

# Reconstructing Soft Robotic Touch via In-Finger Vision

Ning Guo Xudong Han Shuqiao Zhong Zhiyuan Zhou Jian Lin Fang Wan\* Chaoyang Song\*

Ning Guo, Xudong Han

SUSTech Institute of Robotics, Southern University of Science and Technology, Shenzhen, China.

Shuqiao Zhong, Prof. Zhiyuan Zhou, Prof. Jian Lin

Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen, China

Prof. Fang Wan

School of Design, Southern University of Science and Technology, Shenzhen, China.

\*Corresponding Email: wanf@sustech.edu.cn

Prof. Chaoyang Song

Mechanical and Energy Engineering, Southern University of Science and Technology, Shenzhen, China.

\*Corresponding Email: songcy@ieee.org

**Keywords:** *Soft Robotics, Tactile Sensing, Shape Reconstruction, Inverse Problem*

Incorporating authentic tactile interactions into virtual environments presents a notable challenge for the emerging development of soft robotic metamaterials. This study introduces a vision-based approach to learning proprioceptive interactions by simultaneously reconstructing the shape and touch of a Soft Robotic Metamaterial (SRM) during physical engagements. The SRM design has been optimized to the size of a finger with enhanced adaptability in 3D interactions while incorporating a see-through viewing field inside, which can be visually captured by a miniature camera underneath to provide a rich set of image features for touch digitization. Employing constrained geometric optimization, we modeled the proprioceptive process with Aggregated Multi-Handles (AMHs). This approach facilitates real-time, precise, and realistic estimations of the finger's mesh deformation within a virtual environment. We also proposed a data-driven learning model to estimate touch positions, achieving reliable results with impressive  $R^2$  scores of 0.9681, 0.9415, and 0.9541 along the  $x$ ,  $y$ , and  $z$  axes. Furthermore, we have demonstrated the robust performance of our proposed methods in touch-based human-cybernetic interfaces and human-robot collaborative grasping. This study opens the door to future applications in touch-based digital twin interactions through vision-based soft proprioception.

## 1 Introduction

The sense of touch is a fundamental aspect of human interaction with the physical world, complementing our visual perception. Through touch, we gain valuable insights into the properties of objects and environments, enabling us to navigate and interact with the world dynamically [1]. Combining tactile feedback and visual perception is essential for comprehensively understanding the physical world. In human-cybernetic interfaces, touch-based interaction is emerging as a crucial component, connecting human behavior and thinking processes with advanced engineering systems [2]. Recent years have witnessed significant advancements in the field of soft robotics, marked by extensive research efforts towards the enhancement of this technology, including materials science, fabrication techniques, and actuation and sensing methods [3, 4, 5]. These developments are especially pertinent within the human-cybernetic interface and human-cybernetic systems, where the fusion of biological and artificial systems is paramount for multiple sensory-cognitive processing demands. This technology has the potential to revolutionize various industries and provide users with unprecedented levels of engagement and realism.

Soft Robotic Metamaterial (SRM) represents an emerging research field in functional materials, enabling the creation of synthetic structures that possess desired physical properties and embedded intelligence [6, 7, 8, 9]. This convergence, inspired by the intricate sensorimotor system of biological skin, holds immense promise for integrating tactile intelligence into these materials, opening doors for diverse research and real-world applications [10, 11, 12]. However, a significant challenge remains in proprioception due to the substantial deformations these robots can undergo, making traditional sensing and design methods insufficient for precise feedback and control [13, 14]. Various sensing techniques have been integrated into soft robotics to address deformation, force, contact, and temperature monitoring [15, 16, 17, 18]. Yet, accurately identifying contact as the source of sensor responses while integrating material mechanics remains an enduring challenge in the field of robotics [19, 20], epitomizing a classic *inverse problem* within the broader context of mechanical science and engineering [21, 22, 23].

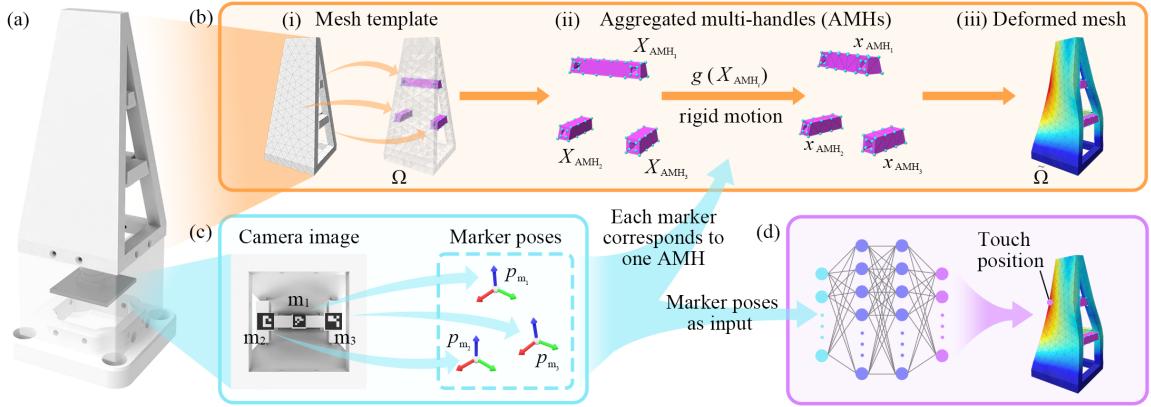
An extensive range of tactile sensors has been engineered [24, 25, 26], embracing several different strategies such as piezoelectric [27], resistance-based [28], and capacitance-based methods [29]. Vision-based tactile sensors have emerged recently as a promising avenue, celebrated for their high resolution and versatile sensing capabilities [30]. These sensors leverage optical techniques to produce rich, detailed data about contact interactions, overcoming many of the limitations of their predecessors [24]. Notable advantages of vision-based systems include their low cost due to the use of off-the-shelf components [31]; easy fabrication, which accelerates development cycles [32]; high durability, ensuring longevity even in challenging operational contexts [33]; and the ability to perform multi-modal measurements [34], capturing a comprehensive picture of tactile interactions [35]. These characteristics position vision-based tactile sensors as a strategic choice for advanced robotic applications requiring nuanced touch interactions [36, 37]. **Table 1** compares design shape, geometric reconstruction method, and deployed camera system between several state-of-the-art vision-based tactile sensors and our proposed sensing system. Our innovation contributes to this diverse ecosystem by marrying the Shape From Template (SFT) methodology with a monocular camera system. This approach facilitates the complex 3D reconstruction of soft robotic metamaterial, setting our work apart.

**Table 1:** A comparison between state-of-the-art vision-based tactile sensors and our work.

Sensors Name	Contact Medium Morphology	Geometric Reconstruction	Camera System
GelSight [38]	Flat Elastomer	Shape from Photometric Stereo	Monocular Camera
GelForce [39]	Flat Elastomer	Shape from Elasticity	Monocular Camera
Feng’s Sensor [40]	Flat Elastomer	Shape from Defocus	Monocular Camera
GeoStereo [35]	Flat Elastomer	Shape from Stereo	Binocular Camera
FingerVision [41]	Flat Elastomer	-	Monocular Camera
Soft-bubble [42]	Dome-Shaped Membrane	Shape from 3D Point Clouds	Depth Camera
TacTip [43]	Dome-Shaped Elastomer	-	Monocular Camera
DenseTact [44]	Sphere-Shaped Elastomer	Shape from Photometric Stereo	Monocular Camera
<b>This work</b>	<b>Soft Robotic Metamaterial</b>	<b>Shape from Template</b>	Monocular Camera

Measuring deformation followed by extracting contact information constitutes a robust methodology for achieving comprehensive contact knowledge and ensuring the reliability of tactile sensors [45]. This strategy allows for a detailed understanding of the tactile dynamics at play, enhancing the sensor’s precision and utility in complex robotic applications [46]. Nevertheless, an alternative paradigm exists within learning-based methods, which offers a distinct approach by directly learning the mapping from raw sensory signals to contact information [47]. Thuruthel et al. [48] used LSTM networks to estimate force and displacement at the soft fingertip using strain sensors. Narang et al. [49] applied variational autoencoders to learn latent representations from the FEM simulation of a multimodal tactile sensor and demonstrated the sim-to-real transfer for real-time applications. Furthermore, Scimeca et al. [50] used a fully connected feed-forward network to reconstruct soft material deformations with tactile arrays. By employing advanced machine learning algorithms, these approaches simplify the process of tactile sensing and open up possibilities for faster, potentially more adaptable sensory feedback mechanisms in robotic systems, thereby broadening the horizons of tactile perception and interaction in automated applications [51]. Incorporating these advancements in sensing and modeling techniques within the human-cybernetic interface and human-cybernetic systems opens the door to more sophisticated interactions, enhanced cognitive processing, and the seamless integration of soft robots into various domains, including healthcare and assistive technologies.

This paper introduces a soft finger integrated with a vision-based proprioception technique for accurate shape and touch estimation during interactive processes, as depicted in **Figure 1**. The design of this finger has been engineered to enhance its adaptability, allowing it to effectively handle bending, twisting, and enveloping actions during interactions. Our approach leverages vision-based proprioceptive sensing to assess shape deformation and identify touch locations precisely. To estimate the shape of the soft finger, we employ constrained geometric optimization to track the poses of ArUco markers [52], which are



**Figure 1: Overview of the vision-based proprioceptive soft finger.** (a) A soft robotic metamaterial with in-finger vision. The markers' poses (c) obtained by the camera are delivered to two models, including (b) a constrained geometric optimization approach using aggregated multi-handle constraints to estimate the shape and (d) a data-driven learning model to estimate touch position.

captured by a monocular camera positioned beneath the finger. In addressing the issue of touch estimation, we theoretically formulate optimal boundary condition estimation under hyperelasticity PDE constraint and propose an adjoint method to resolve surface contact force estimation. Additionally, we developed a data-driven learning model for the soft finger to achieve accurate touch interaction estimation for practical implementation. The robustness and effectiveness of our proposed approach have undergone rigorous evaluation through additional tasks focused on touch-based interaction and intelligent soft grasping. The results of these tasks highlight our approach's reliability and efficacy, demonstrating outstanding proprioceptive sensing capabilities with potential in touch-based human-robot interactions. The contributions of this work are as follows:

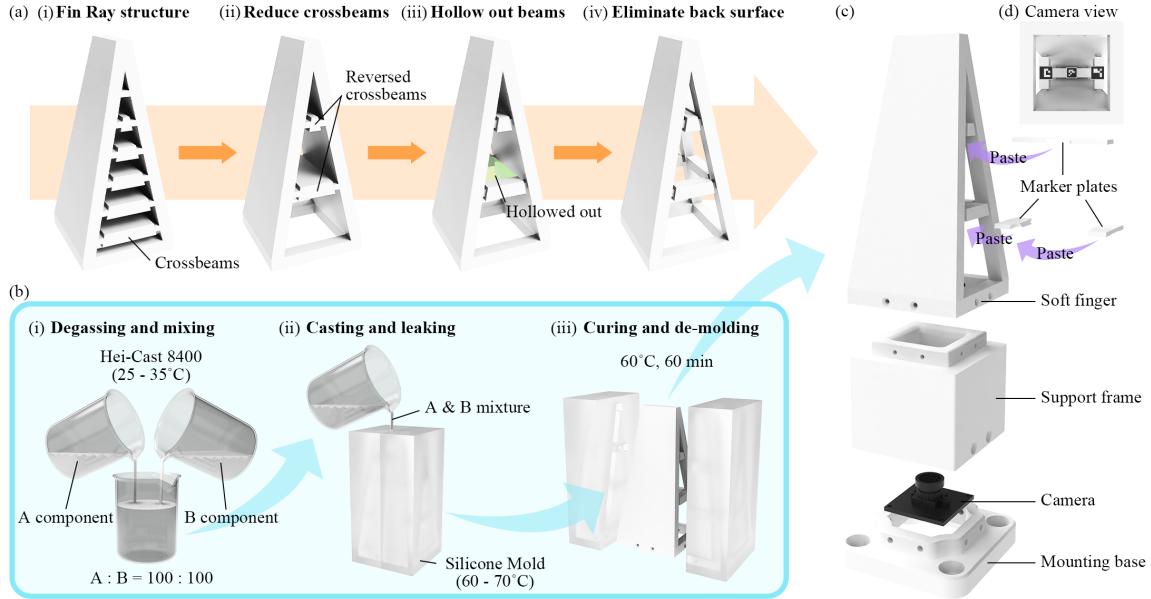
- Developed a soft robotic metamaterial (SRM) finger with a fully enclosed finger surface, featuring a monocular camera for vision-based shape and touch estimation, and introduced a novel approach for accurate shape estimation using geometrically motivated distortion energy minimization.
- Implemented a data-driven method for touch estimation with high accuracy and conducted comprehensive experiments to validate the reliability and robustness of the proposed SRM finger, demonstrating its potential for applications in touch-based robotic grasping.

## 2 Results

### 2.1 Vision-based Soft Finger Design & Fabrication

Taking inspiration from fish fins for fluidic propulsion [53], the Fin Ray Effect (FRE) offers a compliant planar design characterized by passive adaptation, allowing the finger to envelop objects based on geometric contact in the 2D plane [54, 55]. Previous FRE finger designs usually feature an extruded structure based on a 2D triangular cross-section with multiple horizontal crossbeams [56], as shown in **Figure 2(a-i)**. However, this design presents limitations in spatial adaptation, such as twisting and excessive structural complexity.

We initiated several design enhancements to improve the finger's spatial adaptation. Firstly, we reduced the number of crossbeams, retaining only two for essential structural support in Figure 2(a-ii) to improve its twisting adaptation while preserving its bending and enveloping capabilities. Further improvements were made by cutting excessive materials inside the crossbeam in Figure 2(a-iii), creating a see-through frame with a complete cavity inside to accommodate visual sensing. We also cut excessive materials on the back surface and both sides of the finger, resulting in a trapezoid shape in Figure 2(a-iv).



**Figure 2: Soft finger design, fabrication, and assembly.** (a) The finger design was inspired by (i) the Fin Ray Effect, and we improved the structure by (ii) reducing the crossbeams, (iii) hollowing out the beams, and (iv) eliminating the back surface so that the finger is more capable of bending, twisting, and enveloping during the interaction. (b) The fabrication of the soft finger included three steps: (i) degassing and mixing, (ii) casting and leaking, and (iii) curing and de-molding. (c) A monocular camera is mounted under the finger to capture the images containing the three markers placed under the three beams. (d) The three markers can be tracked unobstructedly in camera view.

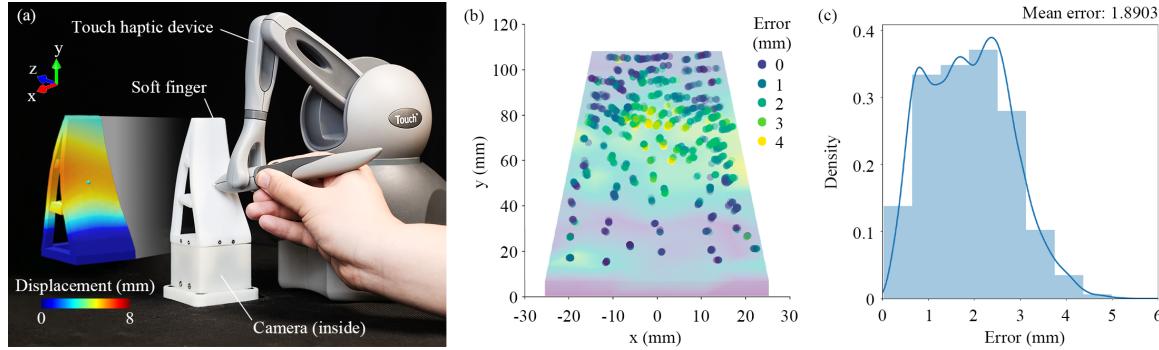
This optimization further enhances its spatial adaptation as these geometries do not directly contact objects. As a result, we developed a soft finger capable of adaptive bending, twisting, and object envelopment following these modifications. See Movie S1 in the Supplementary Materials for a video demonstration.

Figure 2(b) illustrates the fabrication process. We employed Hei-Cast 8400, a three-component polyurethane elastomer. Initially, equal quantities of the A and B components were weighed and separately degassed for 10 minutes at 25–35°C. The C component was omitted, aiming for a hardness level of 90A based on prior experience. Subsequently, the B component was added to the A component under vacuum conditions and left to stand for 30–40 seconds in Figure 2(b-i). Once the two components were thoroughly mixed, the mixture was rapidly cast into a silicone mold at a temperature range of 60–70°C in Figure 2(b-ii). After 90 seconds of mixing, the vacuum was released. The mold was then placed in a 60°C thermostatic oven for 60 minutes before demolding. The resulting soft finger was mounted to a support frame, as shown in Figure 2(c). Beneath the finger, a monocular camera was added to capture images at 60 frames per second. Three markers (8mm wide ArUco markers) were attached to the undersides of three rigid marker plates attached beneath three finger beams. Figure 2(d) shows the arrangement of these markers within the camera’s view. The unobstructed cavity inside lets the camera fully capture the markers’ poses, enabling a data-driven approach for shape and touch estimation during interactions.

## 2.2 Interactive Shape Estimation via Constrained Geometric Optimization

This experiment demonstrates the effectiveness of the proposed method for shape estimation of the soft finger. As shown in **Figure 3(a)**, the soft finger was mounted on the platform, and a Touch™ haptic device was used to contact the surface of the finger. We manually selected 150 contact points on the surface of the finger, which roughly covered the whole surface. At each selected point, the pen-nib of the Touch was used to make contact and push forward five times. During the process, the camera under the finger captured images at 60 fps to get the poses of the three markers, corresponding to the boundary conditions of the AMHs for shape estimation. The Touch also recorded the position of the pen-nib.

This process yielded a rich dataset capturing many deformation states for each contact point. Specifically, for each of the 150 selected points, we obtained data from 5 pushing interactions, each recorded at 60 fps for a minimum duration of 5 seconds. This equates to 300 distinct deformation states captured per contact point ( $5 \text{ seconds} \times 60 \text{ frames per second}$ ) within its physically permissible motion range. When aggregated across all selected points, this methodology produced a comprehensive database of 45,000 unique data entries ( $150 \text{ points} \times 300 \text{ deformation states per point}$ ) for accuracy evaluation. See Movie S2 in the Supplementary Materials for a video demonstration.



**Figure 3: Interactive shape estimation for the soft robotic metamaterial.** (a) The Touch™ haptic device is used to make contact with the finger, and the proposed method synchronously estimates the finger's shape. The pen-nib was pushed forward five times at 150 different positions on the finger's surface. (b) The error distribution between the measured ground-truth position of the pen-nib using the Touch device and the corresponding estimated position of the contact point on the surface is plotted in the x-y plane. (c) The error density is plotted with kernel density estimation, and the mean error is 1.8903 mm.

To evaluate the result of our proposed method for shape estimation, we reconstruct the deformed shape of the soft finger with observed AMHs' rigid motion using the algorithm to be explained in Section 4.1 and compare the estimated position of the contact point on the finger surface with the recorded position of pen-nib along pushing path. Since there is no slip between the finger surface and the pen-nib, recording the pushing position of the pen-nib and selecting points from the path is equivalent to sampling the reconstructed deformation field on the finger surface. As shown in Figure 3(b), the position error distribution between the recorded surface point and the corresponding reconstructed deformed surface point is acceptable, with a more significant error concentrated on the upper part of the finger surface. Figure 3(c) illustrates the statistical characteristics of the error distribution. Though the error distribution does not follow a normal one, the norm of mean error is relatively low at 1.8903 mm, and the error band is also narrow, ranging from 0 mm to 6 mm.

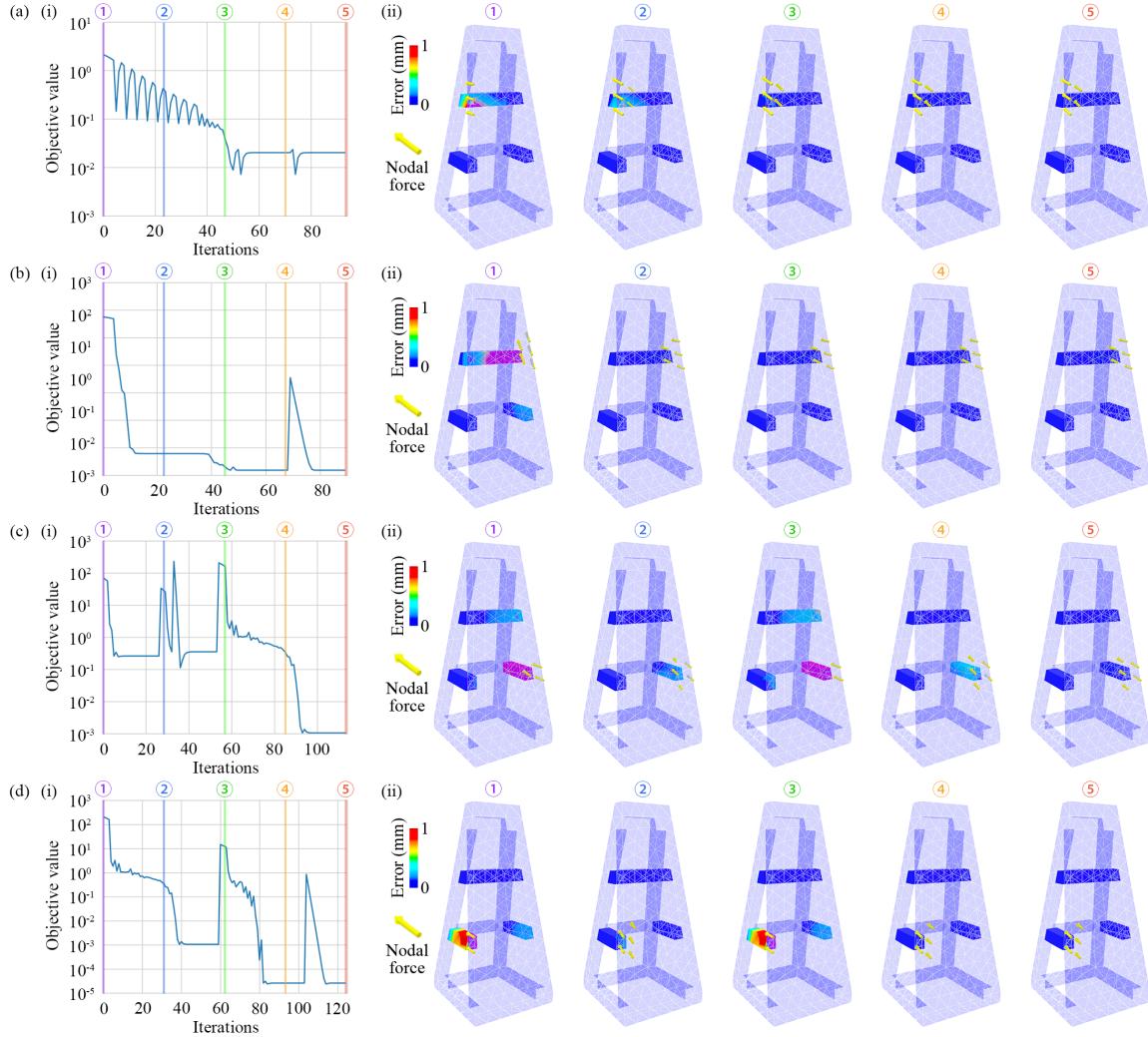
Precisely characterizing the deformation of soft robots is particularly challenging due to their inherently infinite-dimensional nature [57, 58], a difficulty further compounded by limited feature observation. Adding additional markers offers more boundary conditions for shape estimation models, thus enhancing accuracy. Considering our proposed design's structural and cost benefits, obtaining estimation results with acceptable errors using fewer markers remains a notable achievement in our research. In our experimental setup, we utilized the Touch™ haptic device as the source of contact position data. While motion capture systems offer superior positional accuracy, our experience with the soft finger indicates that occlusion of the motion capture camera during deformation can result in data loss in specific scenarios. To mitigate this, we aimed for comprehensive data sources that minimize accuracy loss; hence, we chose to employ the Touch device for interactions with the soft finger.

We implemented the proposed estimation algorithm in C++ and conducted tests on a standard PC with an Intel® Core™ i7 3.8 GHz CPU and 16 GB RAM. Leveraging algorithmic differentiation of the numerical solver, specifically using Eigen [59], our system enables real-time deformable shape estimation with a 3k tetrahedra mesh (up to 20 fps).

## 2.3 Adjoint Method for Optimal Touch Estimation

Reconstructing touch perception involves not only deforming the template mesh and aligning with the visually observed AMHs' motion but also inferring the cause of the deformation. This necessitates transitioning from purely geometric considerations to a mechanical perspective [60].

This section presents the result using an adjoint-based method to reconstruct a surface force distribution that explains the AMHs' motion from visual observation. This reconstruction leverages the discretized model described in the preceding section. When only a few geometric features are observed on a deformed elastic object, it is often complicated, and sometimes impossible, to deduce information about the distribution of contact forces [61]. Such *Inverse problems* [62] frequently arise in the broader context of science and engineering, where the observable components associated with the physical model output are constrained.



**Figure 4: Results of optimal contact nodal forces estimation using the adjoint-based method.** (a), (b), (c), and (d) depict the optimization process of estimating nodal forces for four distinct deformation configurations of the soft finger under varying contact nodal forces. In each case, the objective function value associated with the observed AMHs gradually converges with some oscillation, ultimately leading to optimal nodal forces.

As indicated by the workflow of the adjoint-based method elaborated in Section 4.2, the central component of the computation entails solving the equilibrium state position from the hyperelastic PDE constraint  $C$  using the current estimated force distribution, as well as evaluating the gradient of  $C$  concerning current equilibrium state position  $\mathbf{x}$ . We highlight that the equilibrium state position  $\mathbf{x}$  can be directly derived from the forward FEM simulation, and the gradient term  $\partial C / \partial \mathbf{x}$  coincides with the tangent stiffness matrix at that equilibrium state. Hence, to implement the adjoint-based method, we utilize

an academically accessible, high-speed FEM simulation tool to automatically resolve the PDE constraint and assemble the tangent stiffness matrix. Then, we deploy the conjugate gradient method for solving the adjoint state  $\lambda$  and apply the Limited-memory BFGS (Broyden–Fletcher–Goldfarb–Shanno) method to iteratively update the estimated force [63].

We generate four synthetic test cases using a discretized mesh model described in Section 4.1.1 and a Neo-Hookean hyperelastic constitutive model ( $E = 10$  Mpa,  $\mu = 0.47$ ), as is shown in **Figure 4**. We only permit admissible non-zero nodal forces at a small portion of vertices on the contact surface. Otherwise, the optimization process can quickly fail to converge, primarily due to applying a nonlinear hyperelasticity model. Apart from robustness concerns, each scenario takes nearly thirty seconds to achieve optimal estimation convergence, which may not be acceptable for real-time perception applications.

It is pertinent to note that while the BFGS algorithm is robust, it exhibits sensitivity to initial values and a propensity for slow convergence in the context of our application. These characteristics render it suboptimal for real-time sensing applications, as the computational demands cannot meet the latency requirements for immediate feedback or adjustments in practical settings. Recognizing these constraints, our research transitions from a deep, model-based analysis to a learning-based approach. The insights gained from the model-based exploration, particularly the principles of solid mechanics underlying the contact sensing problem, serve as helpful guidance. This steered our endeavors towards implementing comprehensive FEM simulations that take into consideration the realistic elements of contact sensing while also yielding significant and reliable training data.

Adopting a learning-based method involves offline training a machine learning model with the data generated from these informed simulations. Once trained, the model can infer or predict contact forces in real time, satisfying the critical requirement for real-time sensing in practical applications. This learning-based approach leverages deep insights into the mechanics of contact sensing obtained from the model-based analysis, ensuring that the training data is physically meaningful and aligned with real-world conditions.

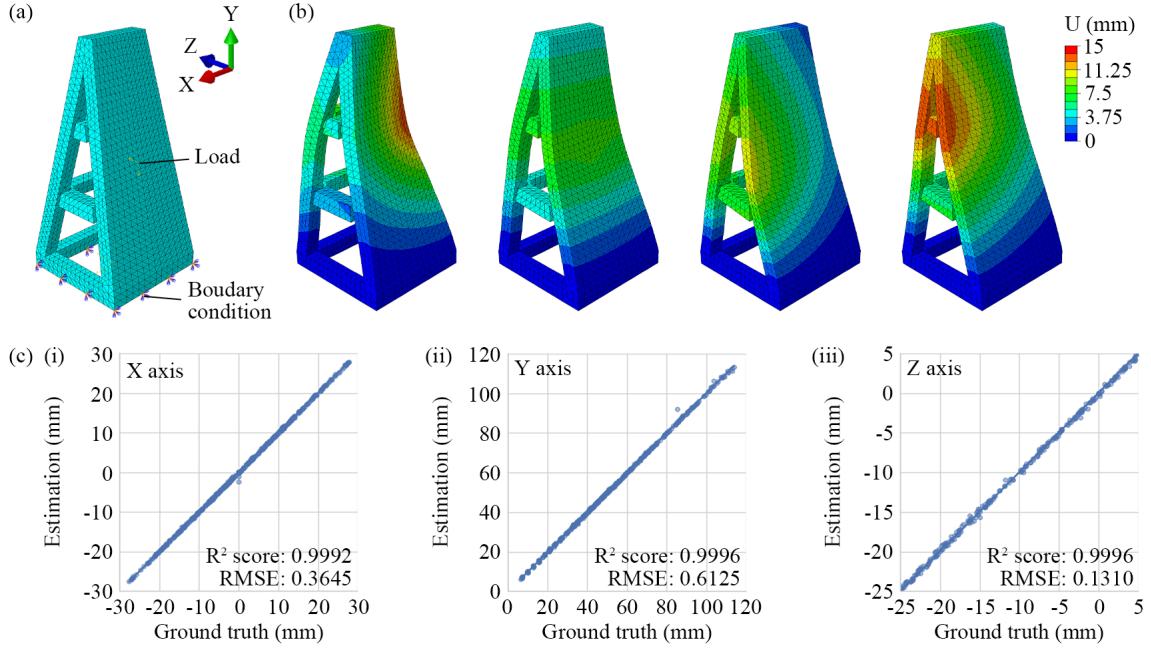
## 2.4 Estimation of Touch Position through Machine Learning

A data-driven method is employed to address the problem of touch position sensing. The input to this method remains the poses of the three markers obtained through the camera positioned at the bottom of the finger. A multi-layer perceptron (MLP) is selected as a regression model between the AMHs' poses and touch position. The output is a three-dimensional vector representing the touch position on the finger surface. This section illustrates the method of touch position estimation by training a neural network with simulated data from ABAQUS. The trained model is then transferred to and tested on actual data from real-world experiments.

### 2.4.1 Learning Touch Position using Synthetic Data from Simulation

As is shown in **Figure 5(a)**, we establish a fixed bottom boundary condition and simulate point contact with a concentrated load. Through uniformly sampling different load positions and forces, we capture the motion of the AMHs across various deformation configurations, as shown in **Figure 5(b)**. The comparison between the neural network output and the ground truth from the simulation is summarized in **Figure 5(c)** with coefficients of determination ( $R^2$ ) and root mean square errors (RMSE) of three dimensions reported. Analysis indicates a high correlation between the model's output of touch position along the  $x$ ,  $y$ , and  $z$  axes and the ground truth acquired using noise-free synthetic data from the simulation, with  $R^2$  scores of 0.9992, 0.9996, and 0.9996, respectively. Interestingly, the RMSE for the touch position estimation along the  $y$ -axis is notably larger at approximately 0.6125 compared to 0.3645 along the  $x$ -axis and 0.1310 along the  $z$ -axis.

The disparity in RMSE values observed across different axes can primarily be attributed to the scale-dependent characteristics of RMSE [64] despite its common use as a metric for assessing model performance [65]. Specifically, in our dataset, the range of variation is approximately 120mm on the  $y$ -axis, in



**Figure 5: Evaluation of the touch estimating model using simulation data.** (a) The FEM simulation setup for data generation. (b) Deformation configurations of the soft finger resulting from simulated point contact loading. Estimations of  $x$ ,  $y$ , and  $z$  versus the ground truth are shown in (c-i), (c-ii), and (c-iii).  $R^2$  scores along three directions are 0.9992, 0.9996, and 0.9996, and RMSE are 0.3645, 0.6125, and 0.1310.

contrast to 60mm on the  $x$ -axis and 30mm on the  $z$ -axis. This variance in scale significantly influences the RMSE values, resulting in 0.6125mm for the  $y$ -axis, 0.3645mm for the  $x$ -axis, and 0.1310mm for the  $z$ -axis. Additionally, it is crucial to consider the physical and mechanical properties of the soft robotic metamaterial structure, which might exhibit different behaviors or responses under varying stress or deformation across different axes [66]. These characteristics will likely complicate the accurate prediction of touch positions on the  $y$ -axis compared to the  $x$  and  $z$  axes.

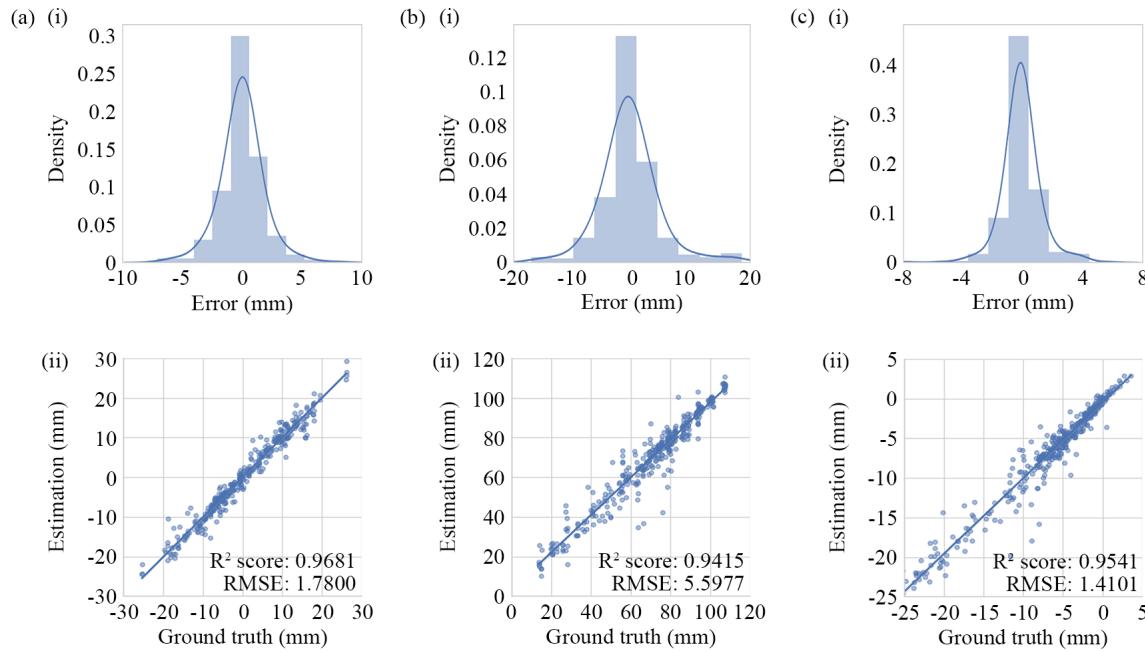
A benchmark analysis has been conducted to demonstrate our MLP-based approach's efficacy further, and the results of the  $R^2$  scores and the Root Mean Square Errors (RMSE) for each evaluated model are presented. As shown in **Table 2**, the MLP model outperforms the other methods in both metrics, showcasing its superior ability to predict the 3D contact locations from the input vectors accurately. Specifically, the MLP model achieves the highest  $R^2$  score, indicating a higher proportion of variance explained by the model, and the lowest RMSE, signifying lower prediction errors on average compared to Linear Regression, SVM, KNN, and Decision Trees. Furthermore, we have noted comparable scale-dependent variations in RMSE across various learning models. The consistency of these findings across different models highlights the significant impact of data scale on the RMSE metric, revealing its limitations as a performance metric in environments where data scale varies [67].

**Table 2:** Benchmark for different learning models.

Models	$R^2$ Score			RMSE		
	$x$	$y$	$z$	$x$	$y$	$z$
Linear regression (LR)	0.9040	0.9683	0.9683	1.9636	5.2502	1.1160
Support vector machines (SVM)	0.9423	0.9486	0.9927	1.4443	5.9299	0.5279
K-nearest neighbors (KNN)	0.9686	0.9915	0.9915	1.1336	2.7786	0.5906
Decision trees (DT)	0.9782	0.9955	0.9956	0.9401	2.3344	0.4352
<b>Multi-layer perception (MLP)</b>	<b>0.9992</b>	<b>0.9996</b>	<b>0.9996</b>	<b>0.3645</b>	<b>0.6125</b>	<b>0.1310</b>

#### 2.4.2 Estimation of Touch Perception using Real-World Data

As shown in **Figure 6**, the distributions of estimation errors along the  $x$  and  $z$  axes exhibit a more pronounced concentration than that along the  $y$ -axis. Compared with the model evaluation with simulation data, the  $R^2$  scores with actual experiment data, though still substantial, show a marginal decrease to 0.9681, 0.9415, and 0.9541, respectively. The discrepancy in performance, as evidenced by the slight decrease in the  $R^2$  score, can primarily be attributed to the gap between the simulated and real-world datasets. The simulation data, generated using ABAQUS FEM simulation, benefits from being noise-free and highly controlled. Such data allows for precise learning and prediction, as the variables and conditions are simplified versions of real-world phenomena, leading to potentially higher performance metrics. In contrast, the real-world data collected for training the same MLP model encompasses a broader spectrum of variability and randomness inherent to physical interactions.



**Figure 6: Evaluation of the touch estimating model using real-world data.** Estimations of  $x$ ,  $y$ , and  $z$  versus the ground truth are shown in (a), (b), and (c). And  $R^2$  scores along three directions are 0.9681, 0.9415, and 0.9541.

Furthermore, the RMSE is higher at 1.7800, 5.5977, and 1.4101 mm, in contrast to the simulation case. The observed increase in RMSE when evaluating the model with real-world data highlights potential limitations in the model's generalizability and adaptability to real-world conditions. It underscores the necessity for continuous refinement of predictive models, including developing models that are robust to noise and capable of capturing the complexity of real-world phenomena more accurately. Additionally, it points to the importance of improving data collection and preprocessing techniques to enhance the quality of real-world datasets.

During data collection and model training, we noticed that estimating contact positions across the entire surface of the finger could lead to non-unique solutions. This indicates a disparity between the simulation model and the fabricated soft robotic metamaterial. We attribute this to the constrained number of visual observations from a practical usage perspective, similar to the challenges in shape estimation.

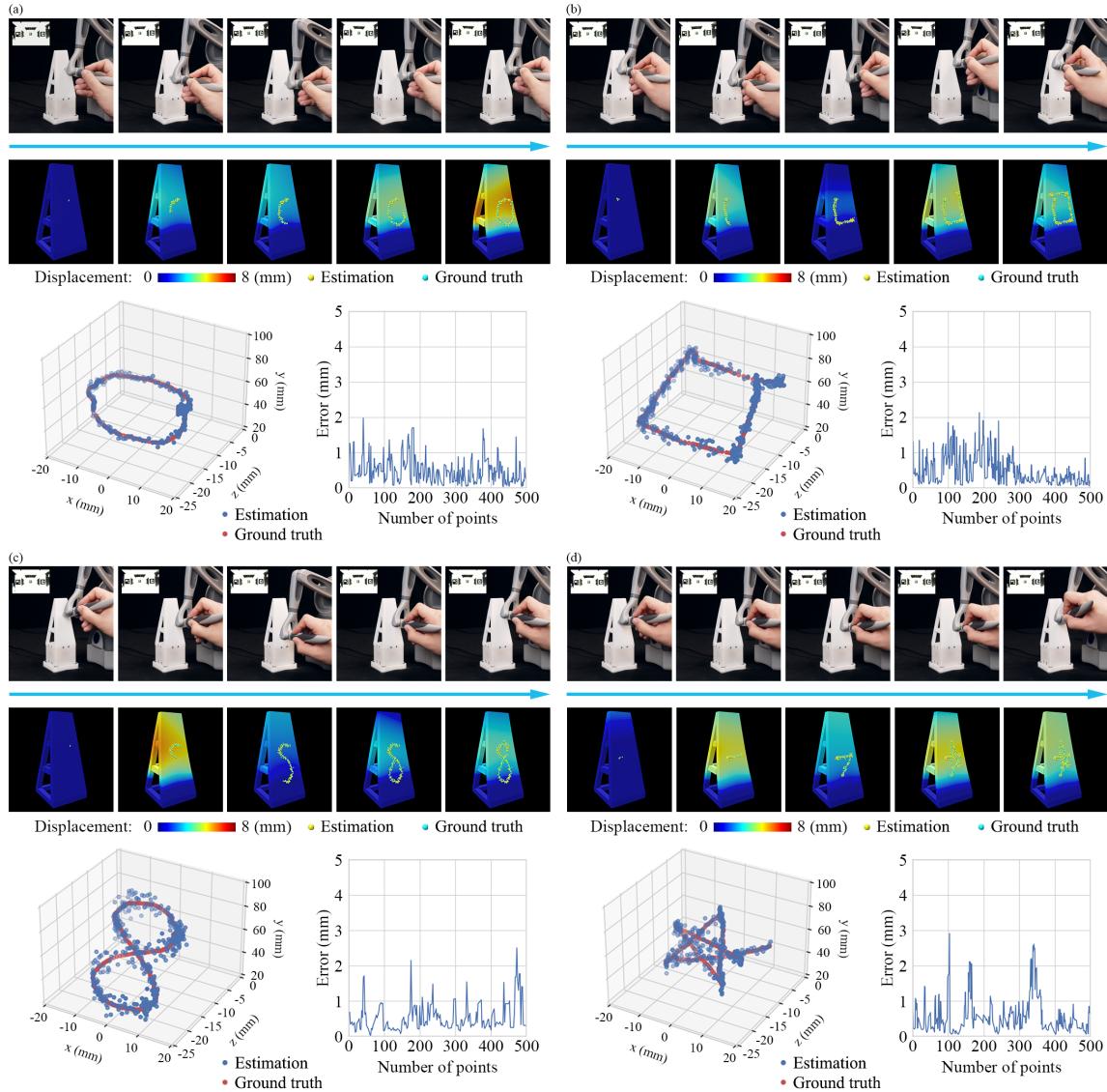
## 2.5 Demonstration for Touch Reconstruction

This section demonstrates human-robot interaction and tactile-based adaptive grasping experiments, showcasing the potential applications of the vision-based soft robotic metamaterial we have introduced.

### 2.5.1 Touch-based Human-robot Interaction

Touch-based interaction involves using tactile stimuli to communicate, control, or feedback in various human-machine interfaces [68], adding an intuitive and immersive dimension to user experiences across different domains, ranging from consumer electronics to industrial control systems [69]. In this experiment, we assessed the efficacy of touch interaction using the proposed touch perception-enabled soft finger, explicitly focusing on real-time dynamic touch estimation by detecting the contact path on the surface of the soft finger.

We employed the pen-nib of the Touch™ haptic device to trace paths of various shapes while consistently maintaining contact with the surface of the finger. Throughout this process, real-time estimation of contact positions was performed using a learning-based touch estimation model. Illustrated in **Figure 7**, we utilized four distinct path shapes—specifically, (a) circle, (b) square, (c) number 8, and (d) star to assess the precision of touch estimation during dynamic interactions. As touch-induced deformation of the fin-



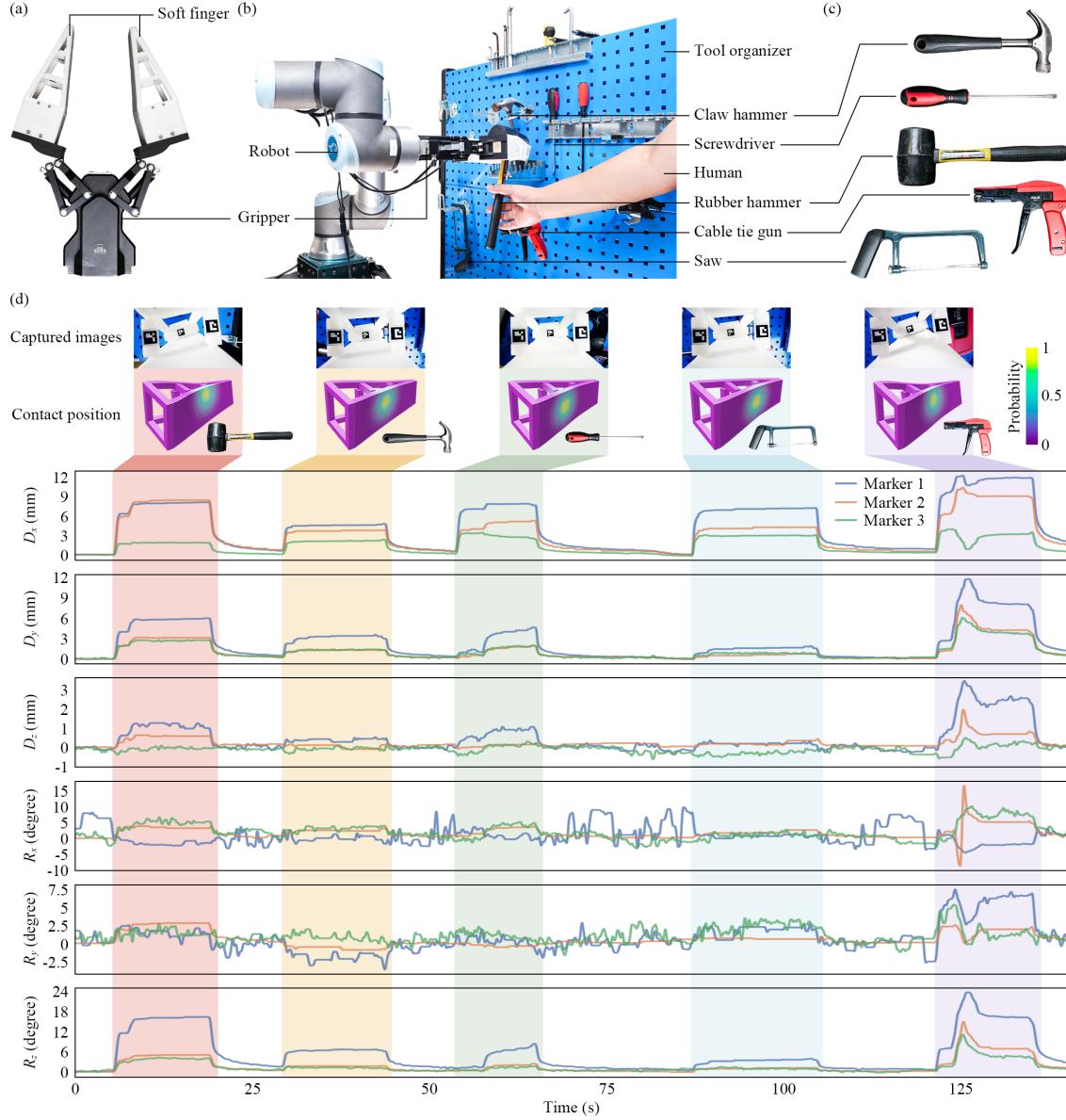
**Figure 7: Comparison between the estimated contact path and the ground truth for different shapes.** We use the pen-nib of the Touch™ haptic device to follow paths of different shapes as (a) circle, (b) square, (c) number 8, and (d) star. The estimated points and the ground truth are plotted in each figure for the four paths, and their errors are also plotted.

ger and the pen-nib made continuous contact with the finger surface, the paths traversed by the pen-nib existed in three-dimensional space. The results demonstrate that the estimated touchpoints align well

with the actual paths. While a few outliers are evident, they do not significantly impact the pen-nib's path shape estimation. See Movie S3 in the Supplementary Materials for a video demonstration.

### 2.5.2 Soft Intelligent Grasping with Touch Estimation

Incorporating touch perception-enabled soft fingers into an industrial robotic gripper allows these systems to dynamically adapt their grasp in real-time based on tactile input during object contact [38]. This integration allows for a more refined and adaptable touch, facilitating robust, gentle, and intelligent interaction with the environment.



**Figure 8: Object handovers in human-robot collaboration.** (a) The proposed soft finger is integrated with a rigid industrial gripper. (b) A collaborative robot, UR, can grasp objects from a tool organizer and transfer them to a human worker. (c) Target objects. (d) The observed pose of markers and the estimated contact position from the soft finger were demonstrated during the experiment.

As illustrated in **Figure 8**, we outline an industrial human-robot collaboration scenario [70], where the robot is instructed to grasp diverse tools, each with unique geometries from a tool organizer and transfer them to a human worker. Equipped with the designed soft robotic fingers (Figure 8(a)), a rigid industrial gripper effectively demonstrates its adaptable capability to securely grasp a variety of tools, as shown in Figure 8(c). As depicted in Figure 8(d), real-time touch estimation offers essential information

for the grasping system to accomplish the human-robot collaboration task. See Movie S4 in the Supplementary Materials for a video demonstration.

In **Figure 9(a)**, we explicitly illustrate the grasping adjustment capability using touchpoint estimation. The gripper can search for the most stable grasping point based on touchpoint estimation from the soft finger. As depicted in Figure 9(b), we implemented a straightforward regrasping strategy to locate a grasping point nearest to the center of the soft finger surface, which we considered the most stable. In the experiment, the gripper effectively located the target grasping position and securely gripped the pincer after five adjustment attempts. See Movie S5 in the Supplementary Materials for a demonstration.

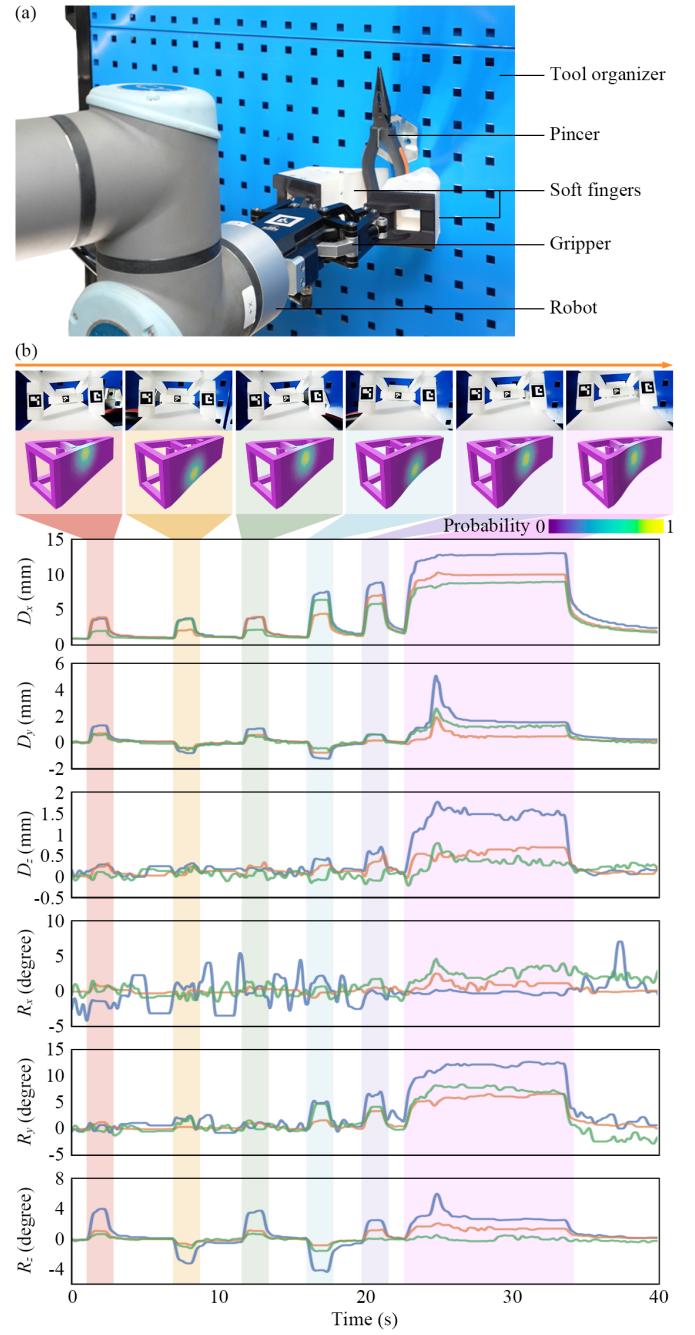
### 2.5.3 Reconstruct Soft Grasping of Fragile Objects

To demonstrate applications that utilize the dual benefits of soft finger deformation for both proprioception and exteroception, we conducted further experiments on the intricate task of grasping different soft objects in **Figure 10**. These experiments are designed to highlight the practical relevance of our research in real-life scenarios, where the subtle modulation of grasping force is crucial to avoid damaging the objects being manipulated.

We selected various objects recognized for their softness and fragility to damage, including a strawberry, a banana, a lime, a mango, and an empty soft drink can. Each object presents distinct challenges in shape, size, and fragility, thereby comprehensively evaluating our system's capability in delicate grasping.

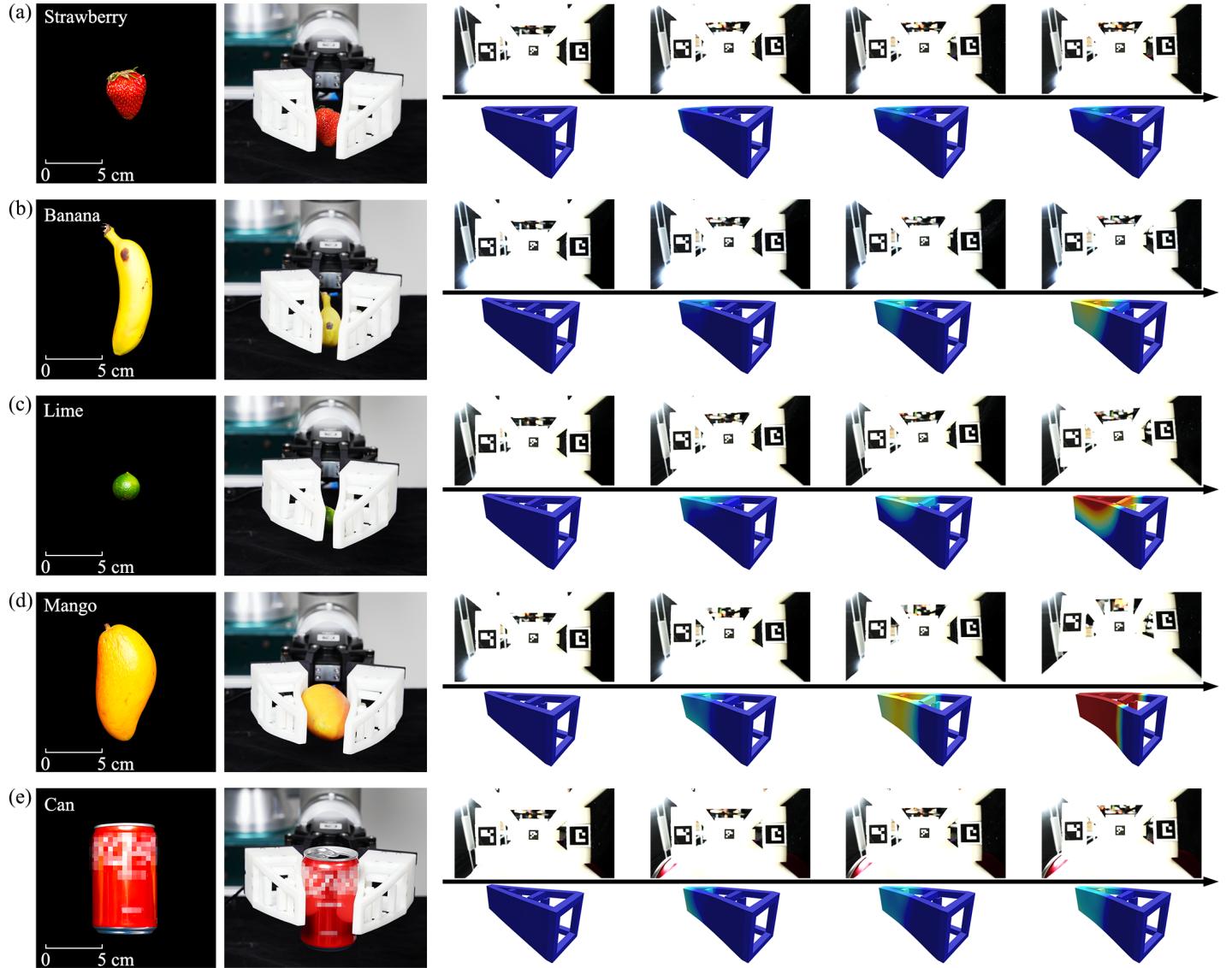
In these experiments, the robotic gripper was programmed to accept positional commands, relying exclusively on the deformation feedback obtained from the soft fingers to modulate the gripping force. There is a direct correlation between the degree of deformation and the contact force exerted by the gripper. By interpreting the deformation data in real-time, the system dynamically adjusted the grip to ensure that each object was held securely yet gently enough to avoid causing any damage. This approach underlines the concept of *soft grasping*, which emphasizes the importance of a delicate touch in robotic manipulation, particularly when interacting with fragile or easily deformable objects.

These experiments, which achieved damage-free grasping of fragile objects, demonstrate the proposed system's potential applications in enhancing robotic grippers' tactile sensitivity and adaptability. Our



**Figure 9: Adaptive grasping with touch estimation.**  
 (a) A collaborative robot, UR, locates the desired grasping point on a pincer based on the information obtained from estimating the soft finger's touch. (b) The pose of markers and the estimated touch point were presented during the experiment. Various stages of a simple grasping strategy utilizing touchpoint estimation are highlighted in different colors.

system works towards more delicate and nuanced robotic grasping by harnessing the feedback on proprioception and exteroception provided by the soft finger deformation. See Movie S6 for demonstrations on gently grasping various delicate objects and Movie S7 for re-grasping an orange facilitated by soft finger touch reconstruction.



**Figure 10: Delicate grasping enhanced by touch estimation.** The first column lists the objects tested, including (a) a strawberry, (b) a banana, (c) a lime, (d) a mango, and (e) an empty soft drink can. The second column is the screenshot of the experiment with further details in the supplementary video. The third column contains screenshots of the in-finger vision with estimated deformation of the soft finger during grasping.

### 3 Discussions

This research introduces an innovative soft finger design proficient in passive 3D deformation. It is well-suited for enveloping objects and offers an ample internal space devoid of occlusions. This configuration facilitates vision-based proprioceptive sensing of shape deformation and touch positioning.

The proprioceptive sensing approach relies on geometry-based optimization, approximating various physical interactions between the soft finger and the environment as geometry-based positional constraints for three sets of vertices on the finger's template mesh. These constraints are observed using an off-the-shelf ArUco marker detector. By enforcing the positional constraints introduced by three deformation

handles based on vertex aggregation on the template mesh of the soft finger, we reconstruct the deformed shape of the finger through the minimization of a constrained energy function. The reconstructed deformation field, uniformly sampled on the finger's surface, demonstrates remarkable accuracy, with a mean error norm of only 1.89 mm. Although the optimization algorithm is relatively comprehensive, the computation speed is limited to 20 Hz for 4k tetrahedron elements, and there is room for improvement through parallel or GPU computation.

When estimating the touch position responsible for the soft finger deformation, we transition from purely geometric considerations to a mechanical perspective by incorporating the hyperelastic constitutive model to account for substantial deformation. We can estimate the contact force distribution that best fits the observed AMHs' motion through the proposed adjoint-based optimization method. However, the solution to the optimization problem for the nonlinear constitutive model is time-consuming and sensitive to the initial guess.

In contrast to model-based optimization, we also propose a data-driven method for practical use. After the proof of concept stage of model training using simulation data, we successfully apply the same neural network model for touch position estimation from synthetic data in simulation to real-world usage, exhibiting minimal performance degradation. This results in impressive  $R^2$  scores of 0.9681, 0.9415, and 0.9541 along the  $x$ ,  $y$ , and  $z$  axes, respectively.

In addressing the broader scope of applications for our research beyond industrial human-robot collaboration scenarios, we have extended our exploratory efforts to include the intelligent soft grasping of a wide array of delicate objects. This initiative capitalizes on the nuanced capabilities afforded by touch reconstruction to facilitate delicate interactions with objects that necessitate a refined approach to manipulation. Through these additional experiments, we aim to showcase the adaptability and precision of our method in managing a diverse collection of items, underscoring its practicality for sensitive applications.

Our future work plans to further expand the application of our proposed vision-based proprioceptive soft finger for accurately visualizing the shape and touch of adaptive grasping in challenging tasks such as adaptive and sensitive grasping in the deep-water [71], where the integration of tactile sensing and real-time visualization of the finger adaptation is complex to achieve using the exiting sensing methods and finger designs. We also intend to conduct in-hand and in-finger manipulation experiments to explore the influence of finger proprioceptive capabilities on control strategies. Moreover, the fusion of vision-based sensing with other sensing modalities, such as optical fibers and magnetic sensors, can create a multi-modal sensing system for enhanced human-robot interactions with a rich level of digitized informatics.

## 4 Materials & Methods

### 4.1 Soft Robotic Metamaterial Deformation Characterization and Optimization

This section introduces an approach based on constrained geometric optimization for estimating the deforming shape of a soft finger in 3D. Considering the intricate volumetric form of the soft finger depicted in Figure 1(a), we define a 3D-space domain, denoted as  $\Omega \in \mathbb{R}^3$ , filled with a hyperelastic material. Depending on the applied boundary conditions, the distribution of internal elastic energy within the finger can exhibit significant variations. Our objective in this shape sensing method is to geometrically deduce the smooth deformation map  $\Phi : \Omega \rightarrow \tilde{\Omega}$  using a monocular camera. This map transforms the soft finger from its initial configuration  $\Omega$  to a deformed configuration  $\tilde{\Omega}$  by minimizing its internal energy. Our proposed soft finger shape estimation approach relies on finite element discretization and selecting an appropriate energy function that characterizes deformation.

#### 4.1.1 Deformation Modeling

Figure 1(b-i) shows the discretized soft finger model using tetrahedral elements, yielding a template mesh  $\mathcal{M} = \{\mathcal{V}, \mathcal{T}\}$  consisting of finite vertex points  $\mathcal{V} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbb{R}^3\}$  and a collection of tetrahedra

elements  $\mathcal{T} = \{t_1, t_2, \dots, t_m\}$ . When the soft finger undergoes deformation, a set of chosen linearly approximated local deformation maps are applied to  $\mathcal{M}$  over each tetrahedron element  $t_j$  through an affine transformation:

$$\Phi_{t_j}(\mathbf{X}) = \mathbf{F}_{t_j}\mathbf{X} + \mathbf{b}_{t_j}, \quad (1)$$

where  $\mathbf{X} \in \mathbb{R}^3$  represents all points inside element  $t_j$ ,  $\mathbf{F}_{t_j} \in \mathbb{R}^{3 \times 3}$  is the differential part of the deformation map, and  $\mathbf{b}_{t_j} \in \mathbb{R}^3$  is the translation part. This piecewise linear deformation map is chosen for computational efficiency. For the selected linear tetrahedral elements, the deformation gradient  $\mathbf{F}_{t_j}$  can be expressed as a linear combination of deformed element vertices' locations  $\mathbf{x}_{t_j}$  as :

$$\mathbf{F}_{t_j}(\mathbf{x}_{t_j}) = \frac{\partial \Phi|_{t_j}}{\partial \mathbf{X}} = \mathbf{D}_s(\mathbf{x}_{t_j}) \cdot \mathbf{D}_m^{-1}(\mathbf{X}_{t_j}), \quad (2)$$

where

$$\mathbf{D}_s(\mathbf{x}_{t_j}) = [\mathbf{x}_{t_j}^2 - \mathbf{x}_{t_j}^1 \quad \mathbf{x}_{t_j}^3 - \mathbf{x}_{t_j}^1 \quad \mathbf{x}_{t_j}^4 - \mathbf{x}_{t_j}^1], \quad (3)$$

$$\mathbf{D}_m(\mathbf{X}_{t_j}) = [\mathbf{X}_{t_j}^2 - \mathbf{X}_{t_j}^1 \quad \mathbf{X}_{t_j}^3 - \mathbf{X}_{t_j}^1 \quad \mathbf{X}_{t_j}^4 - \mathbf{X}_{t_j}^1]. \quad (4)$$

For the discretized tetrahedral mesh  $\mathcal{M}$ , a collection of deformation maps  $\{\Phi_{t_j}\}_{t_j \in \mathcal{T}}$  should uniquely determine the deformed shape of the soft finger [72]. In the soft finger shape reconstruction problem, we proposed to choose a geometrically motivated, rotation-invariant isometric distortion metric, the symmetric Dirichlet energy, for computational efficiency [73]:

$$\Psi(\mathbf{F}) = \|\mathbf{F}\|_{\mathcal{F}}^2 + \|\mathbf{F}^{-1}\|_{\mathcal{F}}^2, \quad (5)$$

where  $\|\cdot\|_{\mathcal{F}}$  is the Frobenius norm. Since the deformation should be irrelevant to translation, the accumulated discrete element energy function only takes the augmentation of deformation gradients as:

$$E(\mathbf{x}) = \sum_{t_j \in \mathcal{T}} \Psi(\mathbf{F}_{t_j}(\mathbf{x})), \quad (6)$$

where  $\mathbf{x} \in \mathbb{R}^{3 \times n}$  contains all the discretized vertices' locations of  $\mathcal{M}$ .

#### 4.1.2 Rigidity-aware Aggregated Multi-Handle Constraints

Deformable shape reconstruction from 2D image observations is a well-known ill-posed inverse problem and is actively researched in the computer vision community [74]. To make this problem tractable in this study, discretized template mesh and rigidity-aware aggregated multi-handle (AMH) constraints are introduced to reconstruct the deformed shape of the soft finger reliably from the camera. As shown in Figure 1(b-ii), three vertex groups act as deformation handles [75] that drive the template mesh to deform towards the target shape. Each handle undergoes rigid motion  $g : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  separately as:

$$\mathbf{x}_{\text{AMH}_i} = g(\mathbf{X}_{\text{AMH}_i}), i \in \{1, 2, 3\}, \quad (7)$$

where  $\mathbf{x}_{\text{AMH}_i}$  contains deformed positions of the vertex group in the target shape, as shown in Figure 1(b-iii), and  $\mathbf{X}_{\text{AMH}_i}$  contains undeformed positions of the corresponding vertex group in the template mesh. The rigid transformation  $g$  is estimated using ArUco markers shown in Figure 1(c).

#### 4.1.3 Shape Estimation by Distortion Energy Minimization

With the discrete energy function of the soft finger  $\mathcal{M}$  and observed AMH constraints, soft finger shape estimation can be directly translated into a constrained geometry optimization problem. Instead of considering kinematic constraints of AMHs as *hard* boundary conditions, they are enforced by appending quadratic penalty functions to the original function in Equation (6), as:

$$\tilde{E}(\mathbf{x}; \mathbf{g}, \boldsymbol{\omega}) = \sum_{t_j \in \mathcal{T}} \Psi(\mathbf{F}_{t_j}(\mathbf{x})) + \sum_{i=1}^3 \omega_i \|\mathbf{x}_{\text{AMH}_i} - g_i(\mathbf{X}_{\text{AMH}_i})\|^2, \quad (8)$$

The deformed shape can be estimated by minimizing the augmented energy function:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \tilde{E}(\mathbf{x}; \mathbf{g}, \boldsymbol{\omega}). \quad (9)$$

Note that a more significant penalty weight  $\omega_i$  results in better constraint satisfaction, but for poor numerical conditions, we set  $\omega_i = 10^5$  for our soft finger. We optimize the problem in Equation (9) via projected Newton, as shown in **Algorithm 1**, to solve the vertices' positions in the deformed shape.

---

**Algorithm 1** Deformed Shape Estimation Algorithm

---

**Input:** Rigid Motion of AMHs  $g_i, i \in \{1, 2, 3\}$   
**Output:** Estimated Vertices of Deformed Shape  $\mathbf{y}_d$

**Require:**

- $\omega_i = 10^5, i \in \{1, 2, 3\}$
- Vertices positions of current shape  $\mathbf{y}_0$
- Initial feasible solution  $\mathbf{x}_0 = \mathbf{y}_0$
- Convergence tolerance  $\epsilon = 1e^{-4}$
- Maximum number of iterations  $N_{\max} = 1000$
- $k \leftarrow 0$
- Compute gradient  $\nabla \tilde{E}(\mathbf{x}_k)$  and Hessian  $\nabla^2 \tilde{E}(\mathbf{x}_k)$
- while**  $\|\nabla \tilde{E}(\mathbf{x}_k)\| > \epsilon$  and  $k < N_{\max}$  **do**

  - Solve  $\nabla^2 \tilde{E}(\mathbf{x}_k) \Delta \mathbf{x}_k = -\nabla \tilde{E}(\mathbf{x}_k)$  for  $\Delta \mathbf{x}_k$
  - Project  $\Delta \mathbf{x}_k$  onto the feasible region
  - Update iterate:  $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \Delta \mathbf{x}_k$
  - $k \leftarrow k + 1$

- end while**
- iteration stop  $\mathbf{y}_d = \mathbf{x}_k$

---

## 4.2 Optimal Touch Estimation Formulation

The sense of touch position is another aspect of proprioception involving estimating the contact location. This paper focuses on the contact between the finger surface and objects like pen nibs, which resemble point contact. By accurately determining where and how contact occurs between surfaces or objects, tactile perception systems can better comprehend and respond to physical stimuli, enhancing their ability to recognize textures, objects, and forces.

To tackle this optimal touch estimation problem, we consider the governing equation of a solid body with hyperelastic deformation.

$$\begin{aligned} \nabla_{\mathbf{X}} \cdot \mathbf{P} + \mathbf{f} &= \mathbf{0} \text{ in } \Omega, \\ \mathbf{u} &= \bar{\mathbf{u}} \text{ on } \partial\Omega_D, \\ \mathbf{P} \cdot \mathbf{N} &= \bar{\mathbf{t}} \text{ on } \partial\Omega_N, \end{aligned} \quad (10)$$

where  $\nabla_{\mathbf{X}} \cdot \mathbf{P}$  denotes the divergence operator applying on the first Piola-Kirchhoff stress concerning  $\mathbf{X}$  on the initial configuration [76].  $\mathbf{u}$  denotes the displacement field to be solved for.  $\bar{\mathbf{u}}$  and  $\bar{\mathbf{t}}$  are prescribed values for Dirichlet boundary  $\partial\Omega_D$  and Neumann boundary  $\partial\Omega_N$ .  $\mathbf{N}$  represents the outward normal unit vector, and body force is denoted as  $\mathbf{f}$ .

Touch estimation is equivalent to finding a surface force distribution  $\bar{\mathbf{t}}$ , which results in a displacement field that best accounts for the visual observation. We follow the discretize-then-optimize [77] approach to this constrained PDE optimization [78]. Given the discretized vertex positions of the initial configuration  $\Omega$  of the soft finger, solving Equation (10) for the displacement  $\mathbf{u}$  is tantamount to resolving the vertex positions  $\mathbf{x}$  in the deformed configuration  $\tilde{\Omega}$ . Thus, the optimization reads:

$$\begin{aligned} \min_{\mathbf{t} \in T} \quad J(\mathbf{x}(\bar{\mathbf{t}}), \bar{\mathbf{t}}) &= \sum_{i \in I_{AMH}} \|\mathbf{x}(\bar{\mathbf{t}})_i - \mathbf{x}_i\|^2, \\ \text{s.t.} \quad C(\mathbf{x}, \bar{\mathbf{t}}) &= 0, \end{aligned} \quad (11)$$

where  $J$  is the loss function that measures the discrepancy between PDE model outputs  $\mathbf{x}$  and observed outputs  $\mathbf{x}_i$  at AMH constraint vertex groups. The set  $I_{AMH}$  contains all the vertex indices constrained by AMHs described in Section 4.1.2. The constraint function  $C$  represents the discretized governing Equation (10) and  $\bar{\mathbf{t}}$  is estimated from the admissible contact force distribution set  $T$ .

To leverage efficient gradient-based optimization algorithms, we adopted the adjoint state method for computing the gradients of the objective function concerning contact force distribution variables

$$\frac{dJ}{d\bar{\mathbf{t}}} = \frac{\partial J}{\partial \mathbf{x}} \frac{d\mathbf{x}}{d\bar{\mathbf{t}}} + \frac{\partial J}{\partial \bar{\mathbf{t}}}. \quad (12)$$

We then take the derivative of the constraint function concerning estimated variables

$$\frac{dC}{d\bar{\mathbf{t}}} = \frac{\partial C}{\partial \mathbf{x}} \frac{d\mathbf{x}}{d\bar{\mathbf{t}}} + \frac{\partial C}{\partial \bar{\mathbf{t}}} = \mathbf{0}. \quad (13)$$

and we have,

$$\frac{d\mathbf{x}}{d\bar{\mathbf{t}}} = -(\frac{\partial C}{\partial \mathbf{x}})^{-1} \frac{\partial C}{\partial \bar{\mathbf{t}}}. \quad (14)$$

Then, substitute Equation (14) to Equation (12), we have,

$$\frac{dJ}{d\bar{\mathbf{t}}} = -\frac{\partial J}{\partial \mathbf{x}} (\frac{\partial C}{\partial \mathbf{x}})^{-1} \frac{\partial C}{\partial \bar{\mathbf{t}}} + \frac{\partial J}{\partial \bar{\mathbf{t}}}. \quad (15)$$

We can define the adjoint variable  $\boldsymbol{\lambda}$  as

$$\frac{\partial C}{\partial \mathbf{x}} \boldsymbol{\lambda} = \frac{\partial J}{\partial \mathbf{x}}. \quad (16)$$

Finally, substitute  $\boldsymbol{\lambda}$  to Equation (15) and we have the gradient we want

$$\frac{dJ}{d\bar{\mathbf{t}}} = -\boldsymbol{\lambda} \frac{\partial C}{\partial \bar{\mathbf{t}}} + \frac{\partial J}{\partial \bar{\mathbf{t}}}. \quad (17)$$

Based on the defined problem in Equation (11), we present the overall workflow of the optimal touch estimation procedure in **Algorithm 2**. We provide visual representations of numerical results obtained through the adjoint method for touch estimation, using the open-source deformable simulation tool Vega [79]. The results are enclosed in Section 2.3.

---

**Algorithm 2** Optimal Touch Estimation Algorithm

---

```

Input: Initial contact force estimation  $\bar{\mathbf{t}}_{ini}$  and maximum iteration number  $i_{max}$ 
 $\bar{\mathbf{t}} \leftarrow \bar{\mathbf{t}}_{ini}, i \leftarrow 0$ 
while  $i < i_{max}$  do
    Compute the equilibrium state position  $\mathbf{x}$  by solving  $C(\mathbf{x}, \bar{\mathbf{t}}) = 0$ 
    Evaluate  $\frac{\partial C}{\partial \mathbf{x}}$  and  $\frac{\partial J}{\partial \mathbf{x}}$ 
    Compute the adjoint state  $\boldsymbol{\lambda}$  by solving Equation (16)
    Compute  $\frac{dJ}{d\bar{\mathbf{t}}}$  by Equation (17)
    Update  $\bar{\mathbf{t}}$  using  $\frac{dJ}{d\bar{\mathbf{t}}}$  by gradient-based optimizer
     $i \leftarrow i + 1$ 
end while
Output: Estimated optimal contact force distribution  $\bar{\mathbf{t}}_{opt}$ 

```

---

## 4.3 Machine Learning-Assisted Touch Estimation from Visual Observation

### 4.3.1 Synthetic Data Generation using FEM Simulation

We used ABAQUS to simulate the soft finger deformation under different loading forces. The finite element model of the finger contained 4,268 nodes and 15,992 elements (C3D4). We selected 570 nodes on

the finger surface as loading positions. A series of normal forces with 21 values of (1, 6) N were applied at each loading position to simulate pointwise contact on the finger surface. In addition, we recorded the node sets corresponding to the three AMHs and calculated the poses ( $D_x, D_y, D_z, R_x, R_y, R_z$ ) of these node sets. After calculation, except for a few convergence failure samples, we finally obtained 11,222 valid data. The data was divided into training set, validation set, and test set by the ratio of 7:1:2. The input of the MLP model was the poses of the three AMHs, whose dimension is 18. And the output is the loading position, whose dimension is 3. By this MLP model, we can estimate the touch position with the three AMHs' poses. The results are enclosed in Section 2.4.1.

#### 4.3.2 Real-world Data Collection for Model Training

To address the disparity between the simulation model and the real soft robotic metamaterial, we collected a dataset of 50,000 samples, capturing the poses of three markers from the camera and the touch position from the Touch™ haptic device. This collection utilized the same experimental setup as detailed in Section 2.2. We employed the same learning model used for training with simulation data to process real-world data. The model takes the poses of the three markers as input and outputs the touch position, with two hidden layers of dimensions 512 and 256, respectively. All networks were trained on a Xiaomi Mi Notebook Pro equipped with Intel® Core™ i7-8750H, Nvidia® GeForce® GTX 1060, batch size 128, and Adam optimizer [80]. The initial learning rate was set to 0.001 and decreased with the training epoch. The results are enclosed in Section 2.4.2.

#### Acknowledgements

This work was partly supported by the National Natural Science Foundation of China [62206119], Shenzhen Long-Term Support for Higher Education at SUSTech [20231115141649002], SUSTech Virtual Teaching Lab for Machine Intelligence Design and Learning [Y01331838], the Science, Technology, and Innovation Commission of Shenzhen Municipality [JCYJ20220818100417038, JSGG20220831110002004].

## References

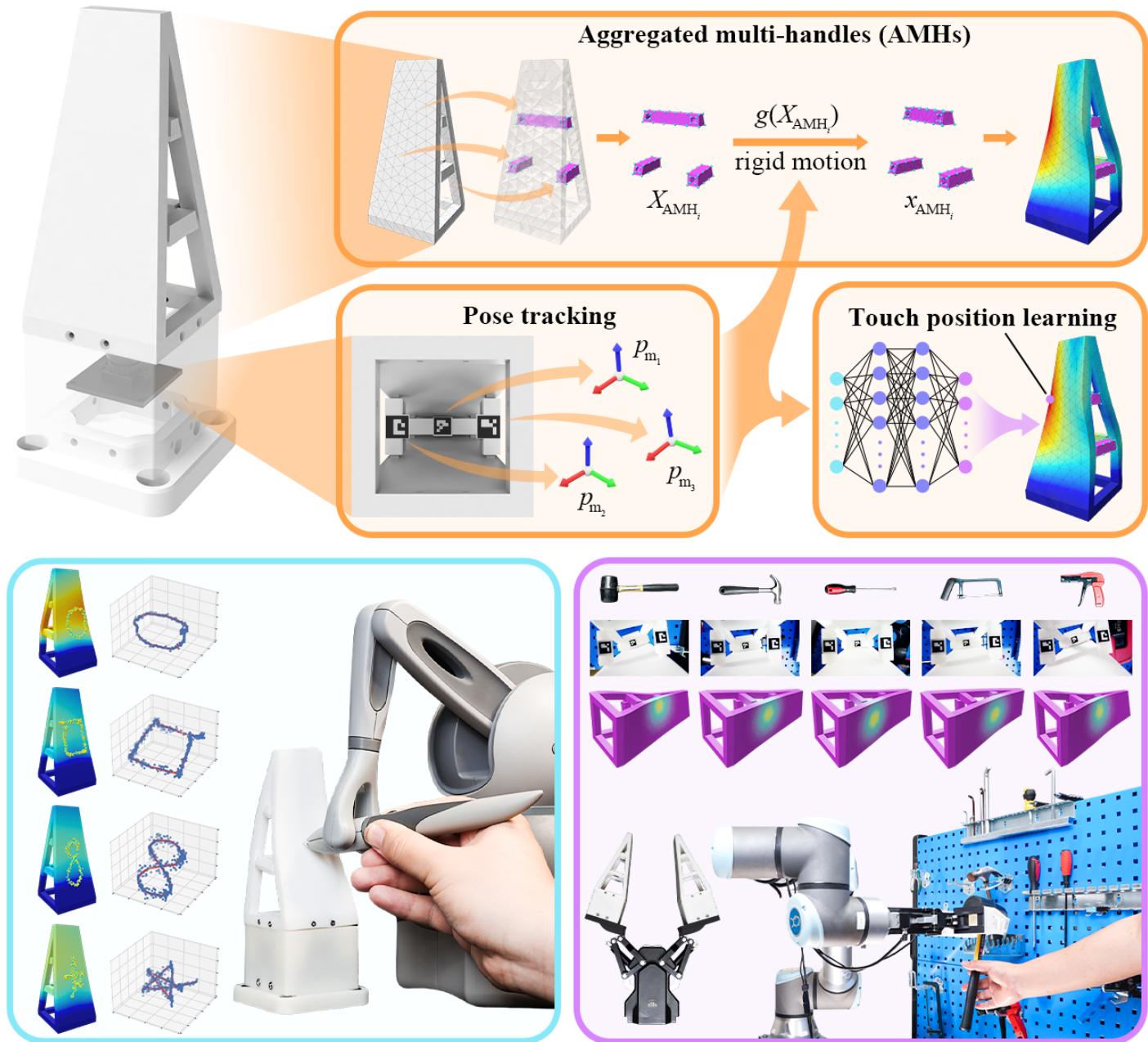
- [1] R. S. Johansson, J. R. Flanagan, *Nature Reviews Neuroscience* **2009**, *10*, 5 345.
- [2] X. Yu, Z. Xie, Y. Yu, J. Lee, A. Vazquez-Guardado, H. Luan, J. Ruban, X. Ning, A. Akhtar, D. Li, et al., *Nature* **2019**, *575*, 7783 473.
- [3] D. Rus, M. T. Tolley, *Nature* **2015**, *521*, 7553 467.
- [4] J. Tapia, E. Knoop, M. Mutný, M. A. Otaduy, M. Bächer, *Soft Robotics* **2020**, *7*, 3 332.
- [5] J. Shintake, V. Cacucciolo, D. Floreano, H. Shea, *Advanced Materials* **2018**, *30*, 29 1707035.
- [6] H. Cui, D. Yao, R. Hensleigh, H. Lu, A. Calderon, Z. Xu, S. Davaria, Z. Wang, P. Mercier, P. Tarazaga, X. R. Zheng, *Science* **2022**, *376*, 6599 1287.
- [7] X. Wang, Z. Meng, C. Q. Chen, *Advanced Science* **2023**, *10*, 13 2206637.
- [8] G. Li, X. Chen, F. Zhou, Y. Liang, Y. Xiao, X. Cao, Z. Zhang, M. Zhang, B. Wu, S. Yin, et al., *Nature* **2021**, *591*, 7848 66.
- [9] T. Wu, Y. Dong, X. Liu, X. Han, Y. Xiao, J. Wei, F. Wan, C. Song, *Materials & Design* **2024**, *238* 112629.
- [10] P. Jiao, A. H. Alavi, *International Materials Reviews* **2021**, *66*, 6 365.
- [11] Q. Ma, T. J. Cui, *PhotoniX* **2020**, *1*, 1 1.
- [12] M. Cianchetti, C. Laschi, A. Menciassi, P. Dario, *Nature Reviews Materials* **2018**, *3*, 6 143.

- [13] H. Wang, M. Totaro, L. Beccai, *Advanced Science* **2018**, *5*, 9 1800541.
- [14] Z. Lin, Z. Wang, W. Zhao, Y. Xu, X. Wang, T. Zhang, Z. Sun, L. Lin, Z. Peng, *Advanced Intelligent Systems* **2023**, *5*, 5 2200329.
- [15] A. Xu, A. K. Mishra, H. Bai, C. A. Aubin, L. Zullo, R. F. Shepherd, *Science Robotics* **2019**, *4*, 34 eaaw6304.
- [16] N. Wettels, V. J. Santos, R. S. Johansson, G. E. Loeb, *Advanced Robotics* **2008**, *22*, 8 829.
- [17] R. Patel, R. Cox, N. Correll, *Autonomous Robots* **2018**, *42* 1443.
- [18] G. Li, S. Liu, L. Wang, R. Zhu, *Science Robotics* **2020**, *5*, 49 eabc8134.
- [19] H. Courtecuisse, J. Allard, P. Kerfriden, S. P. Bordas, S. Cotin, C. Duriez, *Medical Image Analysis* **2014**, *18*, 2 394.
- [20] E. Coevoet, A. Escande, C. Duriez, *IEEE Robotics and Automation Letters* **2017**, *2*, 3 1413.
- [21] M. Bonnet, A. Constantinescu, *Inverse Problems* **2005**, *21*, 2 R1.
- [22] S. Andrieux, T. N. Baranger, *Inverse Problems* **2015**, *31*, 11 115003.
- [23] M. Hematiyan, A. Khosravifard, Y. Shiah, *International Journal of Solids and Structures* **2017**, *106-107* 240.
- [24] S. Zhang, Z. Chen, Y. Gao, W. Wan, J. Shan, H. Xue, F. Sun, Y. Yang, B. Fang, *IEEE Sensors Journal* **2022**, *22*, 22 21410.
- [25] Q. Li, O. Kroemer, Z. Su, F. F. Veiga, M. Kaboli, H. J. Ritter, *IEEE Transactions on Robotics* **2020**, *36*, 6 1619.
- [26] K. Shimonomura, *Sensors* **2019**, *19*, 18.
- [27] J. Zhang, H. Yao, J. Mo, S. Chen, Y. Xie, S. Ma, R. Chen, T. Luo, W. Ling, L. Qin, et al., *Nature communications* **2022**, *13*, 1 5076.
- [28] M. Cheng, X. Huang, C. Ma, Y. Yang, *Journal of Micromechanics and Microengineering* **2009**, *19*, 11 115001.
- [29] T. M. Huh, H. Choi, S. Willcox, S. Moon, M. R. Cutkosky, *IEEE Robotics and Automation Letters* **2020**, *5*, 2 2562.
- [30] A. Yamaguchi, C. G. Atkeson, *Advanced Robotics* **2019**, *33*, 14 661.
- [31] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer, D. Jayaraman, R. Calandra, *IEEE Robotics and Automation Letters* **2020**, *5*, 3 3838.
- [32] U. H. Shah, R. Muthusamy, D. Gan, Y. Zweiri, L. Seneviratne, *Journal of Intelligent & Robotic Systems* **2021**, *102* 1.
- [33] O. C. Kara, N. Ikoma, F. Alambeigi, In *2022 IEEE Sensors*. IEEE, **2022** 1–4.
- [34] J.-T. Lee, D. Bollegala, S. Luo, In *2019 IEEE International Conference on Robotics and Automation (ICRA)*. **2019** 4276–4282.
- [35] S. Cui, R. Wang, J. Hu, J. Wei, S. Wang, Z. Lou, *IEEE Transactions on Industrial Electronics* **2022**, *69*, 6 6015.
- [36] H. Sun, K. J. Kuchenbecker, G. Martius, *Nature Machine Intelligence* **2022**, *4*, 2 135.

- [37] I. Andrussov, H. Sun, K. J. Kuchenbecker, G. Martius, *Advanced Intelligent Systems* **2023**, *5*, 8 2300042.
- [38] W. Yuan, S. Dong, E. H. Adelson, *Sensors* **2017**, *17*, 12 2762.
- [39] K. Kamiyama, K. Vlack, T. Mizota, H. Kajimoto, K. Kawakami, S. Tachi, *IEEE Computer Graphics and Applications* **2005**, *25*, 1 68.
- [40] F. Guo, C. Zhang, Y. Yan, P. Li, Z. Wang, In *IEEE International Symposium on Industrial Electronics (ISIE)*. **2016** 1252–1257.
- [41] A. Yamaguchi, C. G. Atkeson, In *IEEE-RAS International Conference on Humanoid Robotics (Humanoids)*. **2017** 241–248.
- [42] A. Alspach, K. Hashimoto, N. Kuppuswamy, R. Tedrake, In *IEEE International Conference on Soft Robotics (RoboSoft)*. **2019** 597–604.
- [43] N. F. Lepora, *IEEE Sensors Journal* **2021**, *21*, 19 21131.
- [44] W. K. Do, M. Kennedy, In *International Conference on Robotics and Automation (ICRA)*. **2022** 6188–6194.
- [45] M. Li, Y. H. Zhou, T. Li, Y. Jiang, *IEEE Robotics and Automation Letters* **2023**, *8*, 9 6021.
- [46] M. R. Cutkosky, J. Ulmen, *The Human Hand as an Inspiration for Robot Hand Development*, chapter Dynamic Tactile Sensing, Springer, Cham, **2014**.
- [47] V. Kakani, X. Cui, M. Ma, H. Kim, *Sensors* **2021**, *21*, 5 1920.
- [48] T. G. Thuruthel, B. Shih, C. Laschi, M. T. Tolley, *Science Robotics* **2019**, *4*, 26 eaav1488.
- [49] Y. Narang, B. Sundaralingam, M. Macklin, A. Mousavian, D. Fox, In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. **2021** 6444–6451.
- [50] L. Scimeca, J. Hughes, P. Maiolino, F. Iida, *IEEE Robotics and Automation Letters* **2019**, *4*, 3 2479.
- [51] H. Yousef, M. Boukallel, K. Althoefer, *Sensors and Actuators A: Physical* **2011**, *167*, 2 171.
- [52] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, M. J. Marín-Jiménez, *Pattern Recognition* **2014**, *47*, 6 2280.
- [53] O. Pfaff, S. Simeonov, I. Cirovic, P. Stano, *Annals of DAAAM & Proceedings* **2011**, *22*, 1 1247.
- [54] L. Yang, F. Wan, H. Wang, X. Liu, Y. Liu, J. Pan, C. Song, *IEEE Robotics and Automation Letters* **2020**, *5*, 2 1720.
- [55] F. Wan, H. Wang, J. Wu, Y. Liu, S. Ge, C. Song, *IEEE Robotics and Automation Letters* **2020**, *5*, 3 4210.
- [56] K. Elgeneidy, A. Fansa, I. Hussain, K. Goher, In *IEEE International Conference on Soft Robotics (RoboSoft)*. **2020** 779–784.
- [57] L. Chen, C. Yang, H. Wang, D. T. Branson, J. S. Dai, R. Kang, *Mechanism and Machine Theory* **2018**, *130* 109.
- [58] R. Kang, Y. Guo, L. Chen, D. T. Branson, J. S. Dai, *IEEE/ASME Transactions on Mechatronics* **2016**, *22*, 2 751.
- [59] P. Peltzer, J. Lotz, U. Naumann, In *International Conference on Computational Science*. **2020** 690–704.

- [60] A. Ghafoor, J. S. Dai, J. Duffy, *Journal of Mechanical Design* **2004**, *126*, 4 646.
- [61] S. Cotin, G. Mestdagh, Y. Privat, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2024**, *480*, 2281 20230197.
- [62] K. Xu, E. Darve, *Journal of Computational Physics* **2022**, *453* 110938.
- [63] J. Nocedal, S. J. Wright, *Numerical Optimization*, Springer New York, NY, USA, 2nd edition, **2006**.
- [64] J. G. De Gooijer, R. J. Hyndman, *International Journal of Forecasting* **2006**, *22*, 3 443.
- [65] D. Chicco, M. Warrens, G. Jurman, *PeerJ. Computer Science* **2021**, *7* e623.
- [66] H. Zhang, X. Guo, J. Wu, D. Fang, Y. Zhang, *Science Advances* **2018**, *4*, 6 eaar8535.
- [67] T. Chai, R. R. Draxler, *Geoscientific Model Development* **2014**, *7*, 3 1247.
- [68] Z. Sun, Z. Zhang, C. Lee, *Nature Electronics* **2023**, *6* 941.
- [69] H. Yuan, S.-L. Hu, C.-C. Lao, Q.-L. Zhang, B.-K. Liang, In *International Conference on Information Technology and Computer Application (ITCA)*. **2019** 39–42.
- [70] E. Matheson, R. Minto, E. G. G. Zampieri, M. Faccio, G. Rosati, *Robotics* **2019**, *8*, 4.
- [71] N. Guo, X. Han, X. Liu, S. Zhong, Z. Zhou, J. Lin, J. Dai, F. Wan, C. Song, *Advanced Intelligent Systems* **2024**, *6*, 1 2300382.
- [72] M. Rabinovich, R. Poranne, D. Panozzo, O. Sorkine-Hornung, *ACM Transactions on Graphics* **2017**, *36*, 2.
- [73] J. Smith, S. Schaefer, *ACM Transactions on Graphics* **2015**, *34*, 4.
- [74] E. Tretschek, N. Kairanda, M. BR, R. Dabral, A. Kortylewski, B. Egger, M. Habermann, P. Fua, C. Theobalt, V. Golyanik, *Computer Graphics Forum* **2023**, *42*, 2 485.
- [75] M. E. Yumer, S. Chaudhuri, J. K. Hodgins, L. B. Kara, *ACM Transactions on Graphics* **2015**, *34*, 4 1.
- [76] J. Bonet, A. Gil, R. Wood, *Nonlinear Solid Mechanics for Finite Element Analysis: Statics*, Cambridge University Press, **2016**.
- [77] J. Liu, Z. Wang, *Journal of Computational and Applied Mathematics* **2019**, *362* 596.
- [78] D. Z. Huang, K. Xu, C. Farhat, E. Darve, *Journal of Computational Physics* **2020**, *416* 109491.
- [79] F. S. Sin, D. Schroeder, J. Barbić, *Computer Graphics Forum* **2013**, *32*, 1 36.
- [80] D. P. Kingma, J. Ba, In *International Conference on Learning Representations (ICLR)*. **2015** 1–15.

## Table of Contents



This study introduces a vision-based approach to learning proprioceptive interactions by reconstructing the shape and touch of a soft robotic metamaterial during physical engagements. It enables real-time, precise estimations of mesh deformation and touch positions in a virtual environment, showcasing the potential for touch-based digital twin interactions.