This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Tac6D: High-Accuracy Pose and Force Sensing Smart Joint for Hyper-Redundant Manipulators With Camera-Based Tactile Sensing

Zhanxuan Peng, Jieji Ren, Zhenpu Zhu, Jiang Zou, *Member, IEEE*, and Guoying Gu, *Senior Member, IEEE*

***Abstract*—Precise control for cable-driven hyper-redundant manipulators is fundamentally constrained by the lack of reliable state perception. Conventional tension-based sensing suffers from friction and hysteresis, often necessitating excessive cable tension that risks structural buckling. The critical bottleneck lies in the absence of a lightweight, in-situ sensor capable of simultaneous, high-precision tip-contact force and pose perception within the strict spatial and loading constraints of the joint. To address these challenges, we present Tac6D, a lightweight and low-cost smart joint that utilizes a camera-based sandwich structure to achieve high-precision and multimodal perception. Specifically, by observing the mechanical deformation of the gel layer in response to external forces, the bottom camera accurately estimates the joint's pose and external tip-contact forces. To maximize resolution and perception range, we introduce a symmetrical indenter layout with jointly optimized structural parameters. Based on this mechanism, we design a novel multifeature fusion neural network that achieves accurate pose and tip-contact force perception with mean errors of 0.11° and 0.085 N, respectively. We demonstrate the high-performance sensing of Tac6D on hyper-redundant manipulators and successfully achieve pose and tip-contact force feedback control for complex precision operations. Tac6D provides an accurate, versatile, and cost-effective sensing solution for multimodal robotic smart joints.**

***Index Terms*—Camera-based tactile sensing, hyper-redundant manipulators, pose perception, robotic manipulation, tip-contact force perception.**

Received 15 May 2026; revised 27 July 2026 and 11 September 2026; accepted 9 October 2026. Recommended by Technical Editor J. Chen and Senior Editor Y.-J. Pan. This work was supported in part by the National Science and Technology Major Project under Grant 2025ZD1606600, in part by the Science and Technology Commission of Shanghai Municipality under Grant 24511103400 and Grant 25ZR1401191, in part by the National Natural Science Foundation of China under Grant 52505029 and Grant 52625509, in part by the Shanghai Eastern Talent Plan (Youth Project) under Grant QNKJ2025060 and in part by Xplore Prize. *(Zhanxuan Peng and Jieji Ren contributed equally to this work.) (Corresponding author: Guoying Gu.)*

The authors are with the State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China, and also with the Shanghai Key Laboratory of Intelligent Robotics, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: steven2606@sjtu.edu.cn; jiejiren@sjtu.edu.cn; zhuzhenpu@sjtu.edu.cn; zoujiang@sjtu.edu.cn; guguoying@sjtu.edu.cn).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TMECH.2026.3743091.

Digital Object Identifier 10.1109/TMECH.2026.3743091

I. Introduction

CABLE-DRIVEN hyper-redundant manipulators are ideal for compelling applications in confined spaces, such as inspecting aircraft fuel tanks and wing interiors [1]. While their high degrees of freedom (DoFs) and slender structure enable flexible navigation, they pose significant challenges in sensing and control. First, due to serial kinematic error accumulation, even marginal joint-level deviations significantly compromise end-effector positioning, necessitating high-precision pose perception. Second, the inherent low stiffness of long-reach cable-driven structures demands robust force feedback to enable active error compensation and compliant interaction. Traditionally, joint states are inferred from proximal cable tensions. However, this approach is plagued by complex cable-conduit friction and mechanical hysteresis. Such nonlinearities often compel the use of excessive cable pretension [2] to ensure stability, which risks structural buckling and complicates actuator design [3]. Consequently, a critical gap exists for a lightweight, low-cost, and in-situ multimodal sensor that can bypass transmission uncertainties. Addressing the tradeoff between high-fidelity perception and the strict spatial constraints within compact joints remains the primary bottleneck hindering the widespread deployment of hyper-redundant manipulators.

Existing sensing solutions struggle to balance compactness, multimodality, and cost. Conventional discrete sensors, such as magnetic sensors [2] and encoders [4], are typically integrated orthogonally within universal joints to decouple the two rotational DoFs. While effective for pose estimation, magnetic sensors suffer from thermal drift and electromagnetic interference. Encoders, despite their precision, are limited to single-modal kinematic sensing and cannot intrinsically perceive external forces. Fiber optic technologies e.g., fiber Bragg grating (FBG) sensors [5], [6], have been developed for shape reconstruction and tension monitoring in continuum structures. However, their widespread adoption is hindered by high costs and bulky demodulation equipment. Moreover, decoupling external contact forces from internal cable tension remains fundamentally challenging due to nonlinear friction and hysteresis along the cable transmission path [3], [7]. Algorithmic estimation strategies

1083-4435 © 2026 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Shanghai Jiaotong University. Downloaded on November 05,2026 at 09:44:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

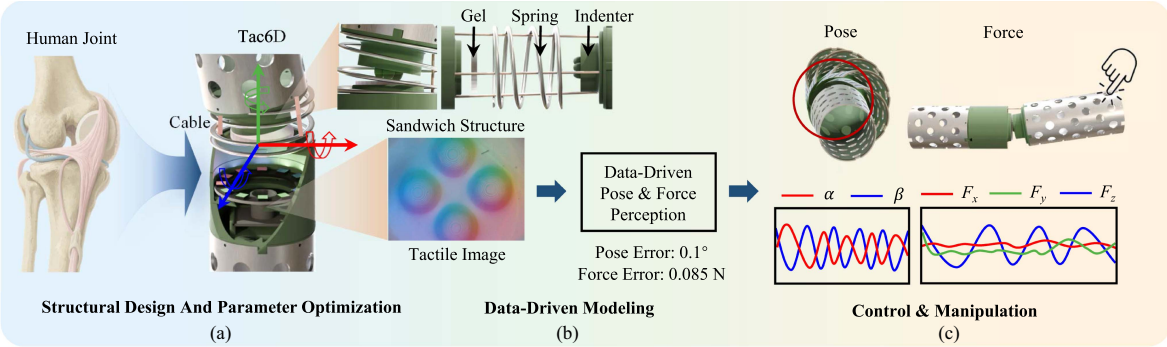


Fig. 1. (a) Camera-based tactile smart joint (Tac6D) is developed for pose and tip-contact force perception of hyper-redundant manipulators. (b) Pose and tip-contact force perception is realized via pose and force estimation networks. The joint outputs high-precision joint posture and end-tip contact force after data acquisition and network training. (c) This allows the manipulator to perform operational tasks with the manipulator's low-level controller. (a) Structural design and parameter optimization. (b) Data-driven modeling. (c) Control & manipulation.

using deep neural networks or dynamic models have been explored to compensate for transmission nonlinearities [8], [9], yet these indirect methods suffer from computational complexity and poor generalization. Commercial multiaxis force/torque (F/T) sensors are prohibitively bulky for confined joint spaces, while external vision-based methods rely on line-of-sight that is often obstructed in confined environments [10]. Consequently, achieving simultaneous pose and force perception within a compact, integrated form factor remains a significant challenge.

Camera-based tactile sensors have emerged as a promising solution for in-situ sensing in hyper-redundant manipulators [11], [12]. By optically capturing soft medium deformation, these sensors provide high-density, high-precision information, enabling simultaneous pose and force perception via a single compact camera. However, integrating such sensors into constrained joint spaces presents three fundamental challenges. First, conventional soft elastomer layers struggle to balance load-bearing capacity with sensing sensitivity, as the gel undergoes excessive deformation or permanent damage under sustained high internal loads (typically 50–100 N cable tensions). Second, simultaneous joint pose and external contact force perception from gel deformation fields remains difficult [13], [14]. Third, achieving adequate measurement resolution and range poses significant constraints [15]. Addressing these challenges requires a systematic approach encompassing mechanical design optimization, sensing mechanism innovation, and data-driven modeling.

To address the aforementioned challenges in simultaneous pose and tip-contact force perception for hyper-redundant manipulators, this article proposes a novel joint-embedded, camera-based tactile module, named "Tac6D." As illustrated in Fig. 1, this sensor provides a lightweight, compact, and information-intensive solution tailored for compact assembly space. The main contributions of this work are summarized as follows.

1) *A joint-level sensing solution* for hyper-redundant manipulator joints that addresses the coexistence of compact spatial constraints and high internal loads.
2) *A functionally decoupled "indenter-spring-gel" architecture* that separates the load-bearing function from the sensing function, enabling high-fidelity tactile perception under sustained high-load conditions that would degrade or damage conventional soft elastomer-based sensors.
3) *A parametric optimization framework* achieving a measurable range of $\pm 10^\circ$ with angular resolution below 0.3°, and *multifeature fusion neural networks* yielding mean angular error of 0.11° and force error of 0.085 N.
4) *Comprehensive experimental validations* demonstrating the sensor's effectiveness in closed-loop control for contact-rich manipulation tasks, including precision peg-in-hole assembly.

The rest of this article is organized as follows: Section II reviews related work on hyper-redundant manipulators and camera-based tactile sensors. Section III presents the proposed framework, including sensing mechanism, structure design, modeling, and perception algorithm. Section IV reports evaluation results and potential applications. Finally, Section V concludes this article with contributions, limitations, and future directions.

II. Related Work

A. Hyper-redundant Manipulators and Sensor Integration

Hyper-redundant manipulators are characterized by high length-to-radius ratios and continuous backbone structures. To minimize distal inertia and maintain a slender profile, these manipulators typically employ cable-driven mechanisms with proximally mounted motors. This configuration reduces structural weight but introduces significant transmission nonlinearities. Cable elongation, friction-induced hysteresis, and coupling effects severely degrade proprioception accuracy based solely on motor-side encoders [2], [3]. Consequently, in-situ measurement methods have gained prominence [6]. Current design practices for cable-driven hyper-redundant manipulators present extremely limited integration of onboard sensing. Only simple sensing schemes are adopted in rigid-link configurations [16], [17]. FBG sensors enable precise shape reconstruction. However, they require extremely dense sampling, leading to high fabrication costs [18]. Hall effect sensors offer a compact and cost-effective alternative [19]. However, the proximity of orthogonal magnetic fields causes severe coupling, making signal decoupling computationally intensive. Furthermore, even minor magnet-sensor axis misalignments cause significant nonlinear errors due to assembly tolerances [20].

Authorized licensed use limited to: Shanghai Jiaotong University. Downloaded on November 05,2026 at 09:44:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
PARAMETERS OF COMMERCIAL SIX-AXIS FORCE SENSORS

Sensor Name	Diameter (mm)	Weight (g)	Load (N)	Price (USD)
ATI Axia80-M8	82.0	280	1500	4000–6000
ATI Omega85	85.0	658	1900	10860–13030
ATI Omega160	160.0	2720	5000	5500–6080
BOTA MiniONE Pro	48.0	57.0	50.0	500–1500
Yuli M3701	15.0	10.0	50	300–1000
Kunwei KWR58	58.0	200	1000	1000–3000
Keprui Micro Sensor	9.60	3.00	40.0	200–800
Weltronics Micro Sensor	9.60	3.00	40.0	200–800
Requirements	$\leq$ 55.0	$\leq$ 100	$\geq$ 150	$\leq$ 100

Prices estimated from manufacturer datasheets and industry quotations (accessed 2026).

Beyond configuration reconstruction, implementing simultaneous force perception presents an even greater challenge due to the conflict between limited integration space and the bulky form factor of commercial sensors. The design requirements in Table I are derived from the SJTU-Snake series specifications [3]: a 55-mm diameter constraint dictated by the joint envelope, a $\leq$ 100-g weight budget to minimize distal inertia, and a $\geq$150 N load capacity to withstand typical cable tensions (50–100 N) with safety margin. As the table shows, commercial F/T sensors that fit within this envelope typically fail to withstand the axial loads from cable tension, while those meeting load requirements are prohibitively large. Furthermore, the crowded internal volume of hyper-redundant joints makes sensor integration challenging, as it risks interfering with cable routing. Researchers often resort to proximal sensing [21], [22], but this indirect approach suffers from severe transmission loss due to cumulative friction along cable paths. Although Coulomb friction models have been proposed for compensation [23], they struggle with the stochastic nature of cable-hole contact states [24], rendering precise distal force estimation through proximal sensing practically unfeasible.

B. Camera-Based Tactile Sensing for Pose and Force

Camera-based tactile sensors offer cost-effectiveness and high information density, making them a prominent solution for robotic perception. By capturing the continuous deformation of a soft elastomer, these sensors reconstruct high-fidelity contact topologies and force distributions [11], which are essential for contact-rich tasks [25]. For hyper-redundant manipulators, purely kinematic models often fail to capture the complex distributed compliance of the structure. The microscale deformation sensitivity of camera-based tactile sensors thus provides a unique advantage, enabling simultaneous monitoring of internal structural states and external interaction forces and making this modality a promising candidate for enhancing the proprioception of continuum robots.

Beyond discrete contact detection, the scalability of camera-based architectures facilitates their integration along continuum robots and hyper-redundant manipulators. Systems, such as ProTac [26] and ConTac [27] demonstrate this by observing internal marker displacements within hyperelastic skins (e.g., Dragon Skin 10) to reconstruct multiaxial forces and local surface curvature. While these designs effectively validate the strategy of using vision for self-state perception and collision avoidance, they predominantly rely on highly compliant elastomers to maximize sensitivity. Consequently, a critical limitation remains. When such soft sensing media are directly subjected to high internal loads (e.g., cable tensions in hyper-redundant manipulators), they undergo excessive deformation that compromises sensing fidelity or leads to permanent damage.

To decode the coupled deformation fields captured by camera-based tactile sensors, data-driven approaches have been extensively validated [28]. Architectures ranging from lightweight multilayer perceptrons (MLPs) to convolutional neural networks (CNNs) [29], [30] are successfully employed to extract high-level representations. However, high-fidelity prediction of complex physical attributes, such as contact force distribution, typically requires deeper and more expressive network structures to model the underlying high-dimensional mappings. Consequently, increasing network depth not only elevates the computational burden for training and inference but can also lead to the attenuation of shallow features essential for precision. Although recent studies have attempted to mitigate this by reintegrating shallow features [31] or employing multiple shallow network ensembles [32], achieving an optimal balance between feature retention and model complexity remains an active research challenge.

In summary, three critical bottlenecks impede the practical deployment of camera-based tactile sensors in hyper-redundant manipulators. First, the high internal tension of cable-driven mechanisms forces excessive deformation of the soft sensing medium, degrading fidelity or causing permanent damage. A mechanism that decouples load capacity from sensing range is therefore needed. Second, sensor resolution is intrinsically tied to the geometric contact interface, requiring rigorous optimization of indenter parameters to maximize information gain from tactile images. Third, the effectiveness of data-driven methods depends heavily on network design. Acquiring reliable datasets spanning the full operational space is difficult due to the complex interplay between internal actuation forces and external contact loads. To address these issues, we propose Tac6D, a camera-based tactile sensing module integrated directly into the joint architecture of the SJTU-Snake III. By replacing the traditional gimbal assembly, Tac6D achieves robust mechanical integration and employs a specialized learning framework to estimate pose and tip-contact forces under quasi-static conditions.

III. METHODS

In this section, the hardware design, the data acquisition process, and data-driven models are introduced. First, an "indenter-spring-gel" sandwich-type structure is applied to withstand cable tension. Second, a parametric indenter design scheme is

Authorized licensed use limited to: Shanghai Jiaotong University. Downloaded on November 05,2026 at 09:44:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

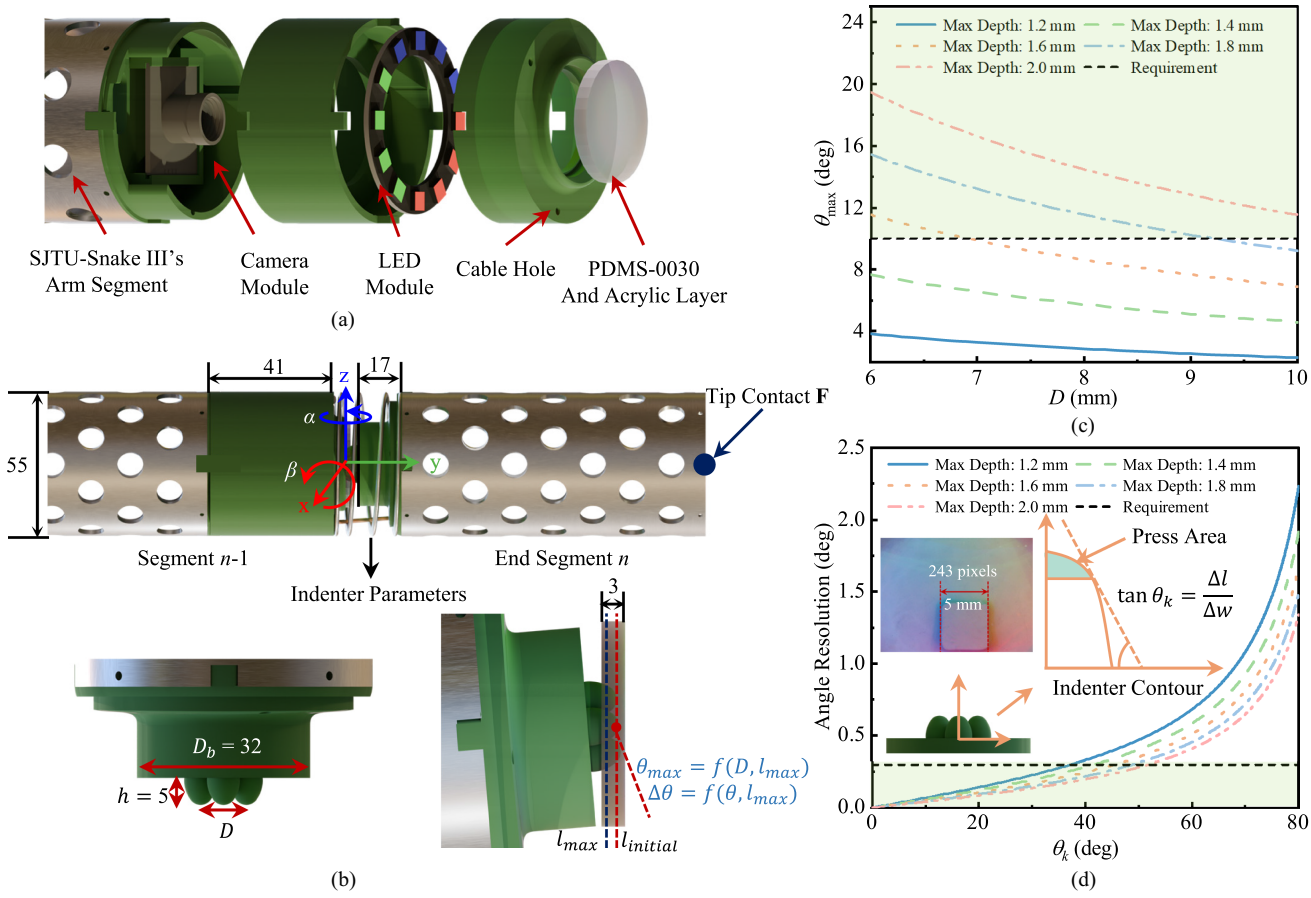


Fig. 2. (a) Exploded view and (b) the "indenter-spring-gel" sandwich structure of the Tac6D joint (unit: mm). To ensure the measurement range and the angle resolution, (c) the indenter distance D and (d) the maximum tangential angle θ_k are parametric optimized.

proposed to control the sensor's maximum measurable range and measurement resolution under the constraints of camera resolution and indenter size. Third, 2-DoF joint pose and force datasets of the manipulator under external loads and no load are simultaneously collected in different postures. Through this design, datasets of force and pose data are collected for model training. Finally, data-driven models are developed, leveraging deep learning to map the gel deformation to the joint pose and contact forces.

A. Hardware Description

The Tac6D sensor is mounted on the SJTU-Snake III manipulator. As illustrated in Fig. 2(a), Tac6D comprises four modules: 1) image acquisition; 2) illumination; 3) tactile perception; and 4) external support. The image acquisition module employs a camera (800× 600 pixels, 42° field of view, resolution $\Delta w = 0.021$ mm/pixel) to capture gel surface deformation. The illumination module features tri-color LEDs (red, green, blue) arranged concentrically with the gel layer to enhance indenter visibility while minimizing reflections. The tactile perception module consists of an acrylic support layer and a PDMS-0030 gel coated with silver graphite, which provides environmental isolation and deformation information. Since the gel layer is sufficiently soft, it accurately reflects the deformation applied by the indenters. The hyperelasticity and hysteresis of the PDMS have minimal influence on sensing fidelity. To protect the soft gel under high cable tensions, a sandwich structure is formed by the gel layer, a stainless-steel spring, and indenters [Fig. 2(b)]. This design mechanically decouples load-bearing from sensing. The spring absorbs the majority of cable tension, shielding the gel from direct loading and maintaining optimal deformation range for accurate sensing. Moreover, the springs can be freely replaced based on the required cable preload force (from 60 to 300 N), allowing the structure to adapt to different pretension demands within the joint. The external support module, fabricated from carbon-fiber-reinforced Poly(Lactic Acid) (PLA-CF), provides structural rigidity and environmental insulation . Detailed component specifications and BOMs are summarized in Table II.

B. Parametric Indenter Design Method

Although Tac6D captures abundant contact information via gel deformation images, its measurable range and resolution are bounded by the indenter size D [Fig. 2(b)], maximum press depth $l_{\max}$, and camera resolution Δw. For SJTU-Snake III, the design targets are: maximum measurable angle $\theta_{\max} \geq 10°$ and angular resolution $\Delta\theta \leq 0.3°$.

Authorized licensed use limited to: Shanghai Jiaotong University. Downloaded on November 05,2026 at 09:44:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE II
COMPONENT SPECIFICATIONS AND BOMS ANALYSIS OF TAC6D

Component	Material/Type	Mass (g)	Cost (USD)
External support	PLA-CF	51.14	1.18
Camera module	–	5.00	2.89
LED strip	WS2812 5050	4.00	0.64
Spring	stainless steel	25.29	1.94
Acrylic layer	PMMA	2.84	0.01
Gel layer	PDMS-0030	0.0024	<0.01
Total	**–**	**88.27**	**6.66**

BOMs: Bill of Materials. Cost reflects laboratory-scale prototyping.

TABLE III
DATA COLLECTION STATISTICS FOR EACH JOINT POSE

Condition	Samples per Pose	Total Samples
Side contact[a]	90	10 890
Top contact[b]	10	1210
Unloaded	25	3025
Data Volume	**125**	**15 125**

a: 9 points × (1 light press + 1 heavy press) × 5 samples.
b: 1 point × (1 light press + 1 heavy press) × 5 samples.

The $\theta_{\max}$ is determined by D and $l_{\max} = 2$ mm. At the initial position, all four indenters press the gel at depth $l_{\text{initial}} = 1$ mm with 5N cable pretension and no external load. Exceeding $l_{\max}$ during rotation may rupture the gel. Assuming rotation about the red point in Fig. 2(b), $\theta_{\max}$ is derived as $\theta_{\max} = \arcsin\frac{2(l_{\max}-l_{\text{initial}})}{D}$. Results are illustrated in Fig. 2(c). By comparing the relationship between D and $\theta_{\max}$ under different $l_{\max}$ conditions, $D = 8.66$ mm is ultimately determined as the desired distance between indenters.

The $\Delta\theta$ depends on Δw, D, and the maximum slope $k_{\max} = \tan\theta_k$ of the indenter contour. The camera resolution $\Delta w = 0.021$ mm/pixel is calibrated by pressing a 5-mm cube into the gel (observed as 243 pixels). For $\theta_{\max} = 10°$, the indenter cross section is approximated as circular with vertical pressing. The angular resolution is calculated as $\Delta l = \Delta w \tan\theta_k, \Delta\theta = \arcsin(\frac{2\Delta l}{D})$, where θ_k corresponds to $k_{\max}$. As shown in Fig. 2(d), $k_{\max} = 0.84$ is selected to achieve the target resolution.

C. Data Collection Methodology

1) Data Requirements: The gel deformation is captured via the camera. As illustrated in Fig. 2(b), the perception of joint pose and contact forces requires establishing the complex mapping between gel deformation and the two physical quantities. The name "Tac6D" refers to its 6-D sensing target: two rotational angles (α, β), the pressing depth d, and three-axis tip-contact force $\mathbf{F} \in \mathbb{R}^3$. However, d remains nearly constant under quasi-static conditions. Specifically, the gel thickness is only 3 mm, which is significantly less than the joint length. Therefore, it can be assumed that the indenter pressing depth remains essentially constant. This assumption is further justified by the stiffness of the spring ($k_s = 4$ N/mm) and the gel layer ($k_g = 0.126$ N/mm), where the gel is approximately 32 times softer. Under quasi-static conditions, a 1-N cable tension fluctuation produces only 0.2-mm indenter displacement. The validated outputs are 2-DoF angular pose and three-axis tip-contact force. Consequently, the data acquisition process primarily focuses on obtaining contact force data and two rotation angles. The current tip-contact force perception framework assumes contact at the distal end-plate [diameter: 30 mm, Fig. 3(a)]. For hyper-redundant manipulators, contacts mainly occur at the peripheral edge rather than the interior of the end-plate; hence, sampling nine points along the 30-mm circular boundary sufficiently covers practical contact scenarios. For arbitrary body contact, decoupling contact force from moment arm would require additional distributed sensing modalities for contact localization. This will be addressed in our future studies.

2) Data Acquisition Process: A calibrated six-axis F/T transducer (F/T sensor of Flexiv Rizon 4, accuracy 0.03 N) provides ground-truth force measurements [Fig. 3(a)]. Joint pose is measured using IMU sensors (Xsens MTi-630-8A1G6) with onboard sensor fusion. An optical motion capture system (Prime13, OptiTrack, USA) provides ground-truth calibration at initial pose configurations, while the IMU tracks relative angular changes during contact events. The IMU is reinitialized between pose configurations to ensure relative measurements from known reference states since OptiTrack markers are easily obscured. Compared to using OptiTrack alone, the angular error when using a combination of IMU and OptiTrack is less than 0.2°. The data acquisition proceeds in two stages. First, the manipulator is positioned at initial angles $\alpha, \beta \in [-10°, -8°, \ldots, 8°, 10°]$ along both axes with 2° step size (121 poses total). Second, the F/T sensor presses against the manipulator tip from nine uniformly distributed points on the peripheral edge (angular interval 40°) and 1 point at the center. This causes the joint to rotate under the combined action of cable tension and spring elasticity. The resulting tactile images, joint angles, and contact forces ($F_x, F_y, F_z \in [0, 10]$ N) are recorded simultaneously. This sampling strategy has three advantages. Mechanical compliance introduces random angular perturbations under external forces, densifying the discrete 2° pose grid in $(\alpha, \beta, \mathbf{F})$ space for regularization. The circular contact pattern covers all contact directions, with the central point as the axial case. Although the pressing depth d is assumed constant, contact force variations during acquisition induce small fluctuations in d, improving robustness to depth changes. Noncontact events are collected at 5× the contact conditions to reflect normal operation (Table III). The force distribution is shown in Fig. 3(b). For data splitting, 2 of the 20 contact conditions per pose are randomly held out as the test set, ensuring unseen contact locations under all trained poses.

D. Pose and Force Data-Driven Modeling

Estimating joint pose and contact forces requires extracting distinct features from tactile images. Physically, the 2-DoF pose is encoded by indenter rotation, which changes the indenter contour size. The tangential forces (F_x, F_z) additionally depend on tangential sliding of the indenters on the gel, producing microstrain patterns distinct from contour changes. A shared

Authorized licensed use limited to: Shanghai Jiaotong University. Downloaded on November 05,2026 at 09:44:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

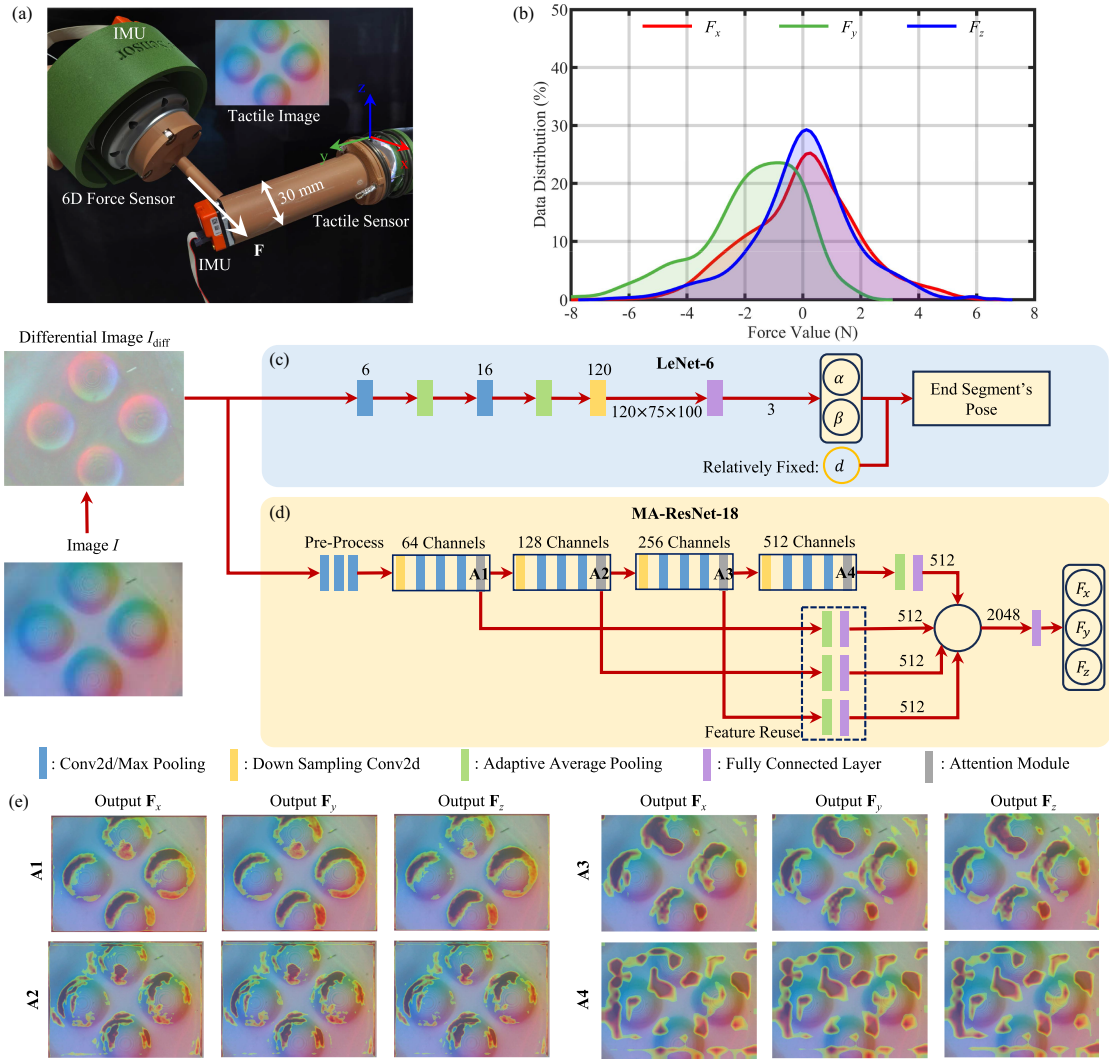


Fig. 3. (a) Experimental snapshot and (b) data distribution of data collection process. The collected data are used to train (c) the pose estimation network (LeNet-6) and (d) the force estimation network (MA-ResNet-18). (e) ScoreCAM heatmaps are generated to visualize the activation distribution of the A1 (64 channels), A2 (128 channels), A3 (256 channels), and A4 (512 channels) layer for each force component prediction.

backbone could misuse these sliding features for pose estimation, causing systematic angle errors. The axial force F_y is coupled with pressing depth, which remains approximately 1 mm under 5 N pretension (gel thickness 3 mm) and is assumed constant. Therefore, a decoupled dual-network architecture is adopted. This design prevents cross-task interference, as the simultaneously collected pose and force data are inherently correlated.

For the 2-DoF pose estimation, a shallow LeNet architecture is employed to balance computational efficiency with representational capacity. To isolate critical contour information and minimize environmental noise, input images are preprocessed as difference maps ($I_{\text{diff}} = I - I_{\text{bg}}$). The network is trained within the PyTorch framework, where the mean squared error loss is minimized via the Adam optimizer with a learning rate of 10^{-5}. In addition, training convergence is accelerated by utilizing automatic mixed precision on an NVIDIA RTX 3060 GPU. Quantitative evaluation on the held-out test set demonstrates high precision, achieving mean absolute errors (MAEs) of 0.11° and 0.06° for z-axis and x-axis rotation, respectively.

Triaxial force estimation is performed using a modified attention ResNet-18 (MA-ResNet-18) architecture, as illustrated in Fig. 3(d). To enhance sensitivity to subtle tactile deformations, a convolutional block attention module [33] is integrated following each convolutional module. Feature maps from residual blocks are concatenated via shortcut transitions, providing a high-dimensional fused representation to the terminal fully connected layer for force regression. ScoreCAM heatmaps [Fig. 3(e)] show that shallow layers (A1, A2) concentrate on indenter edges while deeper layers expand toward contour regions, capturing subtle deformation changes. Although A1–A3 show spatially similar activations, deeper channels encode higher level representations (e.g., from raw edges to contour semantics) of the same regions and feature-reuse reduces MAE by 21% (Table IV). The detailed performance is presented in Section IV-B.

Authorized licensed use limited to: Shanghai Jiaotong University. Downloaded on November 05,2026 at 09:44:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE IV
PERFORMANCE COMPARISON OF DIFFERENT NETWORKS

Network	MAE (N)	σ (N)	R^2	Inference Time (ms)
LeNet-6 (RGB†)	0.166	0.223	0.987	2.2
LeNet-6 (diff‡)	0.181	0.212	0.990	2.2
ResNet-50 (RGB†)	0.236	0.408	0.947	6.8
ResNet-50 (diff‡)	0.242	0.270	0.979	6.8
ResNet-18 (RGB†)	0.280	0.388	0.946	3.0
ResNet-18 (diff‡)	0.281	0.334	0.952	3.0
M*-ResNet-18 (diff‡)	0.221	0.185	0.982	5.4
MA$^\#$-ResNet-18 (diff‡)	0.085	0.052	0.996	5.6
PointNet (diff‡) [34]	1.315	1.094	-0.835	1.5
VGG-16 (diff‡) [35]	0.171	0.284	0.982	4.1

† RGB: Input the tactile image I.
‡ diff: Input the differential tactile image $I_{\text{diff}} = I - I_{\text{bg}}$
* M: Modified net without attention layers.
$^\#$ MA: Modified net with attention layers (Ours).

E. Integration and Perception

To enable physical interaction, the Tac6D unit is embedded as the distal articulated segment of a hyper-redundant manipulator. While proximal joints are monitored by Hall-effect sensors, the Tac6D provides terminal feedback: the distal pose $(\alpha_{\text{end}}, \beta_{\text{end}})$ and the 3-DoF tip-contact force $\mathbf{F}$. The cable pretension is maintained at a nominal level by the low-level controller [3] (approximately 5 N per cable, 15 N total). Therefore, the force estimation network trained under this condition operates reliably within the controlled pretension regime. When the pretension level is changed, a new force estimation network needs to be trained at the new pretension to account for the different gel deformation characteristics. The synchronized multimodal signals drive the force-configuration hybrid controller in [3]. The SJTU-Snake III adopts a hierarchical architecture: an industrial PC with an NVIDIA RTX 3060 GPU performs neural inference and motion planning, while embedded microcontrollers serve as motor drivers. The end-to-end latency of 29.3 ms permits a 30-Hz control loop with adequate timing margin. The manipulator operates quasi-statically. Step-response tests under 5° excitation yield a 0.65-Hz closed-loop bandwidth, consistent with this regime. Given the large length-to-diameter ratio and distal inertia, the per-frame angular displacement is typically $< 0.5°$, so tactile images acquired during closed-loop operation remain within the distribution of the quasi-static training dataset (Section III-C).

IV. VALIDATION EXPERIMENTS

To systematically validate the sensing precision and control efficacy of Tac6D, validation experiments are conducted. First, ablation studies are conducted to validate the performance of our data-driven models. Next, a closed-loop trajectory tracking experiment is conducted to evaluate the accuracy of the joint pose estimation. Then, single-point contact experiments are performed to evaluate Tac6D joint's force prediction accuracy under different joint poses. Finally, a precision peg-in-hole assembly task is executed on the SJTU-Snake III to demonstrate the sensor's capability in enabling robust force-guided manipulation for hyper-redundant manipulators.

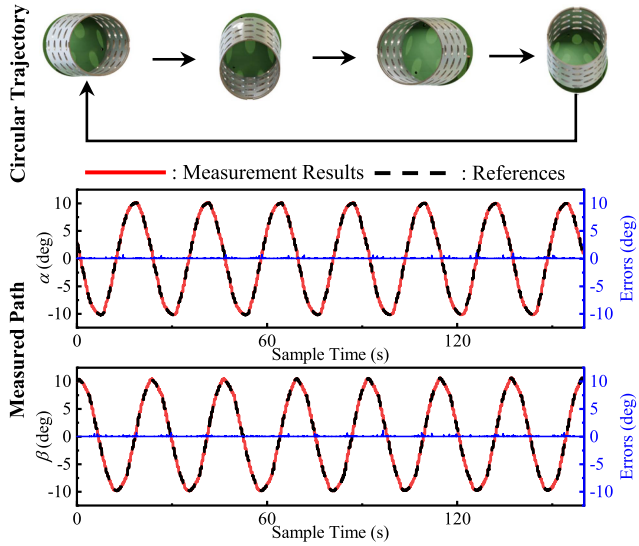


Fig. 4. Trajectory tracking performance of the Tac6D. (top) Schematic of the reference circular trajectory. (bottom) Measurement results and errors for angles. α and β denote the rotation about the z-axis and x-axis, respectively.

A. Performance Validation of the Data-Driven Model

Performance validation experiments are conducted to validate the efficacy of the proposed MA-ResNet-18 structure, employing the same hyperparameters and training protocol defined in Section III-D. The predictive performance is quantified on the test set $\mathcal{D}_{\text{test}}$ using the MAE, the error standard deviation (σ), and the coefficient of determination (R^2), evaluated per axis $k \in \{x, y, z\}$ with true and predicted forces $F_{k,i}$ and $\hat{F}_{k,i}$. Table IV compares MA-ResNet-18 with representative architectures for force estimation, including PointNet and general-purpose CNNs (VGG-16, ResNet-18/50, LeNet-6). Using the differential image I_{diff} visibly stabilizes σ, suppressing error variance and improving robustness to static background noise. The comparison reveals two trends. The shallow LeNet-6 outperforms the deep ResNet-50, which highlights the importance of low-level geometric features. Yet ResNet-50 still exceeds ResNet-18, which indicates that a certain depth remains necessary for complex photometric modeling. To reconcile these requirements, M-ResNet-18 introduces feature-reuse shortcuts that recover fine-grained spatial details lost in deeper layers. MA-ResNet-18 further adds attention modules to prioritize task-relevant regions across feature scales. As a result, it attains the lowest MAE (0.085 N) at an inference time of 5.6 ms.

B. Evaluations of Pose Estimation

To evaluate the pose estimation accuracy of the Tac6D joint, the hyper-redundant manipulator's end segment is commanded to follow a circular trajectory defined by $\alpha(t_i) = 10\sin(\frac{4\pi}{3}t_i)$ deg and $\beta(t_i) = 10\cos(\frac{4\pi}{3}t_i)$ deg, where $t_i = \{0.03, 0.06, \cdots\}$ s, and α and β denote the rotation angles about the z-axis and x-axis, respectively. The manipulator operates in a quasi-static motion pattern. Joint poses are estimated from tactile images

Authorized licensed use limited to: Shanghai Jiaotong University. Downloaded on November 05,2026 at 09:44:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

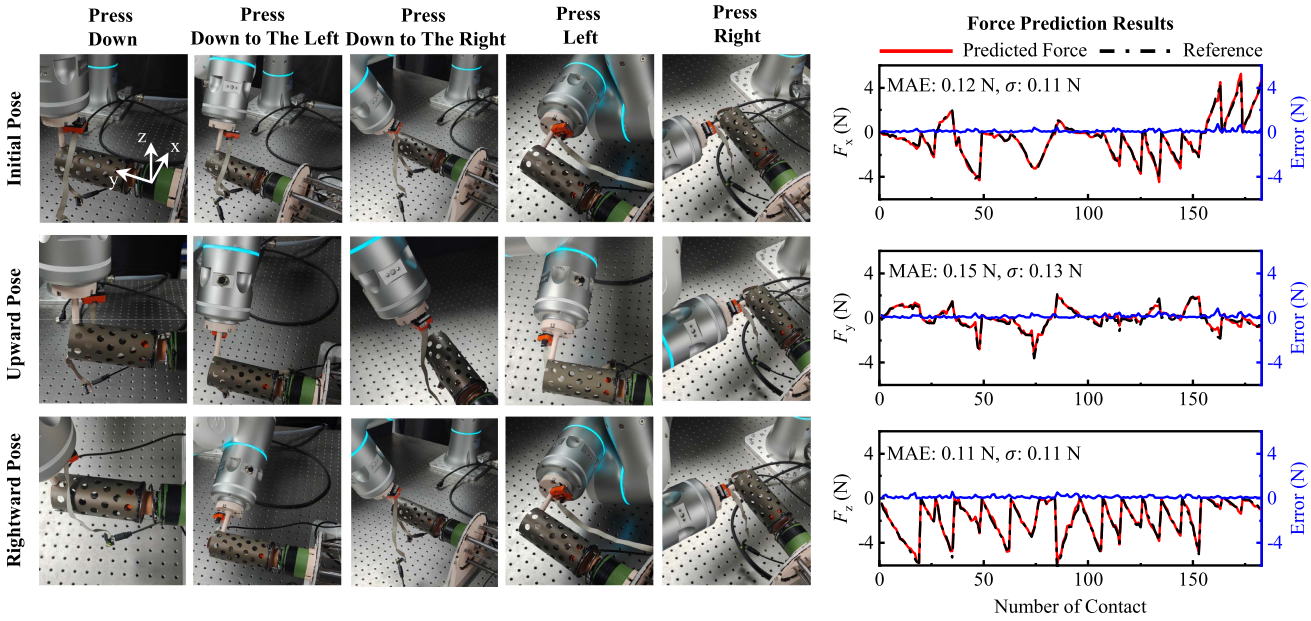


Fig. 5. (left) Experimental snapshots and (right) force prediction results of the contact force perception experiment in (top) the initial pose, (middle) the upward pose and (bottom) the rightward pose.

using the LeNet-6 network. As shown in Fig. 4, ground-truth measurements are obtained via an optical motion capture system (Prime13, OptiTrack, USA). The results show a maximum absolute error of 0.66° and an MAE of 0.04°. Notably, peak deviations occur near the directional reversal points (i.e., the quadrant transitions of the circular trajectory). This phenomenon is primarily attributed to the nonlinear friction and mechanical hysteresis inherent in the cable-driven mechanism. Specifically, during direction changes, the stick-slip effect between the cables and guiding holes induces a momentary kinematic lag, resulting in unmodeled discrepancies. Future work will prioritize friction compensation strategies and structural optimization to mitigate these nonlinearities.

A fatigue test is further conducted to verify that the spring effectively relieves the gel from sustained compressive loads. With all three cables pretensioned to 10 N, the joint continuously follows the same circular trajectory for 9 h without any gel degradation or failure, confirming the effectiveness of the indenter-spring-gel architecture in decoupling load-bearing from contact sensing and ensuring long-term structural reliability. For higher cable tensions, stiffer springs can be substituted to maintain the same indenter pressing depth and thus comparable fatigue life.

C. Evaluations of Force Estimation

Force sensing is first validated via contact experiments at the distal end-plate (Fig. 5). A collaborative arm (Flexiv Rizon 4S) applies linear strokes at five spatial points under three joint poses $(\alpha, \beta) \in \{(0°, 0°), (0°, 8°), (8°, 0°)\}$. Ground-truth forces are recorded by a six-axis F/T sensor and transformed into the Tac6D frame via an IMU (Xsens MTi-630-8A1G6). The MAE stays within 2.5% of the full-scale range across all force components. The maximum error occurs in F_y because the sandwich architecture is stiffer in the y-direction, yielding smaller gel deformations and less prominent visual features.

TABLE V
FORCE ESTIMATION UNDER DIFFERENT CABLE PRETENSIONS

$3T_b$ (N)	$\pm\lvert F\rvert_{\max}$ (N)	Sensitivity[a]	MAE (N)[b]	R^{2b}
15	10.5	0.966	0.083	0.993
30	12.5	0.978	0.066	0.996
60	17.9	0.985	0.076	0.999
360[c]	16.5	0.970	0.126	0.997

[a] Pooled least-squares slope of predicted versus ground-truth force across the three axes.
[b] Each row uses a separately trained network at the corresponding pretension.
[c] Stiffer-spring configuration (cable pretension 120 N).

Then, to quantitatively verify that the Tac6D structure preserves sensing sensitivity under varied cable loads, the tip-contact force experiment is repeated at three baseline cable pretensions $T_b \in \{5, 10, 20\}$ N. The measured mean cable tensions are 4.5, 10.0, and 14.5 N, with peaks up to 37 N during contact. For each pretension, a separate MA-ResNet-18 model is trained using data collected at the corresponding T_b. The summed cable tension $3T_b$, peak absolute contact forces $\pm|F|_{\max}$, sensitivity, pooled MAE, and R^2 are summarized in Table V. The results confirm that the Tac6D architecture preserves sensing sensitivity across different cable loads. The absolute MAE remains nearly constant at 0.07–0.08 N while the operating range expands with T_b. The relative full-scale error decreases from 1.42% at 5 N to 0.73% at 20 N. The sensitivity remains close to unity across all conditions (0.966–0.985). This indicates that the indenter-spring-gel structure maintains faithful force-to-deformation mapping even at higher cable tensions, as the spring absorbs most of the tension and shields the gel from excessive compression. A current limitation is that the network

Authorized licensed use limited to: Shanghai Jiaotong University. Downloaded on November 05,2026 at 09:44:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

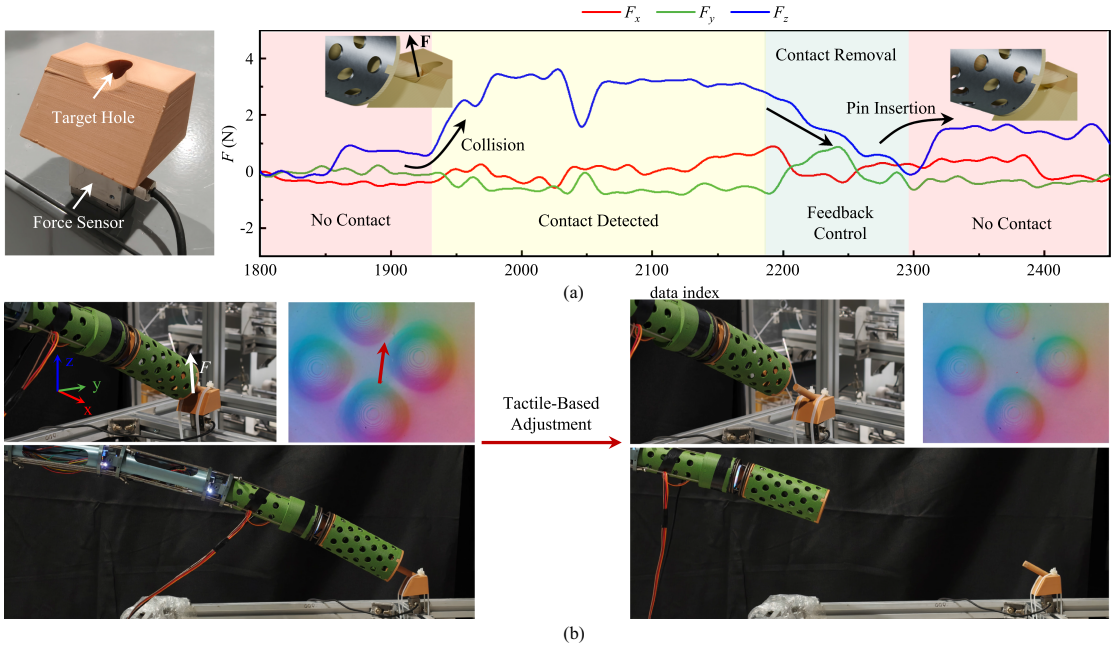


Fig. 6. Peg-in-hole assembly with end-contact force feedback using the SJTU-Snake III. (a) (left) Target hole and (right) information on the prediction of contact forces during the contact process. (b) Snapshots of the insertion process.

trained at one pretension does not directly generalize to another without retraining. Changes in T_b alter the gel deformation characteristics. This limitation primarily affects the force estimation network (MA-ResNet-18) whose input depends on absolute gel deformation. The pose estimation network (LeNet-6) does not require retraining after spring replacement. Only a constant angle bias arises from minor variations in initial indenter pressing depth during spring installation. Subtracting this bias restores the original pose accuracy. Spring substitution is further verified with a stiffer-spring configuration (cable pretension 120 N) using the same data acquisition method (Section III-C) and a separately trained network. The resulting sensitivity (0.970), MAE (0.126 N), and R^2 (0.997) remain comparable to the low-stiffness-spring cases (Table V). This confirms the scalability of the proposed architecture across different spring stiffnesses. Improving cross-pretension generalization will be addressed in future work. Videos are provided in the Supplementary Material.

D. Peg-in-Hole Assembly With End-Contact Force Feedback

As shown in Fig. 6, a precision peg-in-hole experiment is conducted on the SJTU-Snake III platform to evaluate the operational efficacy of the tactile-integrated end joint. The 10-mm pin and 10.2-mm hole yield a 0.2-mm tolerance, with insertion performed at 3 mm/s to maintain quasi-static contact. This speed is typical for precision assembly with hyper-redundant manipulators, whose low stiffness inherently limits operational speeds. Higher speeds would degrade sensing accuracy due to PDMS hysteresis and violation of the quasi-static training assumption. In hyper-redundant mechanisms, cumulative kinematic errors and mechanical vibrations render joint-angle feedback alone insufficient for such high-precision assembly. To mitigate these discrepancies, a tactile-guided control strategy is implemented. During free-space motion, the tactile sensor provides pose feedback (α, β) for closed-loop control. Upon detecting contact forces exceeding a preset threshold, the control logic switches to force minimization mode, actuating the end joint opposite to the measured contact force until lateral forces approach zero. To quantify the advantage of distal tactile feedback, the task is repeated 15 times per condition with randomized initial lateral misalignment of 5–15 mm. The baseline uses a rigid joint (no Tac6D) controlled solely by proximal cable tension via Newton–Euler estimation [3], without distal force feedback. Tac6D achieves a 93.3% success rate (14/15) with an average insertion time of 9.2 s and contact force of 0.82 N (std: 0.74 N, max: 2.34 N), versus 46.7% (7/15), 11.3 s, and 1.72 N (std: 0.74 N, max: 2.64 N) for the baseline. Failed baseline attempts average 4.36 N (max: 6.13 N) due to uncorrected misalignment. A two-sample t-test confirms the difference is statistically significant ($p = 0.023 < 0.05$). The 2× improvement in success rate and 52% reduction in contact force demonstrate that distal tactile feedback is critical for precision assembly with hyper-redundant manipulators. Videos are provided in the Supplementary Material.

E. Structural Reliability Validation of Multiple Tac6D Joints

To verify structural reliability within the kinematic chain, a manipulator with three Tac6D joints is designed. Since proximal joints bear internal loads exceeding 200 N from both cable tensions and distal segment support, stiffer springs are employed for these positions [Fig. 7(a)]. The quick-replaceable sandwich architecture allows spring substitution without major design modifications, adapting to varying joint loads.

As shown in Fig. 7(b), the manipulator is commanded to cycle through horizontal, upward, and downward configurations

Authorized licensed use limited to: Shanghai Jiaotong University. Downloaded on November 05,2026 at 09:44:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

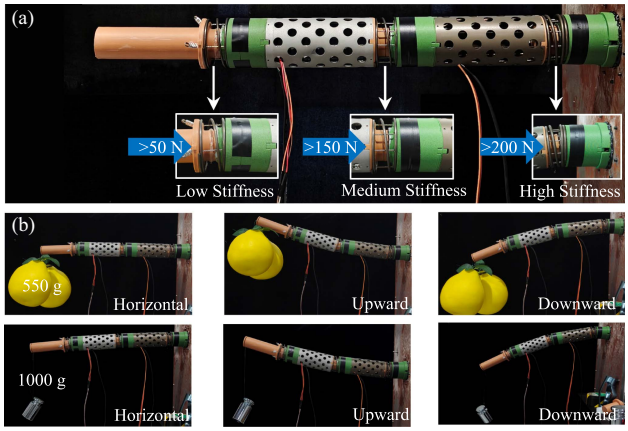


Fig. 7. (a) Hyper-redundant manipulator with multiple Tac6D joints. The closer the spring is to the base, the greater its stiffness. (b) Repeatability evaluation under horizontal, upward, and downward configurations with 550 and 1000 g payloads.

30 times under payload, with end-point positions recorded to evaluate repeatability. Under both 550 and 1000 g payloads, the repeatability error remains consistent at $\pm$ 1 mm. At 1000-g payload, the vertical deflection is 25 mm, compared to 10 mm for the original SJTU-Snake III with three rigid joints, indicating a 2.5$\times$ increase in compliance-induced deflection. While the preserved repeatability demonstrates that Tac6D joints maintain structural integrity, several scalability challenges remain. The softer spring-based joints increase the vertical deflection and limit payload capacity. Each Tac6D joint requires one camera cable and three LED power lines, scaling linearly with joint count. Networks trained for one joint do not directly transfer to others, requiring per-joint data collection of approximately 4 h. Concurrent image processing for three joints reduces the frame rate from 30 Hz to approximately 10 Hz. This rate remains sufficient for quasi-static control, as the sampling-to-bandwidth ratio of 15:1 (10 versus 0.65 Hz bandwidth) well exceeds the guideline [36], and the corresponding zero-order-hold phase lag ($\approx 12^\circ$) preserves stability margins. Future work will address cable routing optimization, transfer learning to reduce calibration overhead, and embedded GPU deployment for parallel inference.

V. Conclusion

This article presents Tac6D, a camera-based tactile smart joint that enables simultaneous pose and tip-contact force perception for hyper-redundant manipulators. The main contributions are as follows. First, an integration approach embeds camera-based tactile sensing into compact joint spaces without compromising the slender form factor. Second, an "indenter-spring-gel" sandwich architecture balances load-bearing capacity with sensing sensitivity, withstanding internal joint loads up to 60 N while maintaining angular resolution below 0.3°. Third, multifeature fusion networks extract pose and force from gel deformation images, achieving mean errors of 0.11° for pose and 0.085 N for tip-contact force. Peg-in-hole assembly experiments confirm that Tac6D provides reliable pose and tip-contact force feedback for autonomous trajectory adjustment and precision operations, validating the effectiveness of the proposed sensing solution. Current limitations include the constrained angular range of $\pm 10^\circ$, the restriction to tip-contact force perception at the distal end-plate, the need to retrain the network when the cable pretension level is changed, and the focus on a single sensor unit. Future work will address these through structural optimization, cross-pretension generalization strategies, and development of modular multijoint sensing systems for full-body proprioception.

References

[1] M. Luo, E. Li, A. Zhang, M. Tan, and Z. Liang, "A bioinspired coiled cable-driven manipulator: Mechatronic design and kinematics planning with multiconstraints," *IEEE/ASME Trans. Mechatron.*, vol. 28, no. 6, pp. 3155–3166, Dec. 2023.
[2] Z. Yang et al., "Design and validation of cable-driven hyper-redundant manipulator with a closed-loop puller-follower controller," *Mechatronics*, vol. 78, 2021, Art. no. 102605.
[3] Z. Zhu, Z. Li, Z. Peng, C. Liu, and G. Gu, "Hybrid tension and configuration control of cable-driven hyper-redundant robots for high accuracy and stability," *IEEE Robot. Automat. Lett.*, vol. 10, no. 6, pp. 5601–5608, Jun. 2025.
[4] Y. Mao et al., "A cable-driven hyper-redundant robot with angular sensing," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, 2021, pp. 1294–1299.
[5] X. Wang et al., "Learning-based visual-strain fusion for eye-in-hand continuum robot pose estimation and control," *IEEE Trans. Robot.*, vol. 39, no. 3, pp. 2448–2467, Jun. 2023.
[6] Z. Yang, L. Yang, Y. Sun, and X. Chen, "A novel contact-aided continuum robotic system: Design, modeling, and validation," *IEEE Trans. Robot.*, vol. 40, pp. 3024–3043, 2024.
[7] A. Gao, N. Liu, M. Shen, M. E. M. K. Abdelaziz, B. Temelkuran, and G.-Z. Yang, "Laser-profiled continuum robot with integrated tension sensing for simultaneous shape and tip force estimation," *Soft Robot.*, vol. 7, no. 4, pp. 421–443, 2020.
[8] S. Choi, Y.-K. Moon, J. Kim, and K. Kim, "Dnn-based force estimation in hyper-redundant manipulators," *Int. J. Precis. Eng. Manuf.*, vol. 25, pp. 2111–2123, 2024.
[9] W. Li, X. Huang, L. Yan, H. Cheng, B. Liang, and W. Xu, "Force sensing and compliance control for a cable-driven redundant manipulator," *IEEE/ASME Trans. Mechatron.*, vol. 29, no. 1, pp. 777–788, Feb. 2024.
[10] S. Tang, Z. Liang, Z. Zhang, X. Duan, H. Wang, and G. Chen, "A redundant parallel continuum manipulator with stiffness-varying and force-sensing capability," *IEEE Trans. Automat. Sci. Eng.*, vol. 22, pp. 3635–3647, 2025.
[11] J. Ren et al., "GelLight: Illumination design, modeling, and optimization for camera-based tactile sensor," *Adv. Intell. Syst.*, vol. 7, no. 5, 2025, Art. no. 2400596.
[12] C. Zhao, J. Liu, and D. Ma, "iFEM2.0: Dense 3-D contact force field reconstruction and assessment for vision-based tactile sensors," *IEEE Trans. Robot.*, vol. 41, pp. 289–305, 2025.
[13] L. V. Duong, "BiTac: A soft vision-based tactile sensor with bidirectional force perception for robots," *IEEE Sensors J.*, vol. 23, no. 9, pp. 9158–9167, May 2023.
[14] W. Du, W. Xu, J. Ren, Z. Yu, and C. Lu, "TacIPC: Intersection- and inversion-free FEM-based elastomer simulation for optical tactile sensors," *IEEE Robot. Automat. Lett.*, vol. 9, no. 3, pp. 2559–2566, Mar. 2024.
[15] K. Althoefer, Y. Ling, W. Li, X. Qian, W. W. Lee, and P. Qi, "A miniaturised camera-based multi-modal tactile sensor," in *Proc. 2023 IEEE Int. Conf. Robot. Automat.*, 2023, pp. 12570–12575.
[16] F. Manara et al., "Tendon-driven hyper-redundant manipulators: A review on design solutions," *IEEE/ASME Trans. Mechatron.*, vol. 30, no. 5, pp. 3852–3867, Oct. 2025.
[17] A. Poka et al., "Alternative solutions to tendon-driven hyper-redundant manipulators: A review on design and actuation strategies," *IEEE/ASME Trans. Mechatron.*, vol. 30, no. 5, pp. 3938–3952, Oct. 2025.
[18] M. Elsherif et al., "Optical fiber sensors: Working principle, applications, and limitations," *Adv. Photon. Res.*, vol. 3, no. 11, 2022, Art. no. 2100371.
[19] H. Wei, Y. Zheng, and G. Gu, "RRT-based path planning for follow-the-leader motion of hyper-redundant manipulators," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 3198–3204.
[20] X. Chen, X. He, Y. Zhang, and B. Gao, "Angle measurement method based on cross-orthogonal hall sensor," in *Proc. 16th IEEE Int. Conf. Electron. Meas. Instrum.*, 2023, pp. 153–158.

Authorized licensed use limited to: Shanghai Jiaotong University. Downloaded on November 05,2026 at 09:44:49 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[21] C. Wang, H. Xie, and H. Yang, "An iterative path-following method for hyper-redundant snake-like manipulator with joint limits," *Ind. Robot, Int. J. Robot. Res. Appl.*, vol. 50, pp. 505–519, 2023.
[22] J. Peng, C. Zhang, D. Meng, and B. Liang, "Trajectory optimization methods of a space hyper-redundant robot based on effective arm-shape measurement," *IEEE Trans. Instrum. Meas.*, vol. 72, 2023, Art no. 5017914.
[23] X. Xu, H. Xie, C. Wang, and H. Yang, "Kinematic and dynamic models of hyper-redundant manipulator based on link eigenvectors," *IEEE/ASME Trans. Mechatron.*, vol. 29, no. 2, pp. 1306–1318, Apr. 2024.
[24] Z. Li, J. Deng, Y. Zheng, C. Liu, and G. Gu, "Cable length estimation for a hyper-redundant robot based on cable hole clearance calculation," in *Proc. IEEE 19th Int. Conf. Automat. Sci. Eng.*, 2023, pp. 1–7.
[25] H. Li, Y. Lin, C. Lu, M. Yang, E. Psomopoulou, and N. F. Lepora, "Classification of vision-based tactile sensors: A review," *IEEE Sensors J.*, vol. 25, no. 19, pp. 35672–35686, Oct. 2025.
[26] Q. K. Luu, D. Q. Nguyen, N. H. Nguyen, N. P. Dam, and V. A. Ho, "Vision-based proximity and tactile sensing for robot arms: Design, perception, and control," *IEEE Trans. Robot.*, vol. 41, pp. 5000–5019, 2025.
[27] T. T. Nguyen, Q. K. Luu, D. Q. Nguyen, and V. A. Ho, "Contac: Continuum-emulated soft skinned arm with vision-based shape sensing and contact-aware manipulation," in *Proc. Conf. Robot. Sci. Syst.*, 2024, pp. 1–12.
[28] B. Fang et al., "Force measurement technology of vision-based tactile sensor," *Adv. Intell. Syst.*, vol. 7, no. 1, 2025, Art. no. 2400290.
[29] T. Zhang, R. Wang, Q. Cao, S. Cui, G. Zheng, and S. Wang, "Flowsight: Vision-based artificial lateral line sensor for water flow perception," *IEEE Trans. Robot.*, vol. 41, pp. 3260–3277, 2025.
[30] Q. Zhang, Z. Zuo, H. Wang, B. Liu, Y. Yilihamu, and L. Wen, "Utact: Underwater vision-based tactile sensor with geometry reconstruction and contact force estimation," *Adv. Robot. Res.*, vol. 2, 2025, Art. no. e202500091.
[31] Y. Zhang et al., "Multi-branch multi-scale channel fusion graph convolutional networks with transfer cost for robotic tactile recognition tasks," *IEEE Trans. Automat. Sci. Eng.*, vol. 22, pp. 11856–11868, 2025.
[32] M. Huang et al., "Two-stream network-driven vision-based tactile sensor for object feature extraction and fusion perception," 2025, *arXiv:2510.12528*.
[33] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.
[34] J. Wang, M. Yao, Y. Wei, X. Guo, A. Zheng, and W. Zhao, "Image-to-force estimation for soft tissue interaction in robotic-assisted surgery using structured light," *IEEE Robot. Automat. Lett.*, vol. 10, no. 8, pp. 7795–7802, Aug. 2025.
[35] D. Yin et al., "Gelevent—A novel high-speed tactile sensor with event camera," *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, Art no. 9511413.
[36] G. F. Franklin, J. D. Powell, and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, vol. 4. Upper Saddle River, NJ, USA: Prentice Hall, 2002.



Zhanxuan Peng is currently working toward the Ph.D. degree in mechanical engineering with Shanghai Jiao Tong University, Shanghai, China.

His research interests include modeling and control of soft continuum robots.



Jieji Ren received the B.Sc. and M.Sc. degrees in optics from the Harbin Institute of Technology, Harbin, China, in 2013 and 2015, respectively, and the Ph.D. degree in mechatronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2022.

Since 2022, he has been with Shanghai Jiao Tong University as an Assistant Researcher. His research interests include camera-based tactile sensing and its applications in soft robotics.



Zhenpu Zhu is currently working toward the Ph.D. degree in mechanical engineering with Shanghai Jiao Tong University, Shanghai, China.

His research interests include modeling and control of hyper-redundant manipulators.



Jiang Zou (Member, IEEE) received the B.E. degree in mechanical engineering from the University of Science and Technology Beijing, Beijing, China, in 2014, and the Ph.D degree in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2019.

He is currently an Associate Professor with the School of Mechanical Engineering, Shanghai Jiao Tong University. He has authored or coauthored more than 30 publications, which have appeared in *Science Robotics*, *Nature Communications*, *National Science Review*, and IEEE/ASME TRANSACTIONS. His research interests include design, modeling, and control of dielectric elastomer actuated soft robots.



Guoying Gu (Senior Member, IEEE) received the B.E. degree (hons.) in electronic science and technology and the Ph.D. degree (hons.) in mechatronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2006 and 2012, respectively.

He was a Humboldt Fellow with the University of Oldenburg, Oldenburg, Germany, and was a Visiting Scholar with the Massachusetts Institute of Technology, Cambridge, MA, USA; the National University of Singapore, Singapore; and Concordia University, Montreal, QC, Canada. He is currently a Distinguished Professor with the School of Mechanical Engineering, Shanghai Jiao Tong University. He has authored or coauthored more than 150 publications, including articles in *Science*, *Science Robotics*, *Nature Biomedical Engineering*, *Nature Reviews Materials*, *Nature Materials*, *Nature Communications*, *Science Advances*, *Advanced Materials*, *Soft Robotics*, *Science China* series, the IEEE/ASME TRANSACTIONS journals, book chapters, and international conference proceedings. His research interests include soft robotics, bioinspired and wearable robots, and smart materials for sensing, actuation, and motion control.

Dr. Gu was the recipient of the National Science Fund for Distinguished Young Scholars and the XPLORER Prize. He is currently an Associate Editor for *Soft Robotics*, *Soft Science*, and IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, and was an Associate Editor for IEEE TRANSACTIONS ON ROBOTICS and IEEE ROBOTICS AND AUTOMATION LETTERS.

Authorized licensed use limited to: Shanghai Jiaotong University. Downloaded on November 05,2026 at 09:44:49 UTC from IEEE Xplore. Restrictions apply.