, while Figure 4h displays the prediction errors, with an average localization error of 1.85 mm. In summary, these results highlight Flexiray's precise multi-modal perception and sensitivity thgoughout dynamic interactions.

Fig. 4: Experiments on gripping force estimation, shape reconstruction, and contact trajectory tracking utilizing the proposed framework.



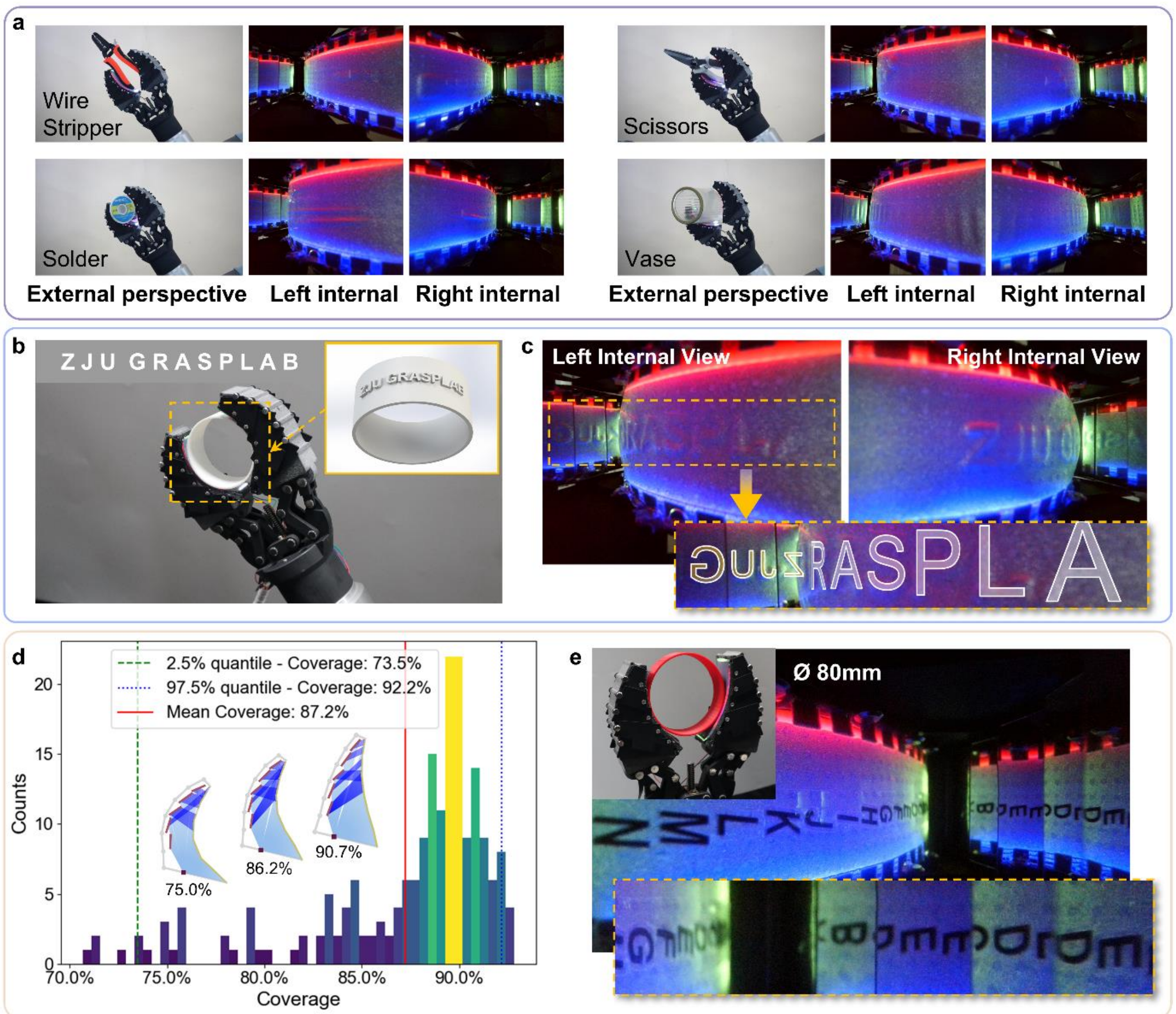
a Dynamic force estimation for enveloping grasps on eggs, cookies, and bread. **b** Seven gripping and releasing actions on a vase at varying heights, yielding tactile images from both fingers. **c** Estimation of the sensing pad surface through interpolation of node positions from both sides of Flexiray. The color mapping reflects the distance from the vase surface. **d** Reconstruction results of the sensing pad at seven different gripping perception positions. **e** Estimated contact points from the reconstructed sensing pad, corresponding to the local shape of the vase. **f** Handwritten digit "8" on the sensing pad, predicting the trajectory of contact points. **g** Comparison of predicted and actual trajectories, with actual data collected from a visual motion capture system that tracks the stylus. **h** Localization error for each contact point.

Texture detection performance and perceptual coverage

To assess the texture detection performance of FlexiRay, a series of large curved or wide-area gripped objects are selected for testing. These objects include tools such as screwdrivers, hot air guns, wire strippers, and scissors. Additionally, curved items of various sizes and shapes, such as solder, bottle caps, vases, pen holders, clips, and mice, are also tested. A two-finger gripper composed of FlexiRay is mounted on a UR5e robotic arm to perform natural grasping experiments and capture texture images without external interference. Several typical demonstrations are shown in Fig. 5a, with the complete set provided in the Supplementary Materials. The results demonstrate that FlexiRay not only conforms seamlessly to and wraps around large curved objects but also accurately captures the surface contours and geometric details during flexible deformations. This showcases the extensive tactile perception capability of FlexiRay, which is not available in current VTS technologies.

The practical effect of the multi-mirror configuration on FOV extension is evaluated using the 3D-printed text ring shown in Fig. 5b. The ring has an outer diameter of 68 mm, a font height of 1.5 mm, and a line width of 1 mm. As shown in the internal view in Fig. 5c, under large deformations, the front beam obstructs the camera's capture of the fingertip region. However, by incorporating mirror-reflective areas, FlexiRay is able to capture the occluded textures, thus providing a more complete perceptual field. Specifically, the direct perception area captures "RASPLA," while the reflective perception area captures "ZJUG," forming a continuous texture pattern "ZJUGRASPLA."

Fig. 5: Texture detection performance and perceptual coverage of FlexiRay.



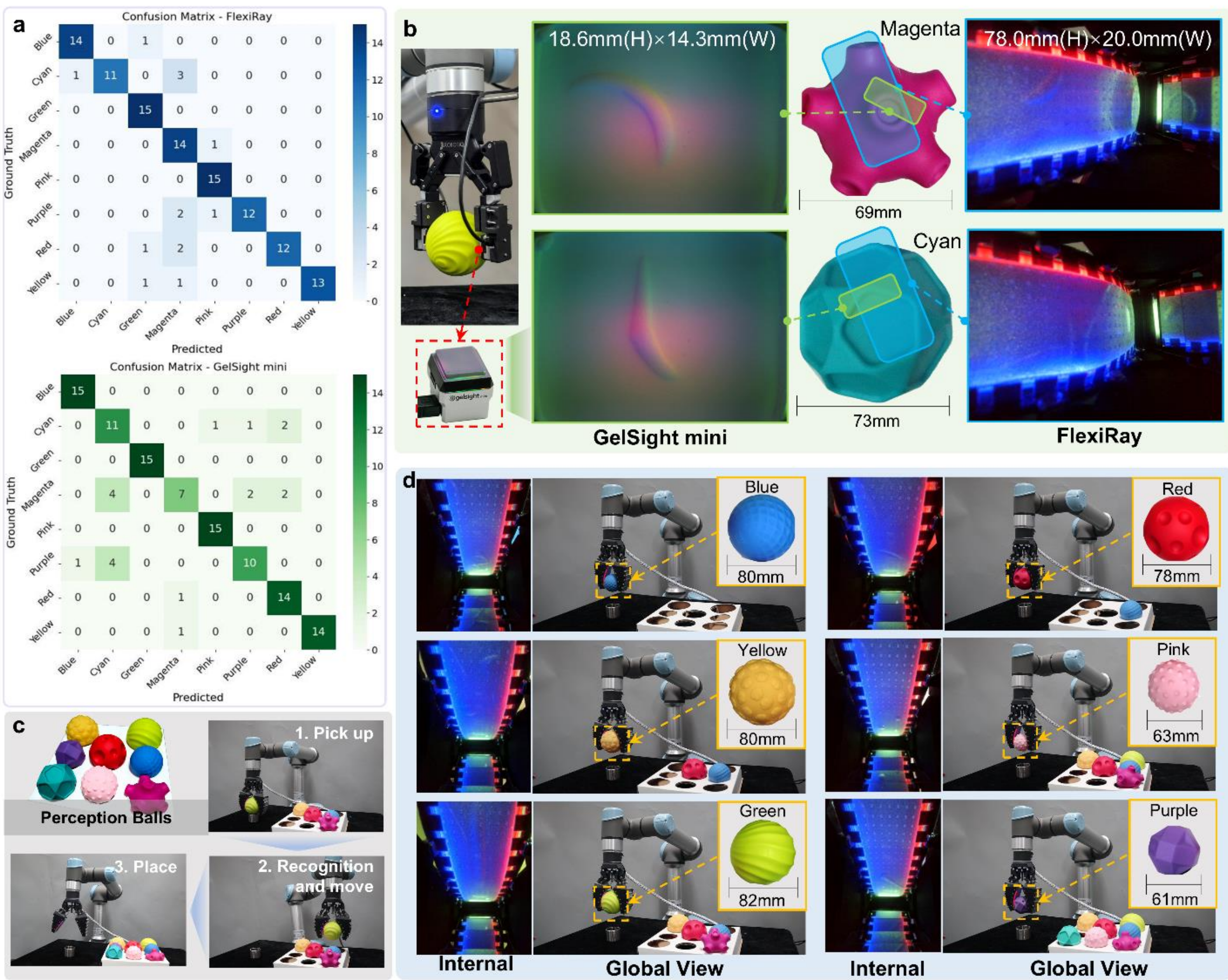
a Raw internal images captured while gripping various objects. **b** 3D-printed text ring. **c** Internal raw image of the text ring grip, illustrating texture details captured by both direct and reflective sensing. **d** Distribution of perceptual coverage across 200 random deformation tests under varying loads. **e** Continuous alphabetic markings indicate the visible areas during the gripping of rings with different diameters, showing a demonstration of the 80mm diameter ring. (Further demonstrations are provided in the Supplementary Materials.)

The perceptual coverage under different deformations is further quantified through both simulation and physical experiments. The simulation is conducted based on the collected hand configuration data at 200 random contact positions and loads. The camera's FOV is uniformly discretized into 300 rays in a 2D plane, and the coverage of each ray on the contact area is calculated, including both direct incidence and mirror-reflected coverage. The contact area is discretized into 100 target points. Considering the varying propagation distance l^c of each ray, the coverage radius of a single ray is defined as $R = l^c\sqrt{2(1-\cos\Delta\theta)}$, where $\Delta\theta$ is the angle between two adjacent rays. The perceptual coverage is assessed by calculating the proportion of target points covered by each ray. The experimental results, shown in Fig. 5d, indicate that the average coverage of the contact area in 200 random tests is 87.2%, with most of the coverage concentrated around 90%. Even under extreme deformations, the perceptual coverage remained above 70.0%. Additionally, for ring-shaped grasps with diameters of 50, 60, 70, and 80 mm, continuous alphabetic markers on the perceptual area visually demonstrate the variation in coverage under different deformations. Figure 5e shows the results for the 80mm diameter, with other results provided in the Supplementary Materials. These results confirm the effectiveness of the designed direct and reflective sensing strategy, along with the optimization method.

Texture-based ball classification

To demonstrate the texture recognition performance of FlexiRay, eight perception balls with different surface textures are selected as classification targets. For each category, 80 randomly collected tactile images of stable grasps are gathered, resulting in a total of 640 samples. The dataset is then divided into training and validation sets in a 4:1 ratio. To highlight the advantages of FlexiRay's large-area flexible sensing capability, the commercial GelSight mini is used as a benchmark for comparison. The GelSight mini also collects 640 samples, and the same training process and model parameters are applied. The training results showed that FlexiRay's classification model achieved an accuracy of 95.83% on the validation set, outperforming GelSight mini, which achieved an accuracy of 94.17%.

Fig. 6: Tactile-based ball classification experiment process and results.



a Confusion matrices for the test results of GelSight mini and FlexiRay. **b** The comparison of perception modes and coverage states between GelSight mini and FlexiRay. **c** Workflow for ball sorting and tactile perception classification. **d** Raw tactile images of various ball types.

To evaluate the accuracy of the trained models, both grippers are randomly tasked with grasping each ball 15 times, resulting in two test sets of 120 samples. The recognition results are presented in the confusion matrix shown in Fig. 6a. FlexiRay achieves an average success rate of 88.3%, with a perfect recognition accuracy of 100% for the green and pink balls. However, the accuracy for the cyan ball is the lowest at 73.3%. Some cyan ball samples are misclassified as the magenta ball, likely due to the similarity in the tactile image features at the edges of both balls. In comparison, GelSight mini achieves an average success rate of 84.2%, 4.1% lower than FlexiRay's performance. Its accuracy for the magenta ball is even lower, reaching only 46.7%, which is half the success rate of FlexiRay. Figure 6b clearly illustrates the tactile sensing differences between the two sensors during grasping. FlexiRay's sensing area (1560 mm^2) is 5.9 times larger than that of GelSight mini (265.98 mm^2), and it is also capable of capturing non-planar contact features. This comparison highlights the advantages of FlexiRay, where its larger sensing area and superior compliance allow it to capture more detailed tactile information in a single grasp, contributing to improved recognition accuracy.

A robot ball sorting task is performed using the trained classification model, as shown in Fig. 6c. During the experiment, balls of random types are placed at the grasping position. The robot relies solely on tactile modality to perceive the surface textures of the balls and sort them into the corresponding positions based on the recognition results. Figure 6d presents additional tactile images captured during the sorting process for each ball. The gripper's performance in this classification task reflects the overall compliance and stability of the FlexiRay, enabling it to adapt to the shape and texture of various complex surfaces. Moreover, the robot is able to accurately classify the balls based exclusively on tactile perception with FlexiRay, highlighting the significance of tactile modality in precise grasping and manipulation.

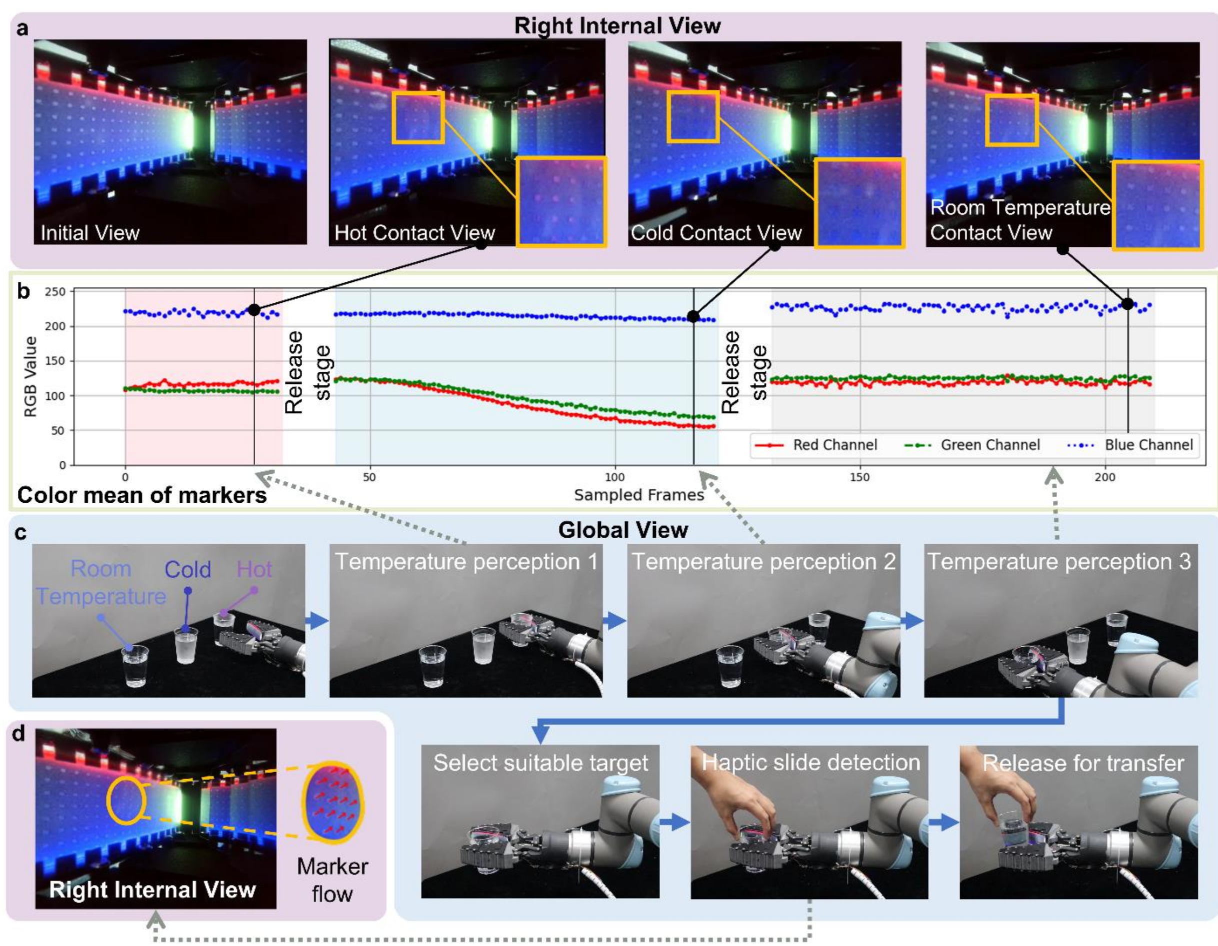
Temperature-aware safe human-robot cup transferring

Humans possess rich sensory capabilities that enable stable and safe interactions. Take cup transfer as an example, using thermoreceptors in the skin, humans can detect thermal stimuli and differentiate between various temperatures. Concurrently, their tactile perception allows them to apply appropriate gripping force to prevent slipping while recognizing external contact states to ensure a successful transfer.Drawing inspiration from these human capabilities, we demonstrate the FlexiRay's ability to replicate similar sensory functions. In this scenario, the robot assists in selecting a cup at the desired temperature and facilitates its handover to a human. The experiment focuses on assessing FlexiRay's temperature perception and tactile feedback mechanisms, investigating how temperature sensing identifies target objects and how tactile feedback controls release timing to enable efficient and safe interactions. The experimental procedure is outlined as follows:

- **Tactile image acquisition of water cups:** At the start of the experiment, three cups of water are placed on a table: hot water (80°C), cold water (4°C), and room temperature water (24°C). The cups are unmarked, and their temperatures cannot be visually distinguished without physical contact. The robot grasps each of the three cups while capturing sensing images from FlexiRay.
- **Temperature perception and recognition:** The contact area of each cup is segmented, and the marker pixels within this region are extracted to compute the average RGB values. The temperature recognition model employs a dual fully connected layer architecture with an input dimension of 256×256, producing output labels to classify the water as hot, cold, or at an optimal temperature. The perception process is maintained for a sufficient duration to allow heat transfer to stabilize, ensuring a reliable temperature modality image.
- **Cup transfer and release triggering:** During the cup transfer process, when a human hand contacts the cup and attempts to take it from the robot's grasp, a sliding signal is generated in the tactile image. FlexiRay detects the slip amount in real-time at the contact area. Once predefined conditions are met, the release mechanism is activated, automatically loosening the gripper to allow the human to safely take the cup.

The tactile image frames and corresponding RGB color changes throughout the experiment are shown in Fig. 7a and Fig. 7b. The results indicate that the robot accurately recognizes the temperature of the water cups in a dynamic environment and successfully delivers the appropriate cup based on the human's preference. During the interaction, FlexiRay reliably detects the sliding signals triggered by human contact and releases the cup at the optimal moment (Fig. 7c and Fig. 7d). In repeated trials, the robot successfully completed the cup transferring and released actions without any incidents of dropping the cup or misjudging the temperature, demonstrating the system's robustness and reliability. This experiment reveals the integrated application of the flexible, large-area VTS developed in this study, highlighting its potential for multimodal perception. It illustrates the multidimensional intelligence of robots in perception, action, and human-robot interaction, with promising applications in more complex tasks such as elderly caregiving and handling objects in household environments.

Fig. 7: Tactile temperature sensing and sliding detection for human-robot cup interaction.



a Raw tactile images. **b** Changes in the RGB values of tactile perception area markers under different temperature contacts. **c** Robotic cup grasping and temperature sensing process. **d** Sliding signal detection.

Discussion

This study presents a novel VTS-integrated flexible robotic gripper, Tactile-integrated FlexiRay, inspired by human tactile modalities and the FinRay effect. The gripper effectively combines enhanced structural compliance with advanced multimodal sensing capabilities, successfully realizing five out of the seven primary human tactile modalities, including contact force, location, texture, temperature, and proprioception, excluding pain and vibration. Specifically, the normal force estimation accuracy reaches 0.14 N, and the spatial proprioception accuracy is 0.17 mm. Through strategic structural design and optimization, FlexiRay demonstrates strong resistance to visual interference even under large deformations. The average effective sensing coverage of the tactile sensing pad across different deformation states is 87.2%, with over 70% maintained during large deformations. To address the visual occlusion gaps that occur during large deformations, we do not limit the compliance of the finger to maintain the optical path. Instead, we treat deformation-induced optical path disturbances as key design parameters. The structure innovatively designs a segmented mirror array and proposes a method for optimizing the internal optical system layout. We characterize the force-deformation patterns of the FRE fingers under different loading conditions in a physical environment. These patterns serve as the basis for determining the optimal position and orientation of the mirrors. This method ensures consistent image acquisition through passively controlled mirror reflections, even under substantial deformation. Therefore, FlexiRay excels at grasping irregular, cylindrical, or spherical objects, significantly increasing the compliant contact area and enhancing the richness and efficiency of the perception data during flexible interactions.

Compared to state-of-the-art VTS technologies, the structure most similar to ours is the GelSight Baby Fin Ray. However, FlexiRay demonstrates a more pronounced compliance advantage. Based on reported data, at a contact force of 7.5 N, the deformation of the GelSight Baby Fin Ray is approximately 3~4 mm. In contrast, FlexiRay exhibits a deformation of about 15 mm in the contact depth direction under the same load, which is more than four times greater than that of the Baby Fin Ray. Additionally, FlexiRay is not constrained by visual occlusion caused by large deformations. The average effective sensing area of the tactile perception pad across different deformation states is approximately 1360 mm², and the effective sensing area under large deformation remains above 1092 mm², an advantage not found in current VTS systems. The combination of high compliance and a large sensing area enables FlexiRay to offer unparalleled flexible visual-tactile perception. It can provide stable conformal grasping while capturing more detailed tactile information from a single interaction. Deriving benefits from this, FlexiRay achieves a success rate of 88.3% in a texture recognition task involving perception balls, surpassing the 84.2% success rate of the GelSight mini. Moreover, in the challenging task of recognizing the magenta ball, FlexiRay's success rate improves by a factor of two. Additionally, with its integrated multimodal perception and deep learning models, FlexiRay not only demonstrates proprioceptive capabilities akin to the human tactile system but also achieves temperature sensing and slip detection, enabling autonomous and safe human-robot interaction in tasks like cup handovers. FlexiRay bestows exceptional sensory capabilities and compliant execution abilities to robots, showing great potential for applications in human daily life assistance. In terms of cost and usability, the affordability of the imaging components and the well-designed manufacturing process give FlexiRay a clear competitive advantage in terms of both manufacturing cost and processing difficulty. Unlike technologies like GelSight, which require strong adhesion between coatings, FlexiRay can be easily fabricated by adding components and casting materials onto a 3D-printed TPU beam skeleton, enabling integrated molding. The modular design of the front, side, and side beams allows for easy replacement of the tactile sensing pad, facilitating maintenance and extending the device's lifespan.

FlexiRay represents a significant advancement in the development of high-resolution, large-coverage, and cost-effective perception capabilities for soft robotic systems. Future research will focus on integrating adjustable skeletal stiffness and advanced perception field materials to further enhance the realism of perception, particularly in texture sensing. Additionally, we plan to expand FlexiRay’s capabilities to multi-finger grippers and explore its applications in dynamic, high-safety tasks such as fruit picking and assistive care, paving the way for the next generation of robust, adaptable, and intelligent human-robot interaction systems.

Methods

Illumination system implementation

- Camera. We chose the camera with high-resolution image capture capability (12 million pixels), compact structure (8 ×8.5 ×5 mm), and wide-angle view (135°). Hence, the camera can be positioned within the gripper to obtain clear internal imagery.

- Illumination. Flexible RGBW (red, green, blue, white) LED lights are embedded in the silicone gel in a series configuration during fabrication. Interlaced side beams are designed to prevent environmental light interference while preserving adaptability.

- Mirrors. The mirrors are strategically mounted on the finger's backbone at specific angles and positions to augment the visual field, ensuring comprehensive visibility even during substantial deformations. To preserve the compliance of the finger without introducing structural interference, the rigid planar mirror is designed in a T-shape. This configuration minimizes the bonded area with the back flexible beam while ensuring that the reflective surface remains sufficiently large.

Skeleton and tactile sensing pad implementation

- Finger. The Fin Ray finger manufactured using TPU materials is capable of conforming to various object shapes without active actuation.

- Tactile Sensing Pad. The tactile sensing pad primarily consists of four layers. We combine Smooth On Inc. 00-30 silicone with thermochromic pigment to form the temperature-sensing layer. For the reflective layer, silicone is mixed with aluminum flake and aluminum powder in a mass ratio of

400:20:3. To enable the acquisition of detailed textural information, a low-hardness (5 A) transparent silicone is used as the elastic layer. Additionally, we employ PDMS as an elastic support to preserve the compliance and deformation capacity of the Fin Ray.

Fabrication and manufacturing

Our primary objective is to endow FRE with the capability to perceive tactile information through a vision-based tactile method while maintaining its compliance and passive deformation abilities. To achieve this, we replaced the traditional acrylic support structure with PDMS material and opted for flexible LED light strips for illumination. The fabrication process is as follows:

First, Smooth On Inc. 00-30 silicone is mixed with aluminum powder and flake following the specified mass ratioas the reflective layer. This mixture is spread with an squeegee applicator to create a thin silicone film with a thickness of 0.15 mm, which is then cured at 100 °C for 1 hour. Subsequently, a laser cutter is utilized to remove an array of round markers. The silicone is then mixed with two thermochromic pigments in a mass ratio of 10:1:1 to form the temperature-sensing layer. The pigments possess thermochromic properties, allowing them to detect temperature variations and undergo color changes in response to temperatures exceeding or falling below the threshold values. Hence, the two thermosensitive powders facilitate a temperature-responsive transition of the thermometric layer: it exhibits deep red above 38°C, shifts to deep blue below 18 °C, and appears light purple under ambient conditions. With the same application method, we can obtain the 0.15 mm thick temperature-sensing layer and bond the two layers together. Afterward, we place the prepared composite layer into the mold and mix a low-hardness, transparent silicone in a ratio of 2.2:1 to create a soft, elastic layer with a hardness (5 shore A) similar to human skin. The silicone is poured into the mold and cured at room temperature for 90 minutes. Subsequently, the RGBW LEDs are connected in series and affixed to the TPU front beam. PDMS is mixed in a ratio of 10:1 to serve as a transparent support structure capable of withstanding large deformations. The TPU front beam and LED lights are placed into the mold. Further, PDMS material is poured to encapsulate these components, resulting in a tactile sensing pad capable of perceiving texture, temperature, and other contact information.

After fabricating the tactile sensing silicone pad, we can assemble the mirrors, camera, side beams, finger framework and the soft pad to construct the FlexiRay. To validate its perception and grasping capabilities, we propose a gripper consisting of two identical FRE fingers. A single stepper motor actuates the gripper through linkage mechanisms, enabling a substantial range of motion for opening and closing actions.

CMA-ES based optics layout optimization

The equilibrium between finger compliance and the robustness of the tactile sensing region presents a significant challenge in the advancement of soft VTS. To mitigate the limitations of FOV and occlusion blind spots associated with large-deformation finger structures, a multi-mirror FOV optimization method is proposed, leveraging CMA-ES[36,37]. Through the optimization of both structural and layout parameters of the camera and mirrors, a visual reflection system composed of multiple planar mirrors is established on the back side of the finger. This configuration facilitates a single camera in achieving comprehensive coverage of the entire deformation range and contact region of the Fin Ray, thereby enabling enhanced visual-tactile perception.

Initially, the displacement of nodes under the applied force F is collected for the Fin Ray, which has identical structural parameters. The displacement is represented by the coordinates of the lateral 2D cross-section, with the back joint nodes denoted as $\{\boldsymbol{N}_\mathrm{i}\}_{i=1}^{n}$ and the tactile sensing area joint nodes labeled as $\{\boldsymbol{P}_i\}_{i=1}^{n}$. Consequently, a mapping from force to deformation is established, represented as $f\colon F \to \{\boldsymbol{N}_i, \boldsymbol{P}_i\}_{i=1}^{n}$. The subsequent part provides a detailed account of the construction and solution of the optical layout optimization problem.

Decision variables:

$$\boldsymbol{x} = \left\{\left(\theta_i^{\mathrm{mir}}, t_i^{\mathrm{mir}}, l_i^{\mathrm{mir}}\right)\right\}_{i=1}^{n-1} \cup \{u, \phi\}, \quad \text{with } \theta_i^{\mathrm{mir}}, t_i^{\mathrm{mir}}, l_i^{\mathrm{mir}}, u, \phi \in \mathbb{R} \tag{1}$$

where $\{\theta_i^{\text{mir}}, t_i^{\text{mir}}, l_i^{\text{mir}}\}$ represents the parameters associated with the mirrors. Specifically, θ_i^{mir} denotes the angle between the mirror and the line segment formed by two adjacent nodes on the back side, t_i^{mir} indicates the midpoint offset distance, and l_i^{mir} signifies the length of the mirror. The camera is positioned along the baseline defined by the back side and the base nodes of the tactile sensing pad. The distance from the camera to the back base node is expressed as a coefficient u, relative to the length of the baseline. Furthermore, the angle between the optical axis of the camera and the baseline is represented by ϕ.

Objective function:

Given a camera with a fixed FOV, the discrete rays produced by the camera can be represented as $\mathcal{R} = \{\boldsymbol{r}_j\}_{j=1}^{m}$. The primary objective of the optimization process is to maximize the coverage of the tactile sensing region by the collection of FOV rays across various deformations. This goal can be achieved through two approaches: direct imaging and single reflections from mirrors.

By sampling K distinct loads from the force-displacement mapping f, a set of joint nodes $\mathcal{D}$ is obtained. Each deformation structure $\boldsymbol{d}$ corresponds to a discrete set of target points $\boldsymbol{p}$, denoted as $\mathcal{P}^d$, within the tactile sensing region. The radius of the illumination range for a single FOV ray is defined as $R \propto l^c$, which depends on the propagation distance l^c. Consequently, the objective function can be expressed as follows:

$$\text{Maximize } f(x) = \frac{1}{K}\sum_{\boldsymbol{d}\in\mathcal{D}}\sum_{\boldsymbol{r}\in\mathcal{R}}\sum_{\boldsymbol{p}\in\mathcal{P}^d} I(\boldsymbol{x}, \boldsymbol{r}, \boldsymbol{p}, R) \tag{2}$$

where $I(\boldsymbol{x}, \boldsymbol{r}, \boldsymbol{p}, R)$ is an indicator function that returns 1 if the target point resides within the coverage range of the ray beam; otherwise, it returns 0.

Constraints:

The following constraints must be adhered to:

- **Geometric constraints.** The lengths, offsets, and rotation angles of each mirror, as well as the position of the camera, must remain within predefined ranges: $l_i^{\text{mir}} \in [l_{\min}, l_{\max}]$, $t_i^{\text{mir}} \in [t_{\min}, t_{\max}]$, $\theta_i^{\text{mir}} \in [\theta_{\min}, \theta_{\max}]$, $u \in (0,1)$.

- **Occlusion constraints.** The potential occlusion between mirrors during the deformation process must be taken into account. If an occlusion exists along the line of sight, the indicator function $I(\boldsymbol{x}, \boldsymbol{r}, \boldsymbol{p}, R)$ returns 0.
- **Safety constraints.** Under all deformation conditions, mirrors must not interfere with the front and back beams. If this condition is not met, the indicator function $I(\boldsymbol{x}, \boldsymbol{r}, \boldsymbol{p}, R)$ returns 0 as a penalty.

Solution process:

The optimization process utilizes the CMA-ES algorithm[38] to search for optimal layout solutions by sampling candidate solutions from a multivariate Gaussian distribution. The multivariate Gaussian distribution is defined as $\mathcal{N}(\boldsymbol{\mu}, \sigma^2\boldsymbol{C})$, where $\boldsymbol{\mu}$ is the mean vector, $\boldsymbol{C}$ is the covariance matrix, and σ represents the step size.

In the y-th generation of the optimization process, for a population size of β, the candidate solutions $\{\boldsymbol{x}_i\}_{i=1}^{\beta}$ are obtained by sampling from the distribution $\mathcal{N}\left(\boldsymbol{\mu}^{(y)}, \left(\sigma^{(y)}\right)^2\boldsymbol{C}^{(y)}\right)$:

$$\boldsymbol{x}_i = \boldsymbol{\mu}^{(y)} + \sigma^{(y)}\sqrt{\boldsymbol{C}^{(y)}}\boldsymbol{s}_i \tag{3}$$

where $\boldsymbol{s}_i \sim \mathcal{N}(0, \boldsymbol{I})$, with $\boldsymbol{I}$ denoting the identity matrix. The k solutions with the highest objective function values are selected as candidates: $\{f(\boldsymbol{x}_{1:k}) \mid f(\boldsymbol{x}_1) \geq f(\boldsymbol{x}_2) \geq \cdots \geq f(\boldsymbol{x}_k) \geq \cdots \geq f(\boldsymbol{x}_\beta)\}$

The evolution paths are updated using the following equations:

$$\boldsymbol{p}_\sigma^{(y+1)} = (1 - c_\sigma)\boldsymbol{p}_\sigma^{(y)} + \sqrt{c_\sigma(2 - c_\sigma)\lambda_w}\sqrt{\boldsymbol{C}^{(y)}}^{-1} \tag{4}$$

$$\boldsymbol{p}_c^{(y+1)} = (1 - c_c)\boldsymbol{p}_c^{(y)} + \sqrt{c_c(2 - c_c)\lambda_w}H_\sigma^{(y+1)} \tag{5}$$

where c_σ and c_c are cumulation factors, respectively. $H_\sigma^{(y+1)}$ is the Heaviside function, and

$$\delta = \sum_{i=1}^{\eta} w_i\sqrt{\boldsymbol{C}^{(y)}}\,\boldsymbol{s}_i, \quad \lambda_w = \frac{1}{\sum_{i=1}^{\eta} w_i^2}, \quad \sum_{i=1}^{\eta} w_i = 1$$

Therefore, the parameters of the multivariate Gaussian distribution for each generation are updated as follows:

$$\boldsymbol{\mu}^{(y+1)} = \boldsymbol{\mu}^{(y)} + c_l \sigma^{(y)} \sqrt{\boldsymbol{C}^{(y)}} \boldsymbol{s}_i \tag{6}$$

$$\sigma^{(y+1)} = \sigma^{(y)} \exp\left(1, c_\sigma \left(\frac{\left\|\boldsymbol{p}_\sigma^{(y+1)}\right\|}{\mathbb{E}[\mathcal{N}(0,\boldsymbol{I})]} - 1\right) / d_\sigma\right) \tag{7}$$

$$\boldsymbol{C}^{(y+1)} = \left(1 + c_1 c_\mu (2 - c_c)\left(1 - H_\sigma^{(y+1)}\right)\right) \boldsymbol{C}^{(y)} + c_1 \Delta_1 + c_\mu \Delta_\mu \tag{8}$$

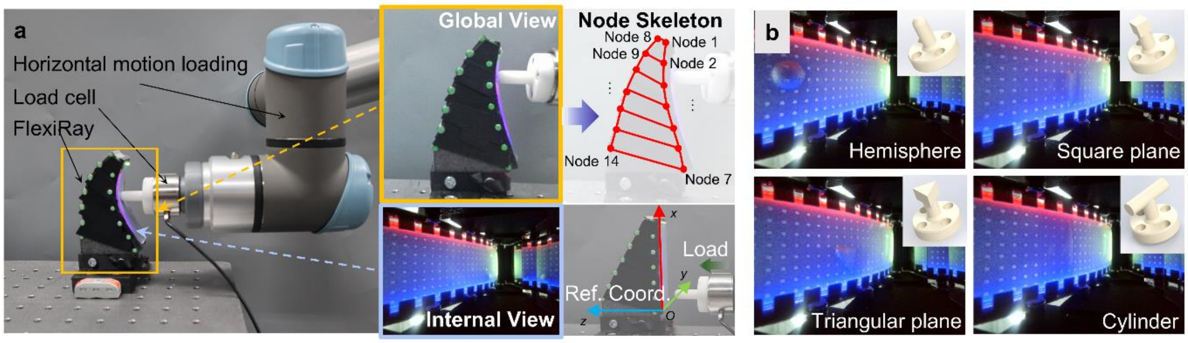
where Δ_1 and Δ_μ correspondingly denote the rank one and $-\mu$ updates. The parameters c_l, c_1, and c_μ represent the learning rates for the mean, rank one, and $-\mu$ updates, respectively. The damping factor d_σ is used for the adaptive accumulation of step sizes.

Starting with initial parameters that satisfy the constraints, the parameters of the multivariate Gaussian distribution are updated iteratively. This process continues until either the expected stopping condition for the objective function is met or the maximum number of optimization generations is reached. The optimized camera and mirror layout parameters are then derived from the distribution.

Proprioception data collection platform and procedures

The dataset collection platform for proprioception tasks is shown in Fig. 8a. The FlexiRay is mounted at a fixed position on the optical platform, with the reference coordinate system established at the lower-left corner of the tactile sensing pad in its initial state. A normal force sensor is attached to the end of the UR5e robotic arm, with the arm's tool coordinate system adjusted so that its z-axis is parallel to the z-axis of the reference coordinate system, allowing a force to be applied along this direction. A 3D-printed contact probe is attached to the force sensor, designed to simulate four contact surface types: hemispherical, square, triangular, and cylindrical, as shown in Fig. 8b. These designs replicate various contact forms, including point, edge, and flat surface interactions. Initially, the probe does not contact the finger. Random initial x and y positions within the feasible contact domain are selected, and a force is applied along the z-axis. Upon detecting contact via the force sensor, synchronized data collection is initiated. This process involves capturing side beam node images using a global camera, recording internal finger images through the embedded camera, and logging force sensor readings. Simultaneously, the robot's end-effector positions are recorded in real time to ensure comprehensive data alignment.

Fig. 8: Data collection platform and procedures.



a Platform setup. **b** Probe type of the load cell.

Machine learning implementation

In terms of model implementation, the PP-LiteSeg semantic segmentation model is built upon the STDC2 backbone, with the objective function using Cross-Entropy Loss. The model is optimized using Stochastic Gradient Descent (SGD) with momentum set to 0.9 and weight decay of 4×10^{-5}. A polynomial decay learning rate scheduler is employed, with an initial learning rate of 0.01 and a decay factor of 0.9. After training, the model's performance evaluation shows a mean Intersection over Union (mIoU) of 89.98%. For the proprioception model, three heads are used to predict normal contact force, contact position, and node localization, each consisting of fully connected layers with 512 and 256 neurons. The output layers have dimensions of $\mathbb{R}^1$, $\mathbb{R}^3$ and $\mathbb{R}^{28}$, respectively. The loss function for all three outputs is Mean Squared Error (MSE), with the learning rate set to 0.001. The dataset comprises 5,000 tactile image samples, which are split into training and validation sets at a 4:1 ratio. Additionally, the classification head of the texture recognition model is composed of fully connected layers with 512 and 256 neurons. The loss function used is Cross-Entropy Loss, and the model is trained with the Adam optimizer, with a learning rate set to 0.001.

Data Availability

The data that support the findings of this study are available within the paper and the Supplementary data files. Other data generated during the current study are available from the corresponding author on reasonable request.

Code Availability

The demo implementation of proprioception model is provided in the Supplementary Dataset files.

References

1. Dahiya, R. S., Metta, G., Valle, M. & Sandini, G. Tactile sensing—from humans to humanoids. *IEEE transactions on robotics* **26**, 1–20 (2009).

2. Felten, D. L., O'Banion, M. K. & Maida, M. S. 9 - Peripheral Nervous System. in *Netter's Atlas of Neuroscience (Third Edition)* (eds. Felten, D. L., O'Banion, M. K. & Maida, M. S.) 153–231 (2016).

3. Handler, A. & Ginty, D. D. The mechanosensory neurons of touch and their mechanisms of activation. *Nature Reviews Neuroscience* **22**, 521–537 (2021).

4. She, Y., Liu, S. Q., Yu, P. & Adelson, E. Exoskeleton-covered soft finger with vision-based proprioception and tactile sensing. in *2020 IEEE International Conference on Robotics and Automation (ICRA)* 10075–10081 (2020).

5. Liu, Z. *et al.* A three-dimensionally architected electronic skin mimicking human mechanosensation. *Science* **384**, 987–994 (2024).

6. Zhou, Z. *et al.* A sensory soft robotic gripper capable of learning-based object recognition and force-controlled grasping. *IEEE Transactions on Automation Science and Engineering* **21**, 844–854 (2022).

7. Yan, Y. *et al.* Soft magnetic skin for super-resolution tactile sensing with force self-decoupling. *Science Robotics* **6**, eabc8801 (2021).

8. Taylor, S., Park, K., Yamsani, S. & Kim, J. Fully 3D printable Robot Hand and Soft Tactile Sensor based on Air-pressure and Capacitive Proximity Sensing. in *2024 IEEE International Conference on Robotics and Automation (ICRA)* 18100–18105 (2024).

9. Shorthose, O., Albini, A., He, L. & Maiolino, P. Design of a 3D-Printed Soft Robotic Hand With Integrated Distributed Tactile Sensing. *IEEE Robotics and Automation Letters* **7**, 3945–3952 (2022).

10. Zhao, H., O'Brien, K., Li, S. & Shepherd, R. F. Optoelectronically innervated soft prosthetic hand via stretchable optical waveguides. *Science Robotics* **1**, eaai7529 (2016).

11. Meerbeek, I. M. V., Sa, C. M. D. & Shepherd, R. F. Soft optoelectronic sensory foams with proprioception. *Science Robotics* **3**, eaau2489 (2018).

12. Wall, V., Zöller, G. & Brock, O. Passive and active acoustic sensing for soft pneumatic actuators. *The International Journal of Robotics Research* **42**, 108–122 (2023).

13. Li, G., Liu, S., Wang, L. & Zhu, R. Skin-inspired quadruple tactile sensors integrated on a robot hand enable object recognition. *Science Robotics* **5**, eabc8134 (2020).

14. Spiegel, J. V. der. Computational sensors: The basis for truly intelligent machines. in *Intelligent Sensors* (ed. Yamasaki, H.) vol. 3 19–37 (Elsevier Science B.V., 1996).

15. Abad, A. C., Reid, D. & Ranasinghe, A. HaptiTemp: A next-generation thermosensitive GelSight-like visuotactile sensor. *IEEE Sensors Journal* **22**, 2722–2734 (2021).

16. Fang, B. *et al.* Force Measurement Technology of Vision-Based Tactile Sensor. *Advanced Intelligent Systems* 2400290 (2024).

17. Xu, W. *et al.* A Vision-Based Tactile Sensing System for Multimodal Contact Information Perception via Neural Network. *IEEE Transactions on Instrumentation and Measurement* (2024).

18. Abad, A. C. & Ranasinghe, A. Visuotactile sensors with emphasis on gelsight sensor: A review. *IEEE Sensors Journal* **20**, 7628–7638 (2020).

19. Johnson, M. K. & Adelson, E. H. Retrographic sensing for the measurement of surface texture and shape. in *2009 IEEE Conference on Computer Vision and Pattern Recognition* 1070–1077 (2009).

20. Kakani, V., Cui, X., Ma, M. & Kim, H. Vision-based tactile sensor mechanism for the estimation of contact position and force distribution using deep learning. *Sensors* **21**, 1920 (2021).

21. Yuan, W., Dong, S. & Adelson, E. H. Gelsight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors* **17**, 2762 (2017).

22. Yuan, W., Mo, Y., Wang, S. & Adelson, E. H. Active clothing material perception using tactile sensing and deep learning. in *2018 IEEE International Conference on Robotics and Automation (ICRA)* 4842–4849 (2018).

23. Li, S. *et al.* M^3Tac: A Multispectral Multimodal Visuotactile Sensor With Beyond-Human Sensory Capabilities. *IEEE Transactions on Robotics* **40**, 4506–4525 (2024).

24. Do, W. K. & Kennedy, M. Densetact: Optical tactile sensor for dense shape reconstruction. in *2022 International Conference on Robotics and Automation (ICRA)* 6188–6194 (2022).

25. Wang, S., She, Y., Romero, B. & Adelson, E. GelSight Wedge: Measuring High-Resolution 3D Contact Geometry with a Compact Robot Finger. in *2021 IEEE International Conference on Robotics and Automation (ICRA)* 6468–6475 (2021).

26. Tippur, M. H. & Adelson, E. H. GelSight360: An omnidirectional camera-based tactile sensor for dexterous robotic manipulation. in *2023 IEEE International Conference on Soft Robotics (RoboSoft)* 1–8 (2023).

27. Lambeta, M. *et al.* Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation. *IEEE Robotics and Automation Letters* **5**, 3838–3845 (2020).

28. Donlon, E. *et al.* Gelslim: A high-resolution, compact, robust, and calibrated tactile-sensing finger. in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* 1927–1934 (2018).

29. Zhao, J. & Adelson, E. H. Gelsight svelte: A human finger-shaped single-camera tactile robot finger with large sensing coverage and proprioceptive sensing. in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* 8979–8984 (2023).

30. Crooks, W., Vukasin, G., O'Sullivan, M., Messner, W. & Rogers, C. Fin ray® effect inspired soft robotic gripper: From the robosoft grand challenge toward optimization. *Frontiers in Robotics and AI* **3**, 70 (2016).

31. Liu, S. Q. & Adelson, E. H. Gelsight fin ray: Incorporating tactile sensing into a soft compliant robotic gripper. in *2022 IEEE 5th International Conference on Soft Robotics (RoboSoft)* 925–931 (2022).

32. Liu, S. Q., Ma, Y. & Adelson, E. H. Gelsight baby fin ray: A compact, compliant, flexible finger with high-resolution tactile sensing. in *2023 IEEE International Conference on Soft Robotics (RoboSoft)* 1–8 (2023).

33. Wang, S., She, Y., Romero, B. & Adelson, E. Gelsight wedge: Measuring high-resolution 3d contact geometry with a compact robot finger. in *2021 IEEE International Conference on Robotics and Automation (ICRA)* 6468–6475 (2021).

34. Sun, H., Kuchenbecker, K. J. & Martius, G. A soft thumb-sized vision-based sensor with accurate all-round force perception. *Nature Machine Intelligence* **4**, 135–145 (2022).

35. Liu, X., Han, X., Hong, W., Wan, F. & Song, C. Proprioceptive learning with soft polyhedral networks. *The International Journal of Robotics Research* 02783649241238765 (2024).

36. Hansen, N., Müller, S. D. & Koumoutsakos, P. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary computation* **11**, 1–18 (2003).

37. Nomura, M., Akimoto, Y. & Ono, I. CMA-ES with Learning Rate Adaptation. *ACM Transactions on Evolutionary Learning* (2024).

38. Nomura, M. & Shibata, M. Cmaes: A simple yet practical Python library for CMA-ES. *arXiv preprint arXiv:2402.01373* (2024).

Acknowledgements (optional)

This work was supported in part by the National Natural Science Foundation of China through the Youth Program under Grant 509109-N72401; in part by the National Natural Science Foundation of China through the high-level Youth Talent Program under Grant 588020-X4240A/008; in part by the China Postdoctoral Science Foundation under Grant 2024M762814.

Ethics declarations

Competing interests

The authors declare no competing interests.