Anchoring Morphological Representations Unlocks Latent Proprioception in Soft Robots

Xudong Han, Ning Guo, Ronghan Xu, Fang Wan, and Chaoyang Song**

This research addresses the need for robust proprioceptive methods that capture the continuous deformations of soft robots without relying on multiple sensors that hinder compliance. The authors propose a vision-based deformation learning strategy called latent proprioception, which anchors the robot's overall deformation state to a single internal reference frame tracked by a miniature onboard camera. Through a multi-modal neural network trained on simulated and real data, the authors unify motion, force, and shape measurements into a shared representation in latent codes, inferring unseen states from readily measured signals. The experimental results show that this approach accurately reconstructs full-body deformations and forces from minimal sensing data, enabling soft robots to adapt to complex object manipulation or safe human interaction tasks. The proposed framework exemplifies how a vision-based deformable learning approach can inform and enhance robotics by reducing sensor complexity and preserving mechanical flexibility. The authors anticipate that such hybrid system codesign will advance robotic capabilities, deepen the understanding of natural movement, and potentially translate back into healthcare and wearable technologies for living beings. This work paves the way for soft robots endowed with greater autonomy and resilience. All codes are available at GitHub: https://github.com/ancorasir/ProSoRo.

1. Introduction

Soft robotics has emerged as a transformative field that leverages highly deformable materials, such as elastomers, gels, and flexible polymers, to construct robots capable of safe and adaptive interactions with complex environments.[1–5] These soft robots excel in tasks requiring compliance and adaptability, allowing them to conform to unstructured terrains,[6–9] manipulate delicate objects,[10–13] and perform challenging functions for rigid robots.[14–17] Applications span from medical and wearable devices that safely interact with human tissue[18–20] to exploration robots that navigate unpredictable terrains.[21,22] The inherent compliance of soft materials enhances safety and enables a new class of robotic functionalities unattainable with traditional rigid designs.[1,5,23]

Despite these advantages, soft robots' continuous and high-dimensional deformations present significant challenges in modeling, control, and sensing.[24,25] Traditional robotic proprioception relies on joint encoders and discrete sensors to measure positions and forces at specific points, providing precise control over rigid structures.[26–28] In contrast, soft robots lack discrete joints and exhibit infinite degrees of freedom, making it difficult to accurately measure and model their state using conventional rigid-body models and sensing techniques.[29,30] The nonlinear and often unpredictable behavior of soft materials complicates the development of accurate models necessary for control and sensing, hindering the deployment of soft robots in applications.[24,29,31]

To address these challenges, researchers have explored embedding numerous sensors within the soft materials, such as stretchable strain sensors,[32,33] fiber optics,[34,35] or conductive composites,[36,37] to capture detailed deformation information throughout the robot's body.[38,39] While effective in increasing observability, this approach can disrupt the mechanical properties of the soft materials, add significant complexity, and reduce the overall robustness of the system due to sensor fragility.[30,40] Another avenue involves leveraging advancements in machine learning to infer internal states from limited sensor data.[41,42] Deep learning and latent space modeling have been employed to extract meaningful representations from high-dimensional sensory inputs.[43–45] Latent representations compress complex deformation data into lower-dimensional spaces, capturing

X. Han, R. Xu, F. Wan
School of Design
Southern University of Science and Technology
Shenzhen 518055, P. R. China
E-mail: wanfang@ieee.org

N. Guo
School of Artificial Intelligence
Shanghai Jiao Tong University
Shanghai 200240, P. R. China

C. Song
Department of Mechanical and Energy Engineering
Southern University of Science and Technology
Shenzhen 518055, P. R. China
E-mail: songcy@ieee.org

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aisy.202500444.

DOI: 10.1002/aisy.202500444

essential features without exhaustive sensing.[46–48] However, these methods often require extensive training data and computational resources and may lack interpretability, limiting their practicality for real-time control and interaction in soft robotics.[38,41] Recent works[49–51] on variational autoencoder present their potential applied to multimodal learning, but mainly limited to supervised image data and lack of relation analysis among latent codes and explicit modalities.

This paper introduces a new proprioceptive method for soft robots by anchoring their morphological representations to a single internal reference frame. At the core of this approach is ProSoRo, a generalizable method that encodes the interactions of the robot's deformable structure by tracking a single internal reference frame using miniature vision. The method captures essential information about overall robot deformation while significantly reducing sensing complexity by tracking how the reference frame moves relative to a fixed boundary. To interpret these measurements, a multimodal variational autoencoder (MVAE) unifies motion, force, and shape data into a shared latent space called latent proprioception. This enables crossmodal inference, allowing the estimation of internal forces and full-body deformations from more straightforward motion-based measurements. Within the latent representation, the robot's fundamental morphing primitives naturally emerge, offering interpretable insights into its deformation patterns. Validation in simulations and real-world experiments demonstrates robust force and shape estimation across various soft robotic structures with high sim-to-real transferability. ProSoRo's capabilities are further showcased through the digitalization and synthesis of complex movements, as well as the execution of sequential contact reasoning during manipulation tasks using a soft-rigid hybrid arm. Altogether, this method provides a scalable and practical solution for soft robot proprioception, bridging the gap between continuous deformations and efficient control strategies and opening avenues for advanced, adaptive robotic systems. The main contributions of this work are listed as follows: A generalizable vision-based approach encoding the interactions of the robot's deformable structure by tracking a single internal reference frame, drastically simplifying sensing requirements while preserving essential information about the deformation state. A multimodal variational autoencoder (MVAE) model integrating motion, force, and shape data into a shared latent space, enabling crossmodal inference and interpretable morphing primitives. Systematic validation on six different structures in both simulated and real-world settings shows robust sim-to-real transfer and effectiveness with these soft robot modules demonstrated in a soft-rigid hybrid arm for delicate manipulation.

Section 2 proposes the method to anchor soft robotic motion to a single internal frame, leading to the design and characterization of ProSoRos. Section 3 introduces a multimodal variational autoencoder (MVAE) that fuses motion, force, and shape data into shared latent codes capturing key morphing primitives. Section 4 evaluates Sim2Real transfer by assessing the MVAE's shape and force estimations. In Section 5, we demonstrate the utility of latent proprioception using a tendon-driven platform. Section 6 showcases sequential contact reasoning in a soft-rigid hybrid arm during pivoting manipulation. Finally, the concluding section discusses this work's limitations, potential refinements, and implications.

2. Anchoring Soft Robotic Motion for Proprioception

2.1. Design & Fabrication of ProSoRo

This study introduces an anchor-based approach that leverages a single internal reference frame to infer the full proprioceptive state of a soft robot, encompassing motion, force, and shape. Our design strategy begins with fundamental geometric primitives, cylinders, octagonal prisms, and quadrangular prisms, foundational elements for soft robotic structures in **Figure 1**(a). The proprioceptive soft robot (ProSoRo) consists of the top frame, marker, metastructure, bottom frame, fill light, camera, indicator light, and mounting base, shown as Figure 1(b). The top frame, bottom frame, and mounting base are fabricated using resin by stereolithography (SLA) 3D printing. The camera (WX605, VISHINSGAE) is mounted in the bottom frame, with an FOV of 150°, a resolution of 1280720 × pixels, and a frame rate of 120 frames per second (fps). By manually adjusting the lens, the focus position falls on the marker (ArUco) pasted on the top frame. The ambient and indicator lights are printed circuit boards (PCBs) with several programmable light-emitting diodes (LEDs).

We fabricated prototypes of six ProSoRo variants to achieve transparency and compliance using clear silicone and polyurethane (Supplementary Movie S1). Figure 1(c) shows prototypes of ProSoRos and images captured by inner cameras. The cylinder, octagonal prism, and quadrangular prism are molded with clear silicone (Solaris, Smooth-On) whose hardness is 15 A, while the omni-neck, origami, and dome are molded with polyurethane rubber (Hei-Cast 8400, H & K) whose hardness is 70 A. They are all molded following similar steps shown in Figure 1(d): first, the mold box is assembled, several pour spouts are placed at the bottom of the box, and a 3D-printed model is pasted on them; second, the silicone rubber is poured into the box and completely envelops the model; third, after the mold cures, the model and the spouts are removed, and the mold is cleaned preparing for molding; fourth, the mold is tied up with a strap and the soft material selected for the metastructure is poured into the mold; finally, after the material cures, we get the metastructure. The metastructure is bonded with the top and the bottom frames using epoxy adhesive, and other parts are assembled with fastening pieces.

These structures exhibit intricate deformation patterns and mechanical responses that are challenging to model and control using conventional methods. The idea is to anchor the robot's morphological representation to a single reference frame within the structure, depicted as a purple point in Figure 1a. This reference frame is a surrogate for the robot's overall deformation state, capturing essential information about its motion relative to a fixed boundary. We can infer high-dimensional proprioceptive data from low-dimensional observations by establishing a consistent relationship between the anchor frame's motion, the forces exerted at the base, and the global shape deformation. We developed Proprioceptive Soft Robots (ProSoRo), an integrated system that combines soft materials with embedded sensing capabilities to realize this concept. Each ProSoRo features a metastructure mounted between a top and bottom frame, with a marker affixed to the top frame as the anchor frame. A miniature monocular camera in the bottom frame tracks the marker's

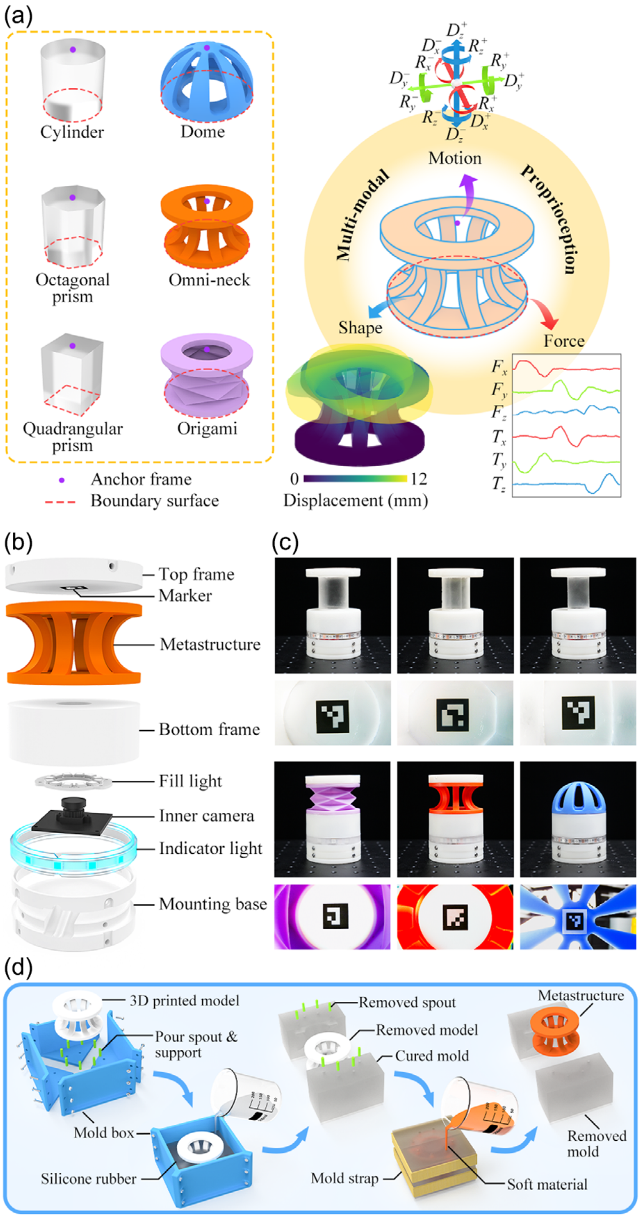

Figure 1. ProSoRo's vision-based anchoring design. a) Exemplified designs of ProSoRo with their proprioception anchored by a single reference frame. b) Integrated design to visually track the spatial deformation anchored at a marker inside. c) Prototypes of ProSoRos and images captured by inner cameras. d) Fabrication process of the ProSoRo's metastructure.

real-time movement. Including programmable fill lights ensures consistent illumination, enhancing the reliability of visual tracking, while indicator lights provide immediate feedback on motion direction and magnitude.

2.2. Mechanical Robustness Characterization

To assess the mechanical robustness of these designs, we conducted a half-million-cycle compression test on the omni-neck ProSoRo. The force versus time curve remained stable between −60 and 20 N throughout the test, demonstrating excellent durability and suitability for long-term applications. The cyclic compression test of the omni-neck ProSoRo was measured with a linear dynamic test (ElectroPuls E3000, Instron) under a cyclic frequency of 5 Hz. The forces along the *z*-axis were collected under a maximum compressing displacement of 3 mm during 0.5 million cycles, plotted versus time as shown in **Figure 2**a.

To test the performance under a single translation or rotation, we used a robot arm (UR10e, Universal Robots) to drive the top surface of the six metastructures, translating and rotating along the *x*-axis. The forces were measured by a force/torque sensor (Nano 25, ATI) mounted under the metastructures, and the values of translations or rotations were read from the robot arm in Figure 2b. Simple geometries exhibited approximately linear force–displacement relationships, whereas complex metastructures like the origami, omni-neck, and dome displayed pronounced nonlinear behaviors. These nonlinearities arise from factors such as hyperelastic material properties, large deformation ranges, and geometric complexities,[5,24] highlighting the limitations of traditional models and underscoring the need for advanced proprioceptive frameworks. Anchoring the morphological representation to a single internal frame simplifies the sensing and computational requirements associated with soft robots. By reducing the dimensionality of the data needed to characterize the robot's state, we circumvent the challenges posed by infinite degrees of freedom and continuous deformation.

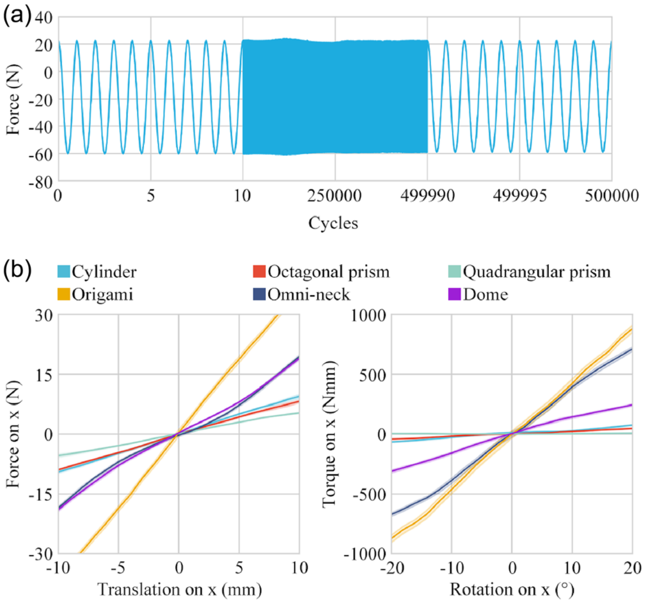

Figure 2. Mechanical robustness and force–displacement behavior of the ProSoRos. a) A half-a-million-cycle compression test along the omni-neck's *z*-axis, demonstrating reliable performance over repeated loading. b) Force–displacement curves for single translation along and rotation about the *x*-axis, revealing distinct force and torque responses under different deformation modes for various ProSoRos.

2.3. A Multimodal Learning Framework for ProSoRo

To harness the full potential of this anchor-based approach, we developed a multimodal learning framework to align ProSoRos' motion, force, and shape into a unified representation based on an anchored observation. It involves three stages of material identification, latent proprioceptive learning, and crossmodal inference in **Figure 3**.

2.3.1. Material Identification with Evolutionary Optimization

Recognizing the impracticality of collecting extensive physical datasets for soft robots, we leveraged finite element analysis (FEA) simulations to generate high-quality training data in the first stage. To evaluate the material properties in the metastructure, stress versus strain curves were measured through uniaxial tensile testing with an electromechanical universal test system (E45.105, MTS). We fitted the curves using different hyperelastic models. To minimize the error of mechanical performance compared with real-world values, an evolution strategy optimization algorithm for material identification has been developed based on the covariance matrix adaptation evolution strategy (CMA-ES).[52] We first deformed the physical ProSoRo to 10 different poses using a robotic arm (Panda, Franka Emika) and recorded boundary conditions on the top surface of the ProSoRo by the tool central point (TCP) of the robot arm and the force on the bottom surface by a force/torque sensor (Nano25, ATI). In Figure 3a, the initial parameters of the material model and the anchor frame's motion from the physical experiment are combined to generate FEA results. The force $[\mathbf{F}, \mathbf{T}]^{\mathrm{T}}$ read from the FEA results and the ground truth $[\mathbf{F}_{\mathrm{gt}}, \mathbf{T}_{\mathrm{gt}}]^{\mathrm{T}}$ measured from physical experiments are compared to calculate the normalized error for updating the material parameters by the ES sampler. After several trials, if the normalized error E is less than the expected threshold E_{e}, or if the number of trials reaches the maximum value, the optimization should be stopped, and the final optimized parameters are obtained. FEA error comparison between with and without optimization can be found in **Figure 4**, showing the effectiveness of reducing the sim-to-real errors, which critically affects the sim-to-real transferability during the following stages.

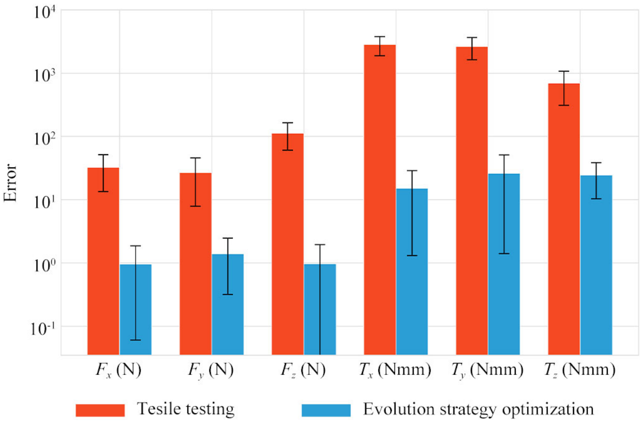

Figure 4. FEA errors on evolutionary optimization for material parameters.

2.3.2. Latent Proprioceptive Learning via MVAE

At the *second* stage, the simulation dataset was generated using the optimized material parameters, which consists of the anchor frame motion $[D_{\mathrm{x}}, D_{\mathrm{y}}, D_{\mathrm{z}}, R_{\mathrm{x}}, R_{\mathrm{y}}, R_{\mathrm{z}}]^{\mathrm{T}}$, the resultant force

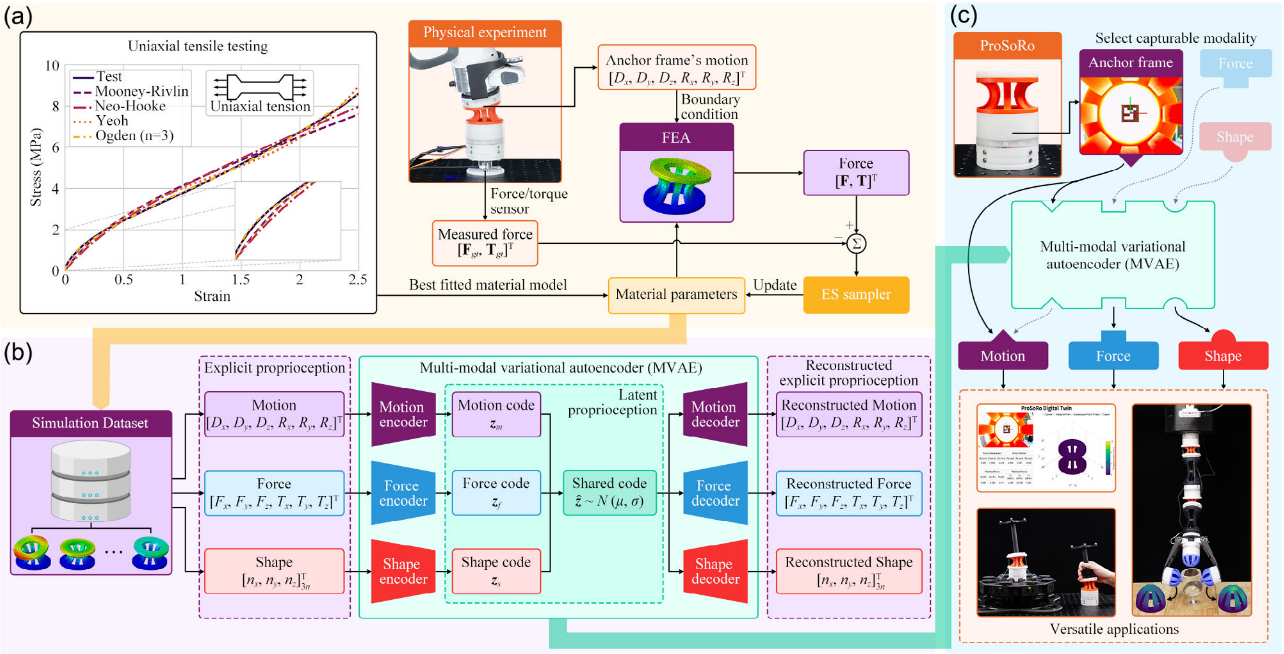

Figure 3. Overview of ProSoRo's learning framework for anchored proprioception in three stages: a) material identification with the evolution strategy optimization, b) latent proprioceptive learning through multimodal variational autoencoder, and c) cross-modal inference for versatile applications.

$\left[F_x, F_y, F_z, T_x, T_y, T_z\right]^T$, and shape deformation represented by displacements of nodes $\left[n_x, n_y, n_z\right]_{3n}^T$ where n is the number of simulated nodes, as the training inputs. To learn these modalities of explicit proprioception, we developed a multimodal variational autoencoder (MVAE) to encode the ProSoRo's proprioception via latent codes that contain the unified representation of proprioception, which can be further decoded to all modalities.

As shown in Figure 3b, the explicit proprioception $\{(\mathbf{x}_i^m, \mathbf{x}_i^f, \mathbf{x}_i^s)\}_{i=1}^N$ are independent and identically distributed as the inputs of MVAE, where $\mathbf{x}^m$, $\mathbf{x}^f$, and $\mathbf{x}^s$ are motion, force, and shape, respectively. We first utilized three encoders $q_{\boldsymbol{\phi}}^m$, $q_{\boldsymbol{\phi}}^f$, and $q_{\boldsymbol{\phi}}^s$ to extract latent codes $\mathbf{z}_i^m$, $\mathbf{z}_i^f$, and $\mathbf{z}_i^s \in \mathbb{R}^{N \times D}$, respectively

$$\mathbf{z}_i^n = q_{\boldsymbol{\phi}}^n(\mathbf{x}_i^n), \quad n \in \{m, f, s\} \tag{1}$$

where $\boldsymbol{\phi}$ is the parameters of encoders, N is the size of data pairs, and D is the latent dimension. Similar to the origin variational autoencoder,[53] the prior over the latent code is assumed to be the centered isotropic multivariate Gaussian $p_{\boldsymbol{\theta}}^n(\mathbf{z}^n) = \mathcal{N}(\mathbf{z}^n; 0, \mathbf{I})$. We let the variational approximate posterior be a multivariable Gaussian with a diagonal covariance

$$\log q_{\boldsymbol{\phi}}^n(\mathbf{z}^n|\mathbf{x}^n) = \log \mathcal{N}(\mathbf{z}^n; \boldsymbol{\mu}^n, \boldsymbol{\sigma}^n\mathbf{I}), \quad n \in \{m, f, s\} \tag{2}$$

where $\boldsymbol{\mu}^n$ and $\boldsymbol{\sigma}^n$ are the mean and standard deviation of the approximate posterior, which are outputs of the encoders. To encourage the approximate posterior to be close to the prior, the Kullback–Leibler (KL) divergence $D_{\mathrm{KL}}(q_{\boldsymbol{\phi}}^n(\mathbf{z}^n|\mathbf{x}^n) \| p_{\boldsymbol{\theta}}^n(\mathbf{z}^n))$, $n \in \{m, f, s\}$ is introduced and can be reformatted as

$$D_{\mathrm{KL}}(\mathcal{N}(\mathbf{z}^n; \boldsymbol{\mu}^n, \boldsymbol{\sigma}^n\mathbf{I}) \| \mathcal{N}(\mathbf{z}^n; 0, \mathbf{I})), \quad n \in \{m, f, s\} \tag{3}$$

Since the data pairs $\{(\mathbf{x}_i^m, \mathbf{x}_i^f, \mathbf{x}_i^s)\}_{i=1}^N$ described the same semantics of ProSoRo with different modalities, the three latent codes $\mathbf{z}_i^m$, $\mathbf{z}_i^f$, and $\mathbf{z}_i^s$ should be aligned to one shared code $\bar{\mathbf{z}}_i$ by minimizing the discrepancies

$$\min \ \|\bar{\mathbf{z}}_i - \mathbf{z}_i^n\|_2^2, \quad n \in \{m, f, s\}. \tag{4}$$

The shared code $\bar{\mathbf{z}}_i$ contained the fused features from all three modalities, denoted as latent proprioception, and could reconstruct origin modalities of explicit proprioception through three specific decoders $p_{\boldsymbol{\theta}}^m$, $p_{\boldsymbol{\theta}}^f$, and $p_{\boldsymbol{\theta}}^s$, respectively

$$\hat{\mathbf{x}}_i^n = p_{\boldsymbol{\theta}}^n(\bar{\mathbf{z}}_i), \quad n \in \{m, f, s\} \tag{5}$$

where $\boldsymbol{\theta}$ is the parameters of decoders. Finally, the objective function contained three parts, including reconstruction loss, KL divergence, and latent loss

$$\begin{aligned}\mathscr{L}(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{x}_i^n) = {} & \underbrace{\|\mathbf{x}_i^n - \hat{\mathbf{x}}_i^n\|_2^2}_{\text{reconstruction loss}} - \underbrace{\alpha D_{\mathrm{KL}}\left(\mathcal{N}\left(\boldsymbol{\mu}_i^n, \boldsymbol{\sigma}_i^n\mathbf{I}\right) \| \mathcal{N}(0, \mathbf{I})\right)}_{\text{KL divergence}} \\ & + \underbrace{\beta\|\bar{\mathbf{z}}_i - \mathbf{z}_i^n\|_2^2}_{\text{latent loss}}\end{aligned} \tag{6}$$

In all experiments, α is 0.1, and β is 1. The reconstruction loss guarantees that the shared code encoded from all modalities can still be decoded to each modality. The KL divergence is used to avoid posterior collapse. The latent loss helps converge all modal information into one unified code shared among different modalities. Therefore, we formulated MVAE with multimodal input and a shared code containing the latent proprioception knowledge of ProSoRo. During training, the latent code is encoded from all modalities and decoded to reconstruct the original features. However, during the deployment, it is flexible to encode from one selected modality and decode to others, realizing crossmodal inference.

2.3.3. Crossmodal Inference for Versatile Applications

Based on the advances of MVAE, we introduce a third stage to enable crossmodal inference where we could use one observed modality as input to infer the other two modalities. As an example of real-world deployment, the anchor frame motion is easy to capture by computer vision. Hence, we use it to obtain $\mathbf{z}_m$ in the shared latent space and estimate the shape modality, which is usually challenging in real-time interactions in soft robotics. We visually capture the ProSoRo's anchor frame as MVAE's input to estimate the force and shape modalities based on the latent knowledge learned from simulation in Figure 3c.

3. Learning Latent Proprioception with Key Morphing Primitives

3.1. Training a Multimodal Variational Autoencoder

Understanding and controlling the complex deformation behaviors of soft robots requires an efficient representation of their high-dimensional proprioceptive states. The multimodal variational autoencoder (MVAE) serves this purpose by encoding motion, force, and shape data into a shared latent space, referred to as latent proprioception. Using the optimized simulation dataset, we trained the MVAE to learn the nonlinear relationships among the proprioceptive modalities. To train the MVAE, we generated simulation data using FEA. The material model was neo-Hookean, whose parameters were based on the optimization results. The mesh type was a four-node tetrahedral element (C3D4). For each ProSoRo, the bottom surface was configured to be completely fixed, and the anchor frame was randomly assigned 100 000 distinct motions (ranging: D_x of (−10, 10) mm, D_y of (−10, 10) mm, D_z of (−3, 3) mm, R_x of (−0.3, 0.3) rad, R_y of (−0.3, 0.3) rad, and R_z of (−0.3, 0.3) rad), which drove the displacements of coupled nodes. Both of these constituted the boundary conditions for FEA. Details of FEA setup can be found in Supplementary Table S1, Supporting Information. The motion of the anchor frame, the force on the bottom surface, and the displacements of surface nodes were collected together as datasets. The datasets were standardized and divided into training and testing datasets with a ratio of 8:2. MVAE was trained for 2,000 epochs on a computer with an NVIDIA Tesla V100 GPU, the Adam optimizer, a batch size of 128, and a learning rate of 0.0001.

During inference, the model takes the motion of the anchor frame as input and estimates the corresponding force and shape states. For the omni-neck ProSoRo, the MVAE's force, torque, and shape estimation closely match the ground truth from the testing dataset in **Figure 5**, achieving R^2 scores above 0.98, 0.99, and 0.99 and low root mean square errors (RMSE) of less than 1.4558 N, 32.6100 Nmm, and 0.1985 mm for forces, torques, and node displacements ranging approximately 36 N, 1, 200 Nmm, and 19 mm.

3.2. Key Morphing Primitives Learned for Omni-Neck ProSoRo

To delve deeper into the latent proprioception space, we analyzed the correlations among the latent codes $\mathbf{z}_\text{m}$, $\mathbf{z}_\text{f}$, and $\mathbf{z}_\text{s}$ generated from motion, force, and shape inputs for omni-neck in **Figure 6**(a). Six components of the latent code, namely z_8, z_{19}, z_{20}, z_{22}, z_{31}, and z_{32}, exhibited high correlation coefficients across the modalities and were identified as "key morphing primitives." The identical set of key morphing primitives in $\mathbf{z}_\text{m}$, $\mathbf{z}_\text{f}$, and $\mathbf{z}_\text{s}$ also indicates that they have been successfully converged to a shared latent space. These primitives are pivotal in encoding the robot's deformation behaviors and are instrumental in cross-modal inference. We visualized the relationships between explicit proprioceptive modalities and the latent codes $\mathbf{z}_\text{m} = [z_1, z_2, \ldots, z_{32}]$ using a chord diagram in Figure 6(b). The width of each connection represents the strength of the correlation to the three modalities, which are uniform and equivalent by normalization, ensuring consistent weights to observe the impact on the latent code. The six key morphing primitive components explain 14.1%, 8.8%, 12.3%, 8.1%, 8.0%, and 13.3% of the variance in the latent proprioception space, respectively, and 64.5% in total. We generated a series of deformation modes for the omni-neck ProSoRo in Figure 6(c) by systematically varying these six latent components.

3.3. Generalization for ProSoRos with Different Designs

The proprioceptive learning performance of all ProSoRo variants tested in this study can be found in **Figure 7**, where all the R^2 scores are above 0.98. This high accuracy confirms the model's ability to capture the essential dynamics of the soft robot's behavior. The error distributions for forces, torques, and shapes across different ProSoRo types follow Gaussian patterns with near-zero means and standard deviation (SD) ranges within [0.17, 1.46] N for forces, [1.46, 46.73] Nmm for torques, and [0.10, 0.20] mm for shapes. While the SDs increased with the structural complexity, the errors remained within acceptable limits for practical applications due to increased nonlinearity and stiffness variations. The performance evaluations indicate that the proposed MVAE framework has successfully learned the three proprioceptive modalities into a shared latent space and could transfer the information from one modality to another, achieving multimodal proprioception inference. The correlation heatmaps and generated meshes for other ProSoRos can be found in Supplementary Figures S1 and S2, Supporting Information. Each key morphing primitive influences the shape in a distinct manifold, providing intuitive interface for manipulating complex deformations.

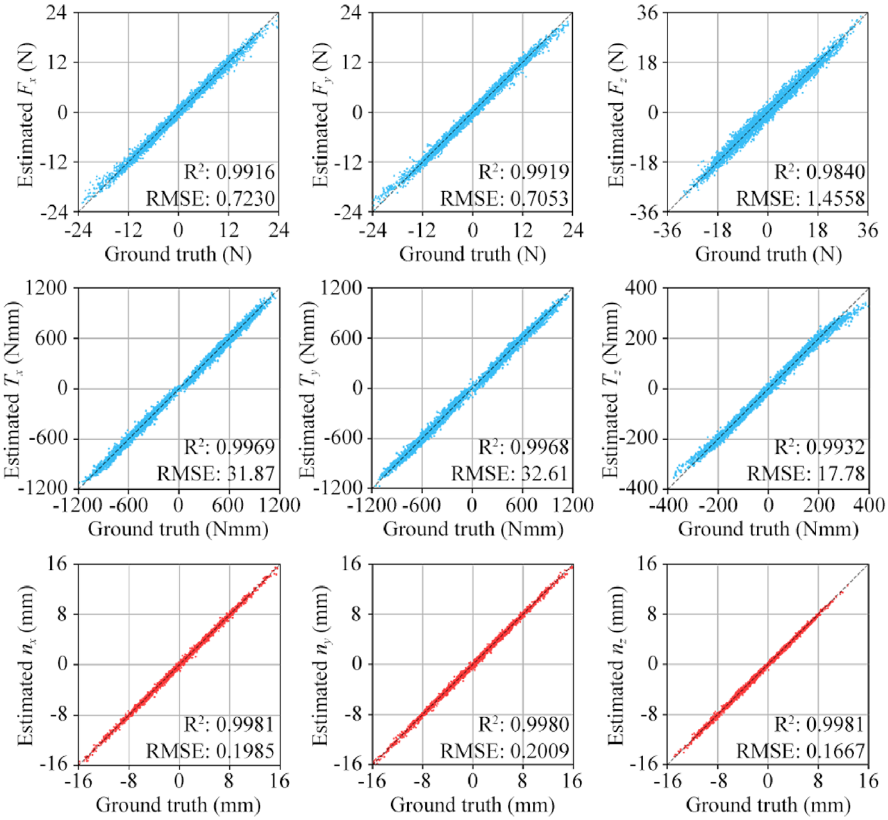

Figure 5. Linear plot comparison of force, torque, and shape (node displacements) estimated by MVAE with motion input versus those from testing data by FEA for omni-neck ProSoRo.

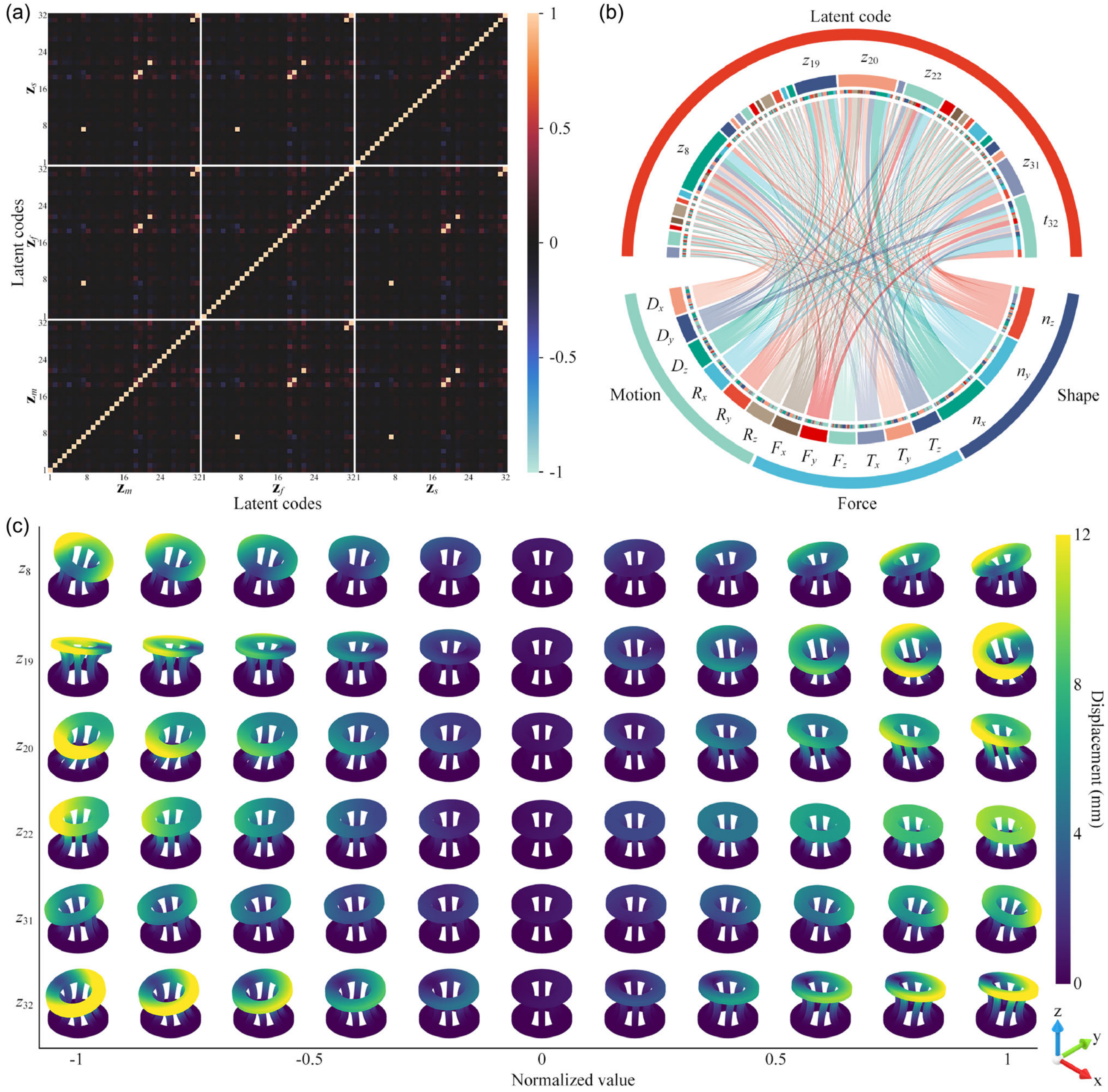

Figure 6. Learning latent proprioception of omni-neck ProSoRo with key morphing primitives. a) Correlation heatmap of latent codes $\mathbf{z}_\mathrm{m}$, $\mathbf{z}_\mathrm{f}$, and $\mathbf{z}_\mathrm{s}$ generated by motion, force, and shape, respectively. b) Chord diagram between the explicit and latent proprioception, of which the components with the six most significant widths are key morphing primitives. c) Mesh results generated by normalized values of the six key morphing primitives.

4. Sim-to-Real Transfer for Crossmodal Inference

4.1. Experiment Setup

The sim-to-real gap is a significant challenge in robotics, and it is essential to ensure that models trained on simulated data perform reasonably well in real-world scenarios.[54–56] To evaluate the MVAE's real-world applicability, we conducted experiments using physical ProSoRo prototypes. Our experimental setup included a robotic arm for controlled manipulation, a 3D scanner for capturing shape data, a camera for tracking the anchor frame, a pose tag for precise motion measurement, and a force/torque sensor for ground truth force data in **Figure** 8a. We fixed the ProSoRo on a force/torque sensor (Nano 25, ATI) and moved the top surface of the ProSoRo to different positions using a robotic arm (Panda, Franka Emika). A pose tag (AprilTag) was fixed to the flange of the robotic arm and tracked by a camera (D435i, Intel RealSense) on the top, representing the motion of the ProSoRo. A 3D scanner (Mini, Revopoint) is placed on a support arm that can rotate 360 degrees around the center

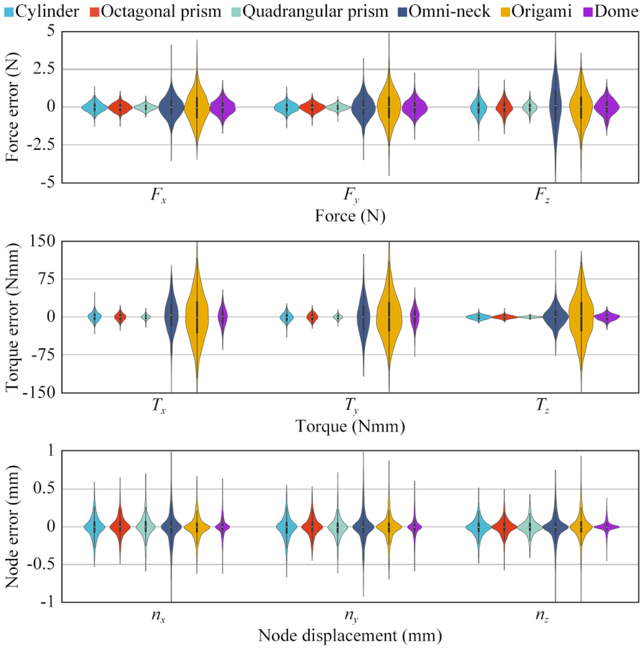

Figure 7. Error distributions of forces, torques, and shapes (node displacements) for different ProSoRos.

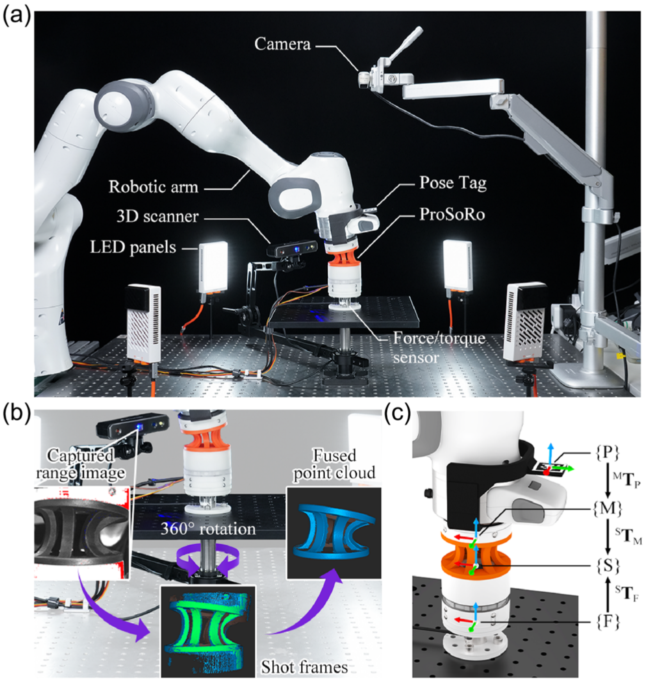

Figure 8. Setup for sim-to-real evaluation of crossmodal inference by MVAE. a) Experiment setup, including robotic arm, 3D scanner, camera, pose tag, force/torque sensor, LED panels, and ProSoRo. b) Range image, single-shot frame (green), overlapping previous point cloud (blue), and fused point cloud of ProSoRo by the 3D scanner. c) Relation among the force/torque sensor's coordinate $\{F\}$, the pose tag's coordinate $\{P\}$, the anchor frame's coordinate $\{M\}$, and the ProSoRo's coordinate $\{S\}$.

to capture point clouds of ProSoRo. We measured the ground truth force using the force/torque sensor. The point clouds were captured under 30 different top surface positions, and the forces were collected when the robotic arm continuously moved for 30 s, respectively. With the ProSoRo's motion measured by the camera and MVAE estimated the shape and force, we compare them with the measured ground truth.

We programmed the robotic arm such that its tool center point (TCP) reached 30 random poses for each ProSoRo and collected corresponding motion and shape data. The ground truth of the ProSoRo's 3D shape is measured by a fused point cloud of its surface, which is obtained by concatenating sequentially captured range images (single-frame accuracy 0.05 mm) while rotating the 3D scanner 360° around the ProSoRo in Figure 8(b). We show the relation among different coordinates, including the force/torque sensor's coordinate $\{F\}$, the pose tag's coordinate $\{P\}$, the anchor frame's coordinate $\{M\}$, and the ProSoRo's coordinate $\{S\}$ in Figure 8(c).

4.2. Results for Morphing Estimations in Shape and Force

The motion of the anchor frame was measured by the pose tag, which is captured by the camera on the top and converted to the ProSoRo's coordinate $\{S\}$. The force/torque sensor readings were converted to $\{S\}$. To compare the shapes estimated by MVAE (purple) and the point clouds (yellow) measured by the 3D scanner in **Figure** 9(a), coarse global registration was first applied based on the relation between the 3D scanner and the ProSoRo's coordinates. Then, the iterative closest point (ICP)

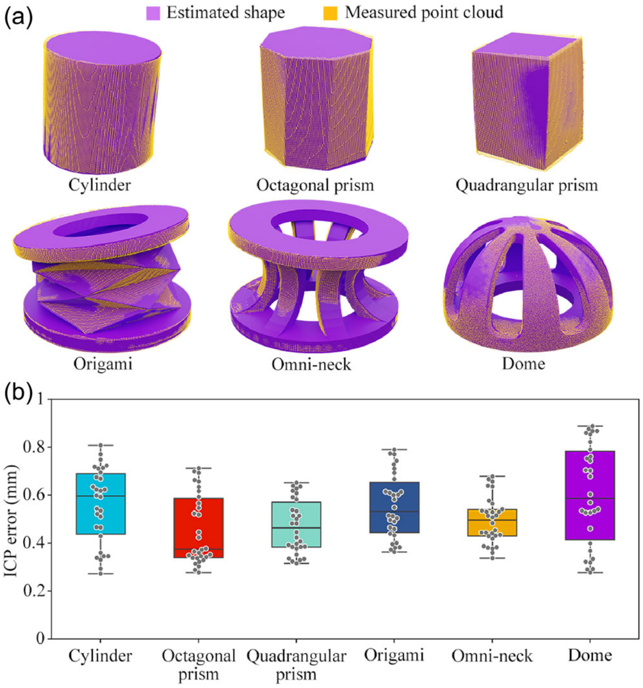

Figure 9. Sim-to-real evaluation of shape estimation by MVAE. a) Examples of shape estimated by MVAE and point cloud measured by the 3D scanner. b) Iterative closest point (ICP) errors between the estimated shape and measured point cloud for the six ProSoRos.

algorithm was used to minimize the difference between the two-point clouds, and the ICP error represented the mean point-to-point distance. We calculated the ICP errors of all cases for six ProSoRos shown as grey points in Figure 9(b), where all ICP errors are lower than 0.9 mm, the mean is between 0.4 to 0.6 mm, and the standard deviation is under 0.2 mm.

4.3. Evaluating Force Estimated Using MVAE

We also pulled the ProSoRo's top surface with a robotic arm. We recorded motion and force/torque in 60 s and found that force/torque estimations from the MVAE closely matched the ground truth measurements. **Figure 10**(a) compares the force estimated by MVAE and the ground truth measured by the force/torque sensor with the omni-neck ProSoRo. Results for other ProSoRos can be found in Supplementary Figure S3, Supporting Information. Although there are some local deviations between the estimated values and the ground truth, the overall trend remains highly synchronized. Figure 10(b) shows the RMSE of force estimation for all six ProSoRos. Through the above evaluation experiments, even though the errors of force/torque and shape become slightly larger from simulation to reality, the results verify that the latent proprioception of MVAE trained by simulation data is transferable to estimate reliable states of ProSoRo in reality, confirming the model's effectiveness in cross-modal inference and Sim2Real transferability.

5. Digitalizing and Synthesizing Omni-Directional Motions

5.1. Experiment Setup with a Tendon-Driven Platform

The capacity to replicate complex motions across different platforms is a powerful tool for advancing soft robotic applications, such as teleoperation and coordinated multirobot systems. We demonstrated this capability by digitizing the motion of a manually operated ProSoRo and synthesizing it on an active ProSoRo mounted on a tendon-driven platform. Our system comprised three synchronized ProSoRos: a manual ProSoRo manipulated by a human operator in **Figure 11**(a), a digital ProSoRo for real-time visualization in Figure 11(b), and an active ProSoRo driven by the tendon platform in Figure 11(c).

The tendon-driven platform for actively driving ProSoRo comprises a mounting base, actuators, tendons, driving drums, and pulleys, as shown in Figure 11(c) and Supplementary Movie S2. The eight actuators (XW540-T140-R, Dynamixel) are uniformly arrayed in a circular configuration and securely fixed to the 3D-printed mounting base. Each actuator has a 3D-printed driving drum installed on its wheel, and all the drums are wrapped with several loops of tendons. The flat shape of the drum is beneficial for maintaining the consistency of the tendons' direction, thereby avoiding friction and enhancing the control precision of the tendons' length. Tendons are polyethylene braided lines with a diameter of 0.36 mm and a strength of 50 lb, which is strong enough to drive the ProSoRo fixed on the center of the mounting base. Unlike the original design shown in Figure 1(b), several circular holes are added to the top and bottom frames of the ProSoRo. Eight pulleys, whose main parts are V-groove bearings with limiters, are fixed on the mounting base along the directions of tendons from the driving drums' outlets. As shown in Figure 11(c), tendons passed through the holes on the bottom

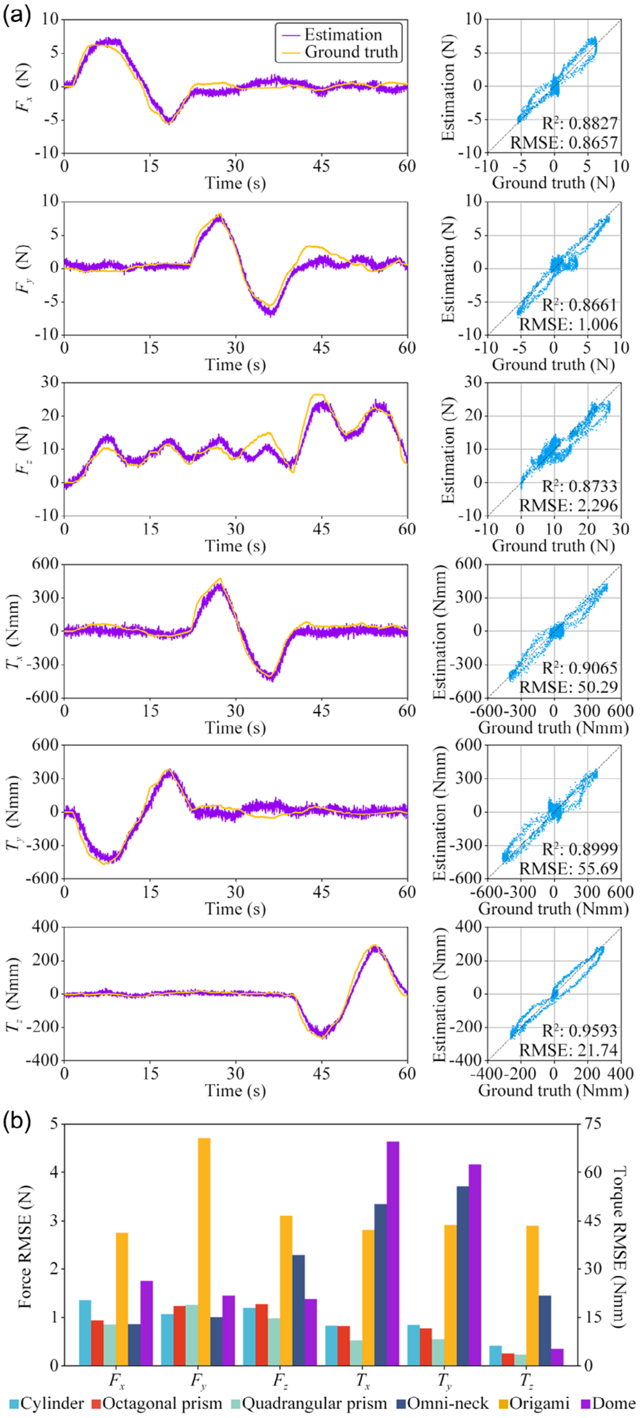

Figure 10. Sim-to-real evaluation of force estimation by MVAE. a) Comparison of force estimated by MVAE and ground truth measured by the force/torque sensor on six dimensions for the omni-neck ProSoRo. b) RMSE of estimated forces on six dimensions for the six ProSoRos.

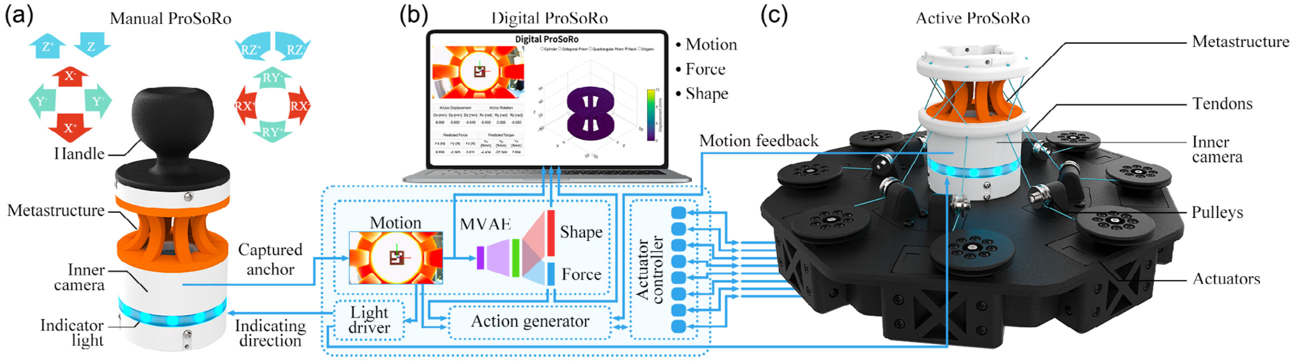

Figure 11. Digitalizing and synthesizing motions of ProSoRos by MVAE. a) Digitalization of ProSoRo's captured anchor motion caused by external manual action. b) Visualization of the proprioception inference on the digital ProSoRo. c) Synthesis of the same omnidirectional motion on the active ProSoRo.

frame, crossed in pairs, and were tied in knots at the holes on the top frame. This tendon layout enables the platform to drive ProSoRo in all six dimensions, including D_x, D_y, D_z, R_x, R_y, and R_z.

The control of the tendon-driven platform, designed for the active manipulation of the ProSoRo, can be articulated as follows: given the observed motion of the manual ProSoRo $\mathbf{x} = \left[D_x, D_y, D_z, R_x, R_y, R_z\right]^{\mathrm{T}} \in \mathbb{R}^6$, the lengths of the tendons of the active ProSoRo $\mathbf{L} = [L_1, L_2, L_3, L_4, L_5, L_6, L_7, L_8]^{\mathrm{T}} \in \mathbb{R}^8$, their relationship can be expressed as

$$\mathbf{L} = f(\mathbf{x}) \qquad f: \mathbb{R}^6 \to \mathbb{R}^8. \tag{7}$$

The calculation of the tendon lengths is given by

$$L_i = \|{}^{S}\mathbf{T}_M \cdot {}^{M}\mathbf{OT}_i - {}^{S}\mathbf{OB}_i\|, \qquad i = 1, 2, \ldots, 8, \tag{8}$$

where ${}^{M}\mathbf{OT}_i$ and ${}^{S}\mathbf{OB}_i$ are fixed vectors from the knots on the top and bottom frame to the origin of their respective reference frames, defined in top frame $\{M\}$ and bottom frame $\{S\}$, and ${}^{S}\mathbf{T}_M$ is the relation between the two frames. Since the motion can be tracked, it is feasible to calculate ${}^{S}\mathbf{T}_M$ and further the tendon lengths $\mathbf{L}$.

To evaluate the performance of the tendon-driven platform, we applied step signals to the platform on six dimensions, respectively, and recorded the motion of the ProSoRo by MoCap (Mars2H, NOKOV), as shown in **Figure 12**. The step translations along the x and y axes vary with a step size of 2 mm, increasing from 0 to 10 mm, then decreasing to -10 mm, and finally returning to 0 mm, while that along the z axis varies with a step size of 1 mm, decreasing from 0 to -5 mm, and then returning to 0 mm. Similarly, the step rotation in each direction varies with a step size of 5°, increasing from 0° to 25°, then decreasing to $-25°$, and finally returning to 0°. A pole with four markers was fixed on the ProSoRo, and MoCap captured the positions of the markers, which were then converted into the motion of the ProSoRo.

5.2. Results for Synthesizing Omnidirectional Motions

When the manual ProSoRo was applied with external action by a human operator, the motion of the anchor frame captured by the inner camera was fed to the MVAE to estimate real-time force/torque and shape. The motion was also sent to the light driver to show the direction and magnitude with the indicator light of ProSoRo as visual feedback to the operator. Detailed proprioception is displayed in digital ProSoRo, a user interface that synchronously receives the motion, force/torque, and shape from MVAE. At the same time, the action of the tendon-driven platform was generated based on motion and force from MVAE and executed through actuators to drive the active ProSoRo to the target motion. To compare the manual and active ProSoRo motions, we added markers to them, as shown in **Figure 13**(a), and recorded their trajectories using the motion capture (MoCap) system. We evaluated the motion replication fidelity using the MoCap system by tracking trajectories for various motion patterns, including circular, square, four-leaved rose, and spiral paths in Figure 13(b)–(e) and Supplementary Movie S3. The trajectories are plotted in the x–y plane and recorded based on the geometric center of the markers. The results show reliable motion performance with RMSE of 2.510, 2.296, 2.076, and 1.596 mm for the four trajectories. Taking the spiral trajectory as an example, Figure 13(f) shows displacements on both x and y axes, and Figure 13(g) shows snapshots every 4 s, corresponding to the purple points in Figure 13(i). The result shows accurate and robust performance for digitalizing and synthesizing ProSoRo's omnidirectional motion empowered by MVAE with latent proprioception knowledge.

6. Applicability to Sequential Contact Reasoning

6.1. Experiment Setup with a Soft-Rigid Hybrid Arm

We reused core components in the tendon-driven platform to build a soft-rigid hybrid arm within a compact design shown in **Figure 14**(a). Eight actuators (XW540-T140-R, Dynamixel) with driving drums were arranged staggered on the upper and lower layers, and the tendons went through the top surface of

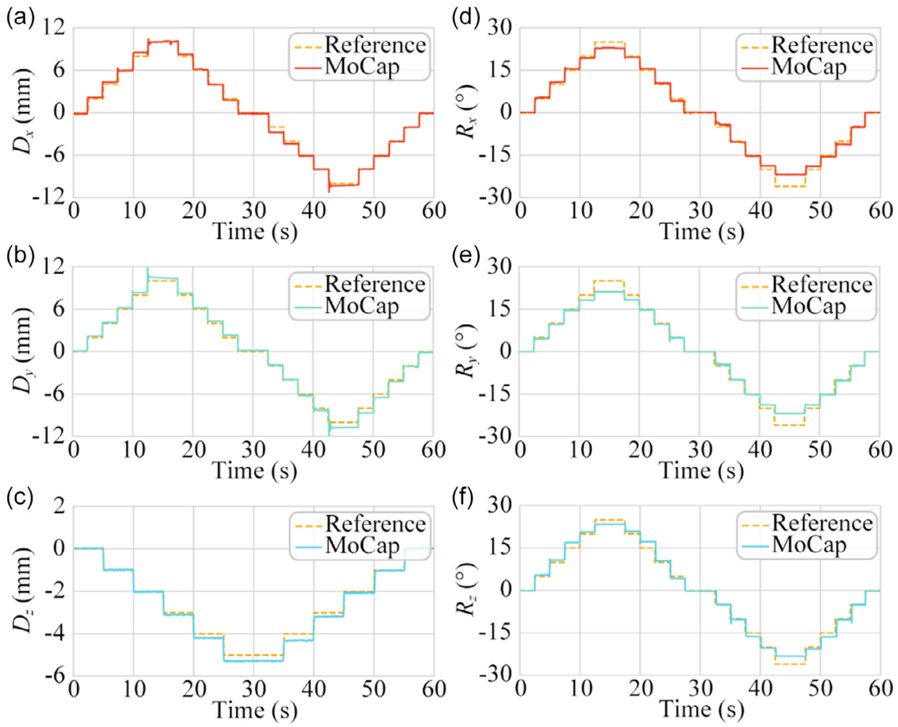

Figure 12. Performance of the tendon-driven platform with single translation a) D_x, b) D_y, c) D_z, and single rotation d) R_x, e) R_y, f) R_z on the ProSoRo, respectively, comparing the input value as a reference and the actual value measured by MoCap.

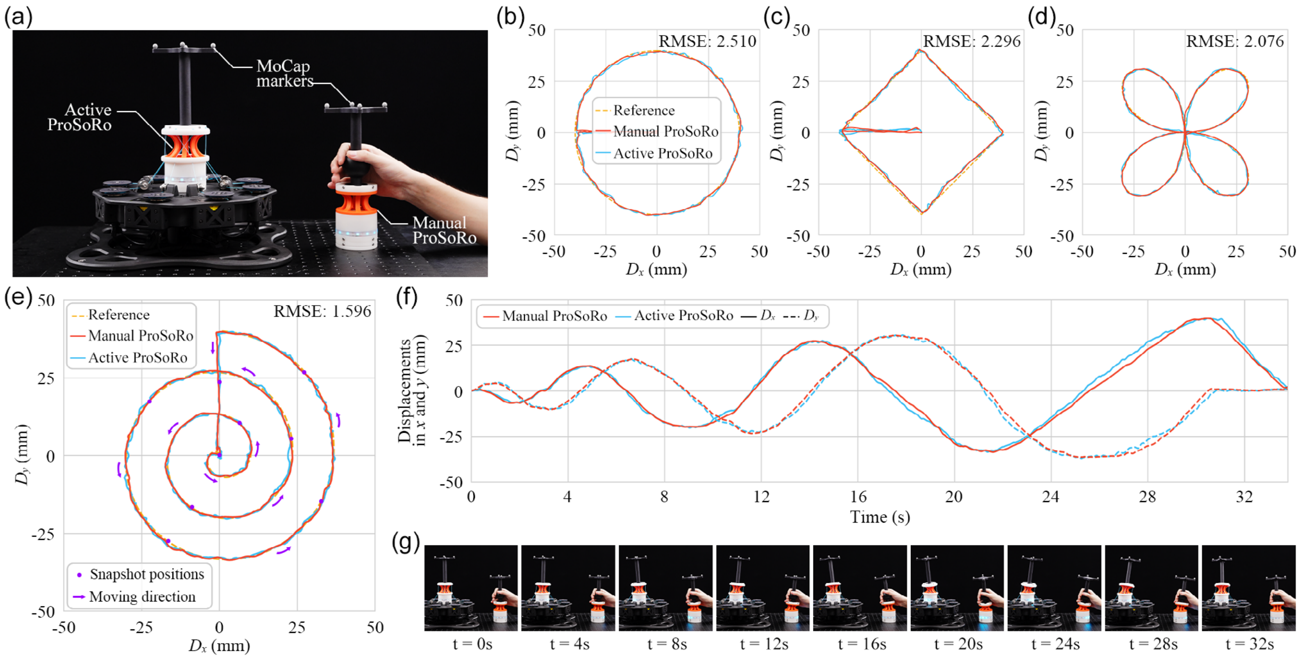

Figure 13. Evaluation of digitalizing and synthesizing motions by MVAE. a) Experiment setup, including active ProSoRo, manual ProSoRo, markers, and MoCap. b–e) Four motion trajectories, including circle, square, four-leaved rose, and spiral, shown as trajectory plots of manual and active ProSoRos in the x–y plane. f) Displacements of the manual and active ProSoRos on both x and y axes, respectively, versus time during the spiral trajectory. g) Snapshots during the spiral trajectory every four seconds, corresponding to the purple points in (e).

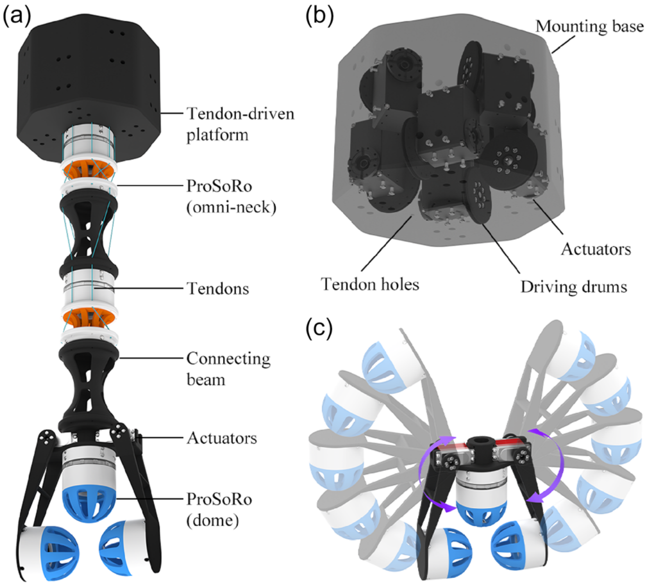

Figure 14. Tendon-driven soft-rigid hybrid arm design. a) A tendon-driven platform, omni-neck ProSoRos, connect beams, actuators, and dome ProSoRos. b) Tendon-driven platform consisting of a mounting base, actuators, and driving drums. c) Large swinging range of the gripper beams for flexible manipulation.

the mounting base, as shown in Figure 14(b). ProSoRos could be mounted on the tendon-driven platform connected by rigid hollow connecting beams. A soft-rigid hybrid gripper was mounted to the end. The gripper comprises a mounting base, two actuators (DG-3150, XUNLONGZHE), two swinging beams, and three dome ProSoRos. The three ProSoRos have adaptive interacting surfaces against the environment. The two actuators, respectively, drove the two swinging beams, and their large swinging ranges in Figure 14(c) allowed flexible manipulation according to objects and environments.

6.2. Sequential Contact Reasoning with ProSoRos

Understanding contact interactions is crucial for soft robots engaged in manipulation tasks, especially in unstructured or dynamic environments. We explored the MVAE's capability to infer contact states during a pivoting manipulation of a wine glass using a soft-rigid hybrid arm in **Figure 15**a.

The dome ProSoRos interacted with the wine glass, deforming in response to contact forces. The anchor frame motions captured by the internal cameras were input into the MVAE to estimate the deformation states. All ProSoRos' proprioceptions were estimated by MVAE and visualized on a user interface for a human operator, including the reconstructed whole-arm state in 3D, third-person point of view from the outer camera,

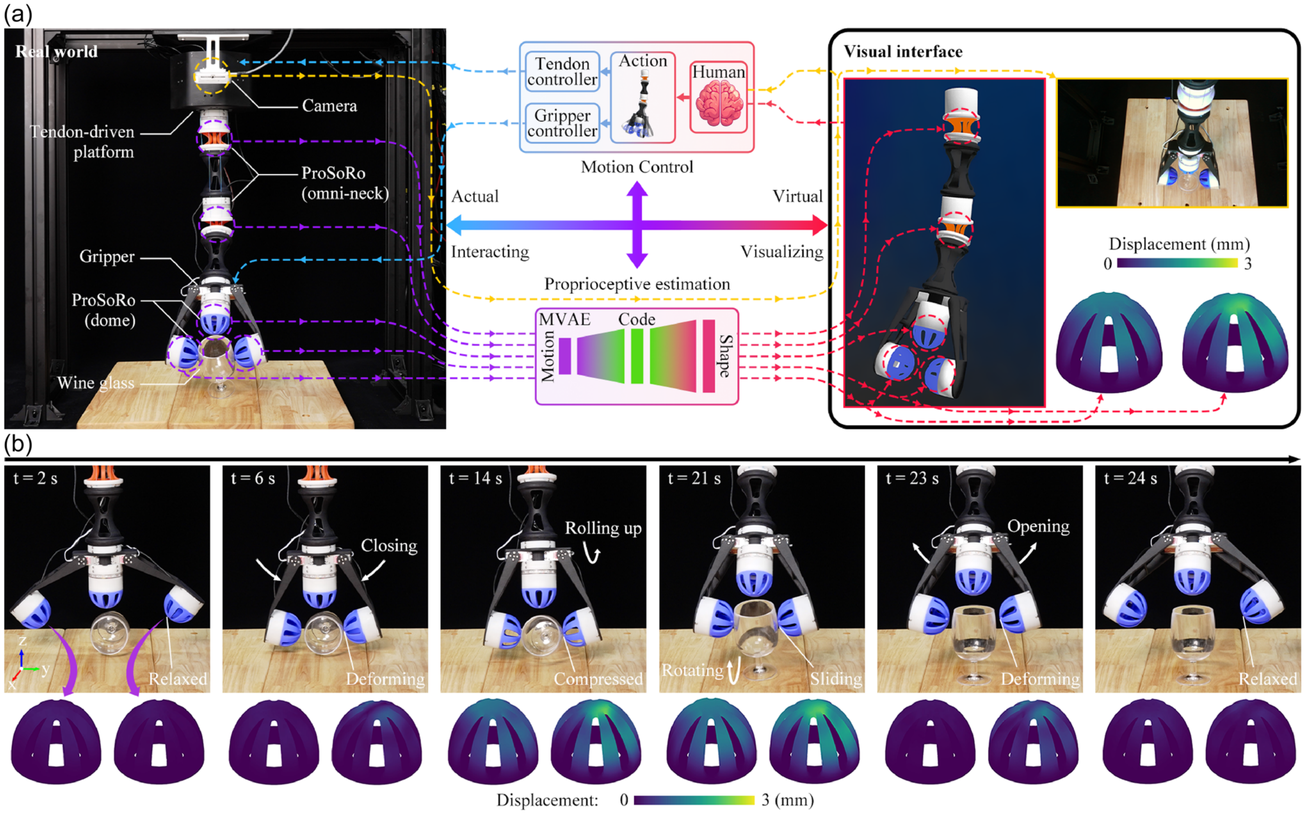

Figure 15. Setup and process for pivoting manipulation using the soft-rigid hybrid arm. a) Setup for pivoting manipulation of a wine glass, where proprioceptive states of ProSoRos are estimated by MVAE and transferred to the visual interface, assisting motion control of the soft-rigid hybrid arm. b) Snapshots of the gripper interacting with the wine glass and estimated shapes of ProSoRos.

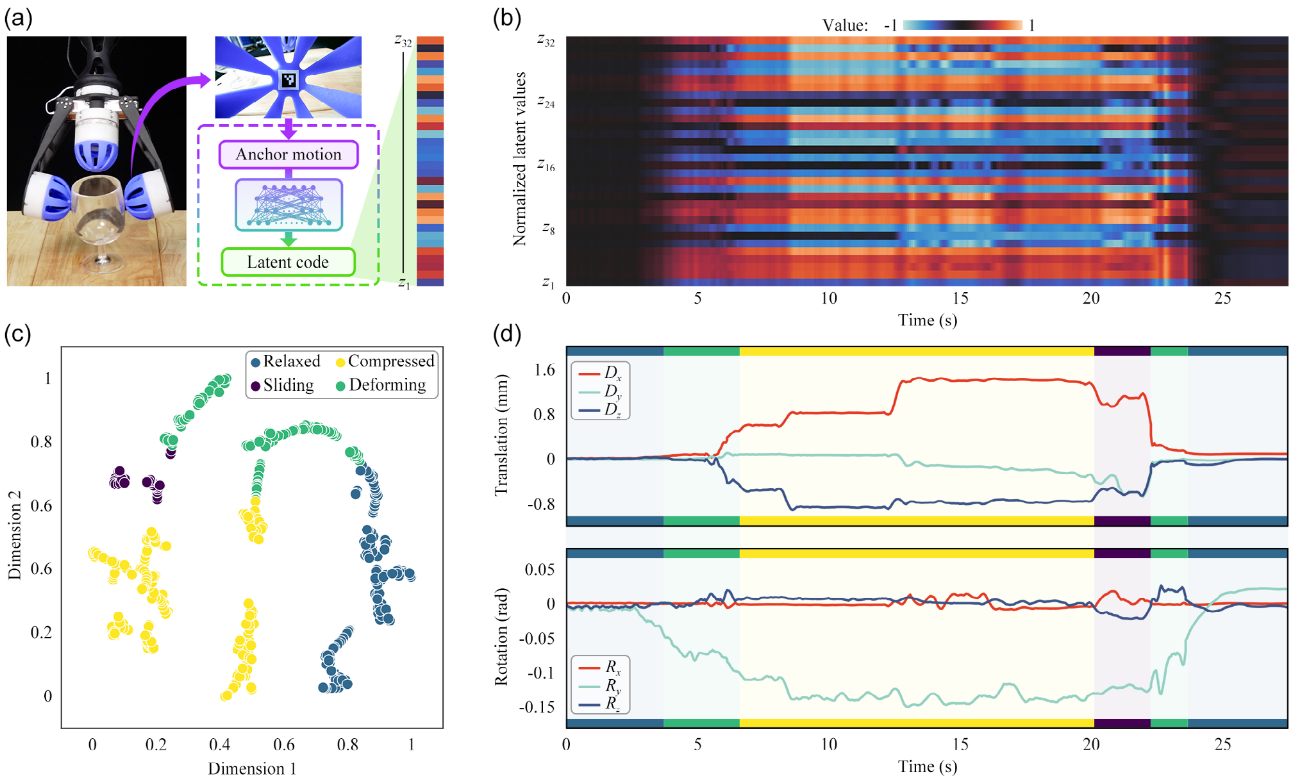

Figure 16. Sequential contact reasoning with latent proprioception. a) Latent proprioception inference from anchor motion by MVAE. b) Plot of normalized values of latent code versus time. c) Cluster distribution of latent codes, representing different contact states. d) Line plots of translation and rotation of ProSoRo's motion versus time, and the whole sequential process is separated into various contact states according to clustering.

and deformed shapes of dome ProSoRos colored by node displacements. Since the wine glass was transparent, estimating the contact states only through the camera view was difficult. Therefore, the estimated shape of the dome was crucial for maintaining an appropriate contact extent. Based on the contact information, the operator sent the following action to the arm to manipulate the wine glass from lying to standing.

In Figure 15b, the first row shows time-sequential snapshots of the gripper interacting with the wine glass, and the second row shows the deformed shapes of the dome ProSoRos estimated by MVAE. Initially, the arm was naturally downward, and the gripper was open. Then, the gripper closed to catch the lying wine glass, and the two dome ProSoRos were compressed with a specific contact extent, allowing the wine glass to rotate in the hand. The colors represented the displacements of each node, facilitating the operator's intuitive judgment. The gripper dragged the wine glass on the plane when the arm rolled up. After the wine glass was lifted to the appropriate height, it rotated to a standing state in the gripper. Then the gripper opened, and the two dome ProSoRos were relaxed. Supplementary Movie S4 shows the complete process.

In **Figure 16**(a), during the sequential contact process, the two dome ProSoRos on the gripper were deformed when contacting with the wine glass, and the motion captured by the inner camera was encoded to latent code z_1 to z_{32} by MVAE. By analyzing the temporal evolution of the latent codes, we observed distinct patterns corresponding to different contact states. Taking the right dome ProSoRo as an example, Figure 16(b) plots the normalized values of latent code versus time, showing that the behavior of each component was different and the time sequence could be separated into several distinct stages. Clustering these latent codes revealed four primary interaction phases: relaxed, compressed, sliding, and deforming in Figure 16(c). We used the *k*-means clustering algorithm to classify the latent codes into four categories. We visualized them by *t*-distributed stochastic neighbor embedding (*t*-SNE). Mapping these clusters onto the motion sequence provided insights into the manipulation process in Figure 16(d), enabling more precise control and adaptation. The dome ProSoRo's motion in terms of translation and rotation versus time is separated into six stages sequentially: relax, deforming, compressed, sliding, deforming, and back to relax. It is observed that the four categories obtained from clustering correspond to the actual interactions on the dome ProSoRo. The results show the contact reasoning capability of the latent proprioception during sequential contact manipulation, providing strong support for researchers to identify contact states and apply them to control strategies.

7. Discussion and Conclusion

In this study, we introduce ProSoRo, a proprioceptive soft robotic system that addresses one of the fundamental challenges in soft robotics: capturing and utilizing the complex, continuous deformations inherent in soft materials for precise control and adaptive interaction. By anchoring the robot's morphological state to a single internal reference frame and leveraging a multi-modal variational autoencoder (MVAE), we have demonstrated that it is possible to infer high-dimensional proprioceptive data, encompassing internal force and whole-body shape, from low-dimensional observations like the motion of the reference frame.

Minimizing the sensing requirements to a single internal frame significantly reduces the design and sensing complexity traditionally associated with soft robots. These robots often require extensive sensor arrays or external measurement systems to monitor their infinite degrees of freedom. By simplifying the sensing to a single internal frame, we preserve the robot's compliance and softness while enabling real-time, accurate estimation of its state. This method provides a practical and efficient solution for proprioception in soft robots, which is crucial for tasks requiring precise control in complex and dynamic environments.

The MVAE framework is central to our system, effectively aligning multiple proprioceptive modalities into a unified latent code termed latent proprioception. We ensured high-quality learning while minimizing the sim-to-real gap by training the MVAE on simulation data generated through finite element analysis (FEA) with optimized material parameters. The MVAE's ability to perform crossmodal inference allows for estimating difficult-to-measure modalities, such as internal forces and whole-body deformations, from easily accessible measurements like the motion of a single internal frame. This enhances the robot's adaptability and functionality, allowing for more complex tasks without additional sensing infrastructure.

One insight from our study is the identification of key morphing primitives within the MVAE's latent code. These primitives represent fundamental deformation modes of the soft robot and serve as intuitive control handles for manipulating complex deformations. We can generate a spectrum of deformation behaviors by systematically varying these latent components, offering a novel perspective on soft robotic systems' intrinsic dimensionality and controllability. Although this motion cannot be decomposed into conventional displacements or rotations along the x, y, or z axes in the Cartesian space, it can be clearly seen that the motion spaces generated by different key morphing primitives have specific manifolds. This understanding enhances the interpretability of the latent code and facilitates the development of more sophisticated control strategies and advanced human–robot interfaces.

Our experimental results underscore the effectiveness and versatility of the proposed approach. We validated the MVAE's ability to transfer learned proprioceptive representations from simulation to real-world scenarios. The estimated forces and shapes closely matched real-world measurements across different ProSoRo variants, achieving high accuracy despite the inherent challenges of soft robotic materials and complex deformations. Furthermore, the framework applies to various soft robotic designs and tasks, from motion replication to advanced manipulation, highlighting its broad utility in assistive technologies, exploration, and industrial automation. By enabling precise, synchronized actions between a manual ProSoRo and an active ProSoRo on a tendon-driven platform, we showcased the system's ability to facilitate intuitive human–robot interfaces and coordinated control. We also demonstrated ProSoRo's ability to infer contact states during a soft-rigid hybrid arm's pivoting manipulation of a wine glass. By analyzing the temporal evolution of latent codes, we identified distinct patterns corresponding to different contact states, including relaxed, compressed, sliding, and deforming, enabling more precise control and adaptation in real time.

While our approach offers significant advancements, certain limitations exist. The effectiveness of using a single internal reference frame assumes that this frame's motion can adequately capture the robot's deformations. For structures with complex, localized deformations or distributed contacts, this assumption may not hold. Future work could explore incorporating additional reference frames or distributed sensing to enhance the system's ability to capture localized behaviors. The quality of the MVAE's learning is contingent on the accuracy of the simulation data. While we optimized material parameters to reduce the sim-to-real gap, unmodeled phenomena or simplifications (e.g., viscoelasticity, hysteresis) in the FEA could lead to discrepancies. Integrating adaptive modeling techniques or online learning algorithms that update the MVAE based on real-world feedback could enhance performance. In addition, the ablation study on MVAE could help us better understand the key factors involved in latent proprioception learning. Our current framework focuses on motion, force, and shape. Expanding the MVAE to include other modalities such as temperature, distributed stress, or acoustic signals could enrich the latent proprioceptive space and enable the robot to interact with its environment more nuancedly.

Notwithstanding the above limitations, our study presents a significant step toward enabling soft robots to achieve proprioceptive awareness and control comparable to their rigid counterparts while retaining the inherent advantages of softness and compliance. By anchoring morphological representations and leveraging latent proprioception through the MVAE, we provide a practical and scalable solution to one of the core challenges in soft robotics. This work lays the foundation for future developments where soft robots can become more autonomous, intelligent, and capable of complex interactions with their environment.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was partly supported by the National Science Foundation of China under grant 62473189 and grant 62206119 and Shenzhen Long-Term Support for Higher Education at SUSTech under grant 20231115141649002.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are openly available in GitHub at https://github.com/ancorasir/ProSoRo, reference number 0.

Keywords

latent representations, multimodal learning, proprioceptions, soft robotics

Received: April 19, 2025
Revised: May 16, 2025
Published online: July 2, 2025

[1] C. Laschi, B. Mazzolai, M. Cianchetti, *Sci. Robot.* **2016**, *1*, eaah3690.
[2] Y. Roh, Y. Lee, D. Lim, D. Gong, S. Hwang, M. Kang, D. Kim, J. Cho, G. Kwon, D. Kang, S. Han, S. Ko, *Adv. Funct. Mater.* **2024**, *34*, 2306079.
[3] S. I. Rich, R. J. Wood, C. Majidi, *Nat. Electron.* **2018**, *1*, 102.
[4] M. Li, A. Pal, A. Aghakhani, A. Pena-Francesch, M. Sitti, *Nat. Rev. Mater.* **2022**, *7*, 235.
[5] O. Yasa, Y. Toshimitsu, M. Y. Michelis, L. S. Jones, M. Filippi, T. Buchner, R. K. Katzschmann, *Annu. Rev. Control, Robot. Auton. Syst.* **2023**, *6*, 1.
[6] G. Li, X. Chen, F. Zhou, Y. Liang, Y. Xiao, X. Cao, Z. Zhang, M. Zhang, B. Wu, S. Yin, Y. Xu, H. Fan, Z. Chen, W. Song, W. Yang, B. Pan, J. Hou, W. Zou, S. He, X. Yang, G. Mao, Z. Jia, H. Zhou, T. Li, S. Qu, Z. Xu, Z. Huang, Y. Luo, T. Xie, J. Gu, et al., *Nature* **2021**, *591*, 66.
[7] S. Wu, Y. Hong, Y. Zhao, J. Yin, Y. Zhu, *Sci. Adv.* **2023**, *9*, eadf8014.
[8] Z. Ren, W. Hu, X. Dong, M. Sitti, *Nat. Commun.* **2019**, *10*, 2703.
[9] W. Hu, G. Z. Lum, M. Mastrangeli, M. Sitti, *Nature* **2018**, *554*, 81.
[10] E. Roels, S. Terryn, J. Brancart, F. Sahraeeazartamar, F. Clemens, G. Van Assche, B. Vanderborght, *Mater. Today Electron.* **2022**, *1*, 100003.
[11] T. Chen, X. Yang, B. Zhang, J. Li, J. Pan, Y. Wang, *Sci. Robot.* **2024**, *9*, eadl0307.
[12] T. Jin, Z. Sun, L. Li, Q. Zhang, M. Zhu, Z. Zhang, G. Yuan, T. Chen, Y. Tian, X. Hou, C. Lee, *Nat. Commun.* **2020**, *11*, 5381.
[13] N. Guo, X. Han, S. Zhong, Z. Zhou, J. Lin, J. S. Dai, F. Wan, C. Song, *IEEE Trans. Robot.* **2024**, *40*, 4684.
[14] X. Chen, X. Zhang, Y. Huang, L. Cao, J. Liu, *J. Field Robot.* **2022**, *39*, 281.
[15] X. Liu, X. Han, W. Hong, F. Wan, C. Song, *Int. J. Robot. Res.* **2024**, *43*, 1916.
[16] Y. Dong, L. Wang, N. Xia, Z. Yang, C. Zhang, C. Pan, D. Jin, J. Zhang, C. Majidi, L. Zhang, *Sci. Adv.* **2022**, *8*, eabn8932.
[17] J. Fras, Y. Noh, M. Macias, H. Wurdemann, K. Althoefer, *2018 IEEE international conference on robotics and automation (ICRA)*, IEEE, **2018**, 1583–1588.
[18] C. Wang, V. R. Puranam, S. Misra, V. K. Venkiteswaran, *IEEE Robot. Autom. Lett.* **2022**, *7*, 5795.
[19] J. Shi, G. Shi, Y. Wu, H. A. Wurdemann, *IEEE Trans. Med. Robot. Bionics* **2024**, *6*, 1309.
[20] J. Qiu, X. Guo, R. Chu, S. Wang, W. Zeng, L. Qu, Y. Zhao, F. Yan, G. Xing, *ACS Appl. Mater. Interfaces* **2019**, *11*, 40716.
[21] Z. Zuo, X. He, H. Wang, Z. Shao, J. Liu, Q. Zhang, F. Pan, L. Wen, *IEEE Robot. Autom. Mag.* **2024**, *31*, 96.
[22] J. Qu, G. Cui, Z. Li, S. Fang, X. Zhang, A. Liu, M. Han, H. Liu, X. Wang, X. Wang, *Adv. Funct. Mater.* **2024**, *34*, 2401311.
[23] D. R. Yao, I. Kim, S. Yin, W. Gao, *Adv.Mater* **2024**, *36*, 2308829.
[24] C. Armanini, F. Boyer, A. T. Mathew, C. Duriez, F. Renda, *IEEE Trans. Robot.* **2023**, *39*, 1728.
[25] C. Zhang, P. Zhu, Y. Lin, Z. Jiao, J. Zou, *Adv. Intell. Syst.* **2020**, *2*, 1900166.
[26] T. J. Prescott, K. Vogeley, A. Wykowska, *Sci. Robot.* **2024**, *9*, eadn2733.
[27] Z. Fu, A. Kumar, A. Agarwal, H. Qi, J. Malik, D. Pathak, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, **2022**, 17273–17283.
[28] S. A. B. Birjandi, J. Kühn, S. Haddadin, *IEEE Robot. Autom. Lett.* **2020**, *5*, 954.
[29] R. L. Truby, C. Della Santina, D. Rus, *IEEE Robot. Autom. Lett.* **2020**, *5*, 3299.
[30] C. Han, Y. Jeong, J. Ahn, T. Kim, J. Choi, J.-H. Ha, H. Kim, S. H. Hwang, S. Jeon, J. Ahn, J. Hong, J. Kim, J. Jeong, I. Park, *Adv. Sci.* **2023**, *10*, 2302775.
[31] H. Wang, M. Totaro, L. Beccai, *Adv. Sci.* **2018**, *5*, 1800541.
[32] Z. Wu, Y. Zhao, Y. Duo, B. Li, L. Li, B. Chen, K. Yang, S. Su, J. Guan, L. Wen, M. Liu, *ACS Appl. Mater. Interfaces* **2024**, *16*, 64222.
[33] S. E. Navarro, S. Nagels, H. Alagi, L.-M. Faller, O. Goury, T. Morales-Bieze, H. Zangl, B. Hein, R. Ramakers, W. Deferme, G. Zheng, C. Duriez, *IEEE Robot. Autom. Lett.* **2020**, *5*, 5621.
[34] H. Bai, S. Li, J. Barreiros, Y. Tu, C. R. Pollock, R. F. Shepherd, *Science* **2020**, *370*, 848.
[35] B. G. Cangan, S. E. Navarro, B. Yang, Y. Zhang, C. Duriez, R. K. Katzschmann, *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, **2022**, 9424–9430.
[36] Q. Wang, Z. Wu, J. Huang, Z. Du, Y. Yue, D. Chen, D. Li, B. Su, *Compos. B: Eng.* **2021**, *223*, 109116.
[37] S. Shu, Z. Wang, P. Chen, J. Zhong, W. Tang, Z. L. Wang, *Adv.Mater* **2023**, *35*, 2211385.
[38] Y. Jiang, S. Yin, J. Dong, O. Kaynak, *IEEE Sens. J.* **2020**, *21*, 12868.
[39] X. An, Y. Cui, X. Dong, Y. Wang, B. Du, X.-J. Liu, H. Zhao, *Cell Rep. Phys. Sci.* **2024**, *5*, 10.
[40] B. Shih, D. Shah, J. Li, T. G. Thuruthel, Y.-L. Park, F. Iida, Z. Bao, R. Kramer-Bottiglio, M. T. Tolley, *Sci. Robot.* **2020**, *5*, eaaz9239.
[41] Q. Sun, Z. Ge, *IEEE Trans. Industr. Inform.* **2021**, *17*, 5853.
[42] S. Zhou, Y. Li, Q. Wang, Z. Lyu, *Cyborg Bionic Syst.* **2024**, *5*, 0105.
[43] D. Bruder, X. Fu, R. B. Gillespie, C. D. Remy, R. Vasudevan, *IEEE Trans. Robot.* **2020**, *37*, 948.
[44] A. Spielberg, A. Zhao, Y. Hu, T. Du, W. Matusik, D. Rus, *Adv. Neural Inf. Process Syst.* **2019**, *32*, 8284.
[45] T. Sugiyama, K. Kutsuzawa, D. Owaki, M. Hayashibe, *Soft Robot.* **2024**, *11*, 105.
[46] P. Zhou, P. Zheng, J. Qi, C. Li, H.-Y. Lee, A. Duan, L. Lu, Z. Li, L. Hu, D. Navarro-Alarcon, *Robot. Comput. Integr. Manuf.* **2024**, *88*, 102727.
[47] T. Aumentado-Armstrong, S. Tsogkas, S. Dickinson, A. Jepson, *Int. J. Comput. Vis.* **2023**, *131*, 1611.
[48] C. Jiang, J. Huang, A. Tagliasacchi, L. J. Guibas, *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 9745.
[49] M. Wu, N. Goodman, *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 5580.
[50] Y. Shi, N. Siddharth, B. Paige, P. Torr, *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 15718.
[51] T. M. Sutter, I. Daunhawer, J. E. Vogt, *arXiv preprint arXiv:2105.02470* **2021**.
[52] N. Hansen, S. D. Müller, P. Koumoutsakos, *Evol. Comput.* **2003**, *11*, 1.
[53] D. P. Kingma, *arXiv preprint arXiv:1312.6114* **2013**.
[54] J. Gao, M. Y. Michelis, A. Spielberg, R. K. Katzschmann, *IEEE Robot. Autom. Lett.* **2024**.
[55] G. Fang, Y. Tian, Z.-X. Yang, J. M. Geraedts, C. C. Wang, *IEEE/ASME Trans. Mechatron.* **2022**, *27*, 5296.
[56] W. Zhu, X. Guo, D. Owaki, K. Kutsuzawa, M. Hayashibe, *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *34*, 3444.