Vision-based Proprioceptive and Force Sensing for Soft Robotic Actuator

by

Annan Zhang

B.Sc., ETH Zürich (2019)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

Author .....
Department of Electrical Engineering and Computer Science
May 13, 2022

Certified by .....
Daniela Rus
Professor of Electrical Engineering and Computer Science
Director, Computer Science and Artificial Intelligence Laboratory
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Vision-based Proprioceptive and Force Sensing for Soft Robotic Actuator

by

Annan Zhang

Submitted to the Department of Electrical Engineering and Computer Science
on May 13, 2022, in partial fulfillment of the
requirements for the degree of
Master of Science in Electrical Engineering and Computer Science

Abstract

Developing reliable control strategies for soft robots requires advances in soft robot perception. Due to their near-infinite degrees of freedom, obtaining useful sensory feedback from soft robots remains a long-standing challenge. Moreover, sensorization methods must be co-developed with more robust approaches to soft robotic actuation. However, current soft robotic sensors pose many performance limitations, and available materials and manufacturing techniques complicate the design of sensorized soft robots. To address these needs, we introduce a vision-based method to sensorize robust, electrically-driven soft robotic actuators constructed from a new class of architected materials. Specifically, we position cameras within the hollow interiors of actuators based on handed shearing auxetics (HSA) to record their deformation. Using external motion capture data as ground truth, we train a convolutional neural network (CNN) that maps the visual feedback to the pose of the actuator's tip. Our model provides predictions of tip pose with sub-millimeter accuracy from only six minutes of training data, while remaining lightweight with 300,000 parameters and an inference time of 18 milliseconds per frame on a single-board computer. We also develop a model that additionally predicts the horizontal tip force acting on the actuator and demonstrate its ability to generalize to previously unseen forces. Overall, our methods present a reliable vision-based approach for designing sensorized soft robots built from electrically-actuated, architected materials.

Thesis Supervisor: Daniela Rus
Title: Professor of Electrical Engineering and Computer Science
Director, Computer Science and Artificial Intelligence Laboratory

Acknowledgments

First and foremost, I would like to thank my advisor, Daniela. She has been a constant source of inspiration and encouragement, always finding the right words that push me to think deeper and to *keep the gradient*. I am especially thankful for the freedom she has given me to explore my interests.

My collaborators made this project possible. Ryan's onboarding and mentorship have contributed a lot to a smooth start during the pandemic year at MIT. Lilly's sharp-mindedness and can-do attitude make her an invaluable colleague. Shuguang's lightheartedness infects every room he enters. I would also like to thank all other lab members and colleagues for their help and Mieke for her organizational genius.

Finally, I would like to thank Jess, my friends, and my parents for their unconditional love and support. I am especially grateful to have found such a great group of friends in my cohort, bonding over the common difficulties we faced because of starting graduate school in 2020.

This work was done at the Computer Science and Artificial Intelligence Laboratory at MIT and supported by the National Science Foundation, EFRI Grant #1830901.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Despite rapidly advancing technologies, robots are still not part of most people's daily lives. While robotic arms are already ubiquitous on factory floors and assembly lines, it is rare to encounter autonomous helpers in the home, supermarket, or hospital. To make progress in integrating robots into people's daily lives, interactions between robots and humans must be fundamentally safe and robust. However, robotic arms used on factory floors are often fundamentally unsafe. The rigidity of these robots, which makes them accurate and useful, also makes them dangerous. As a result, they are often used where there are no humans around, or cages are built around them. So-called soft robots are a new way of thinking about robots. They are made of soft materials with lower stiffness and promise to be fundamentally safe in human-robot interactions [43]. Soft robots are also excellent at dealing with uncertainty because the deformability of the material allows them to adapt to their environment. Applications that exploit this property include grippers [1, 29], robotic manipulators [34, 25], crawling and walking search-and-rescue robots [37, 8], exoskeletons [40], and auxiliary tools for surgery [41, 42].

However, perception has remained a critical challenge in the area of soft robotics. Accurate and robust sensory feedback is fundamental to the development of intelligent soft robots [64]. For example, perception is required to creating robust control strategies for soft robots, especially since soft robots' compliance gives them near-infinite potential degrees of freedom [2]. However, despite recent advances, soft robotic sens-

ing is still considered to be in its infancy [64]. One of the key limitations of sensorizing soft robots is the need to integrate sensing and actuation capabilities together. For example, in popular fluidic elastomer actuators, the soft robot moves by the pressurization of channels in elastomers [36]. Although traditional external motion capture or vision-based systems can be used as a proxy for direct perception [6, 19], true soft perception would require the sensors to be incorporated within and move along with a pneumatic soft robot's body. However, the very actuation scheme that enables the robots to move with compliance makes it difficult to sensorize. While significant progress has been made by embedding sensors within soft robots or encasing soft robots with a sensorized skin [64, 49, 60], these methods still require either significant changes to the fabrication process or have sensors with time-varying, hysteretic responses. These limitations make it difficult to design and control soft robots for dynamic applications, highlighting a clear need for perception strategies that are tightly intertwined with the robot's mode of actuation.

We consider a new approach for sensorizing soft robotic actuators based on handed shearing auxetics (HSAs) using visual feedback from onboard cameras. HSA-based actuators are motorized systems that rely on two counter-rotating cylinders to create compliant motion [3]. Like other soft robotic systems, the complex fabrication and geometry of HSA-based systems make them difficult to sensorize directly. Similarly, cameras have recently been used in a variety of soft robots for endogenous vision-based perception to bypass the challenges of conventional sensing approaches [68, 48, 65, 66]. However, these sensing solutions are not easily incorporated into a moving system given their rigidity.

In this work, we combine the HSA architecture with onboard cameras in a complementary way to achieve accurate and robust proprioception and force sensing (Fig. 1-1). Cameras for vision-based sensing are positioned at the non-rotating, distal end of the HSA actuator, giving a static base for stable visual acquisition while simultaneously providing a tight coupling with the system's actuation with no risk of mechanical interference. Using motion capture (mocap) data of the HSA actuator's tip pose as ground truth, we develop and train a convolutional neural network (CNN)

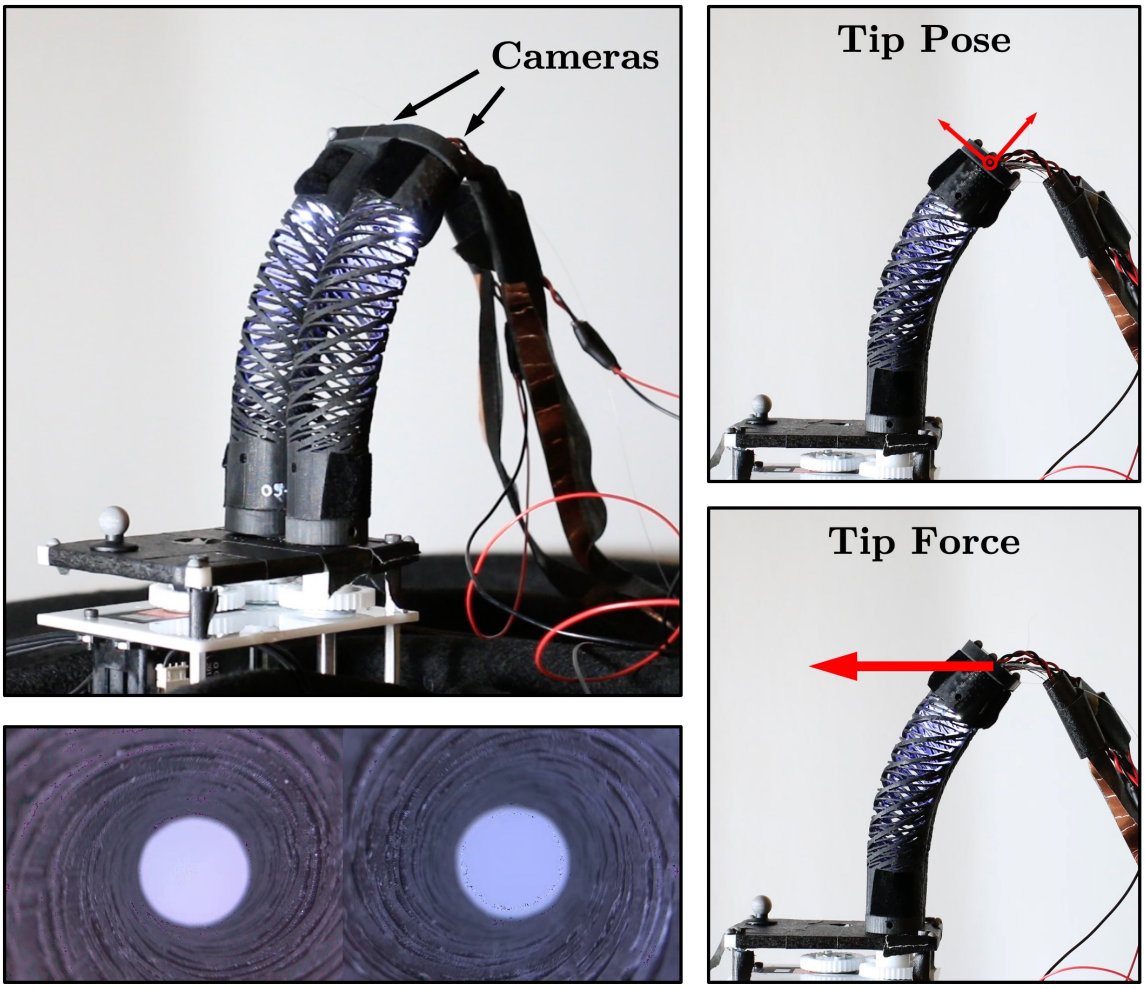

Figure 1-1: **Overview of the sensorization approach.** Proprioception through vision in handed shearing auxetic (HSA) actuators. Cameras are placed at the distal end of the HSA-based actuator (top left) that record the interior of the HSAs (bottom left). A trained CNN takes these camera images as input to predict tip pose (top right) and horizontal tip force (bottom right).

to proprioceptively predict the actuator's tip pose during free, unloaded bending. We then load the HSA actuator with weights and use the same model to predict the contact force acting on the actuator during stationary motifs. The high resolution of the onboard cameras allows us to fully capture the HSAs' complex range of motion, providing accurate and repeatable predictions for tip pose and contact force on previously unseen test data. In the case of free bending, our model predicts the tip position with a mean absolute error (MAE) of 0.10 mm, matching the accuracy at which the mocap system records the real-world data. When predicting contact force and tip pose simultaneously, our model achieves an MAE of 0.04 N on the force and an MAE of 0.27 mm on the positions.

1.1 Contributions

While similar methods have employed deep learning for proprioceptive and tactile feedback in other soft sensorized robots [59, 54, 31, 9, 50], our methods provide more accurate predictions and a more lightweight approach. First, utilizing cameras for sensing enables a lightweight model compared to the recurrent neural networks (RNNs) required for soft sensorized robots with hysteretic, piezoresistive sensors [54, 59]. Our final model for tip pose prediction deployed on a single-board computer takes on average 18 milliseconds to process a frame read in by the cameras. Furthermore, our CNNs heavily reduce the size of the data sets required for training, which consequently reduces the overall training time required for this sensing approach. For example, we achieve our predictions of tip pose and tip force with only a few minutes ($\approx$ 6 min) of video data. Most importantly, the motorized nature of our HSAs yields a more robust strategy for developing soft robots appropriate for long-term deployment compared to fluidically actuated varieties [3]. Finally, HSAs are difficult to sensorize by other methods, and we use vision to integrate perceptive capabilities with a new class of soft robotic actuators. Our methods can be readily applied to other hollow-bodied or sparse structured systems and add to a growing body of work highlighting the unique benefits of vision-based sensing in

(soft) robots [68, 48, 65, 66]. Overall, we contribute:

- A new HSA-based soft robot device with integrated cameras and internal LED arrays for controlled lighting conditions,
- A data-driven deep learning approach for proprioceptive sensing in HSA-based soft robots, and
- A demonstration of simultaneous proprioceptive and contact force sensing in HSA-based soft robots.

1.2 Thesis Outline

This thesis is in large parts based on a manuscript co-written by the thesis author that is being reviewed for publication at the time of writing: "Perception through Vision for Soft Robotic Fingers Based on Handed Shearing Auxetics" [69].

The remainder of this thesis is organized as follows: Chapter 2 provides background on soft robots (Section 2.1) and handed shearing auxetics-based actuators (Section 2.2), and reviews existing approaches to perception for soft robots (Section 2.3). Chapter 3 proposes the sensing approach, introducing hardware design (Section 3.1), sensor and actuator components (Section 3.2), data acquisition methods (Section 3.3), and neural network architecture (Section 3.4). To demonstrate its viability for soft sensing, the sensing pipeline is put to the test in Chapter 4. Experimental setups are proposed (Section 4.1) and the results for tip pose prediction (Section 4.2) and horizontal tip force prediction (Section 4.3) are evaluated. Finally, Chapter 5 concludes the thesis and discusses the limitations of the approach, future work, and lessons learned by the author.

Chapter 2

Related Work

2.1 Soft Robots

Soft robotics is an emerging field that deals with robots made out of mechanically compliant materials which deform significantly in comparison to their counterparts made out of joints and rigid links [43, 57]. Inspired by animals that evolved to have soft bodies with elastic elements like tendons and muscles, researchers are trying to mimic their dexterity and dynamic abilities required for robust interaction with the world [5, 21]. Soft robots designed to be compliant and inherently safe represent a paradigm shift from conventional rigid robots that are primarily designed to be accurate and stiff under loads [5].

Soft robots find applications in a variety of real-world tasks thanks to their compliance and flexibility. They excel in applications where they can squeeze, stretch, grow, and morph – abilities conventional rigid robots typically do not possess [26]. Through their mechanical compliance, soft robots promise robustness and safer human-robot interactions [39, 7, 25], for example when working alongside humans [23]. Soft robots also find applications as medical devices, for example serving as exoskeletons for rehabilitation [40], or assisting in surgery and medical procedures [41, 42]. Furthermore, soft robots demonstrate extraordinarily robust performance on complicated tasks involving uncertainty, like grasping [52, 1, 17, 44, 4, 29] or manipulation [34, 25, 13]. For example, soft vacuum-based grippers can snuggle up against the surfaces of the

grasped objects and grasp a wide range of objects of various unknown shapes and sizes without the need to specify exact motor commands or grasp forces [29]. On the contrary, when using a rigid gripper to grasp a fragile item, the slightest inaccuracies in sensor signals or actuation commands may lead to a breaking item. While rigid robots can be damaged themselves due to their lack of compliance, soft robots are shown to withstand extreme loading conditions, such as heavy blows of a hammer [47, 56, 13]. Soft robots also have the ability to adapt to harsh environments, which makes them viable for search-and-rescue operations. For example, snake-like soft robots can crawl and navigate through various terrains [8, 14, 62], and humanoids can explore areas too dangerous for humans in a post-earthquake scenario [37].

However, the versatility of soft robots comes at a cost. Their low-stiffness materials deform continuously, often described as having practically infinite degrees of freedom [43]. Soft robots are thus heavily underactuated. Furthermore, their materials undergo large deformations and show nonlinear material behavior on the constitutive level like hyperelasticity. Time-varying effects like long-term drift and viscoelasticity result in rate-dependent deformation behavior, stress relaxation, creep, and hysteresis, which make soft robots hard to model.

2.2 Handed Shearing Auxetics

Handed shearing auxetics (HSAs), first introduced in [30], are a class of metamaterials that couple shearing with extension through form. More specifically, the internal cellular pattern of an HSA asymmetrically enables a chiral coupling between extension and rotation. When tessellated on a cylindrical surface, the HSA pattern gives rise to a compliant linear actuator that extends in the axial direction when twisted at its end. Joining four of these linear actuators rigidly at their tops in a 2×2 arrangement and driving them via servo motors at their bottoms yields an electrically actuated robotic platform with four degrees of freedom that is capable of bending in two directions, twisting, and elongating [30]. Adding additional material along a diagonal on the wall of a cylindrical HSA constrains the HSA on one side and causes it to bend instead

of axially elongate. To build a soft robotic finger, two constraint HSA cylinders with left- and right-handed structures are rigidly joined together at their tops and driven by counter-rotating servo motors at their bottoms. Thus, pairing HSAs of opposite handedness gives rise to compliant, soft robotic actuators with actuation behaviors that bypass the need for cumbersome auxiliary hardware components like pistons and compressors required for other soft robotic actuators [3, 58]. These HSA-based grippers were shown to outperform their fluidically-driven counterparts in terms of speed, energy efficiency, strength, compactness, and puncture resistance [3].

Up until recently, fabricating HSA-based actuators involved cutting out the pattern from off-the-shelf polymer tubes on a rotary laser machine. Reliance on customary products, combined with the mechanical requirement to have a material with high elongation at break, restricted the material choice practically to extruded Teflon (polytetrafluoroethylene, PTFE) tubes. Inconsistencies in quality from off-the-shelf material and inaccuracies in the laser cutting equipment limited design flexibility. As shown in recent work, 3D printing HSAs via digital projection lithography enables more flexible design, a greater range of materials, and generally wider adoption of HSAs [58]. State-of-the-art HSA-actuators are printed out of Flexible Polyurethane 50 (FPU 50), a proprietary polymer resin mixture by Carbon Inc. This leap in manufacturing has led to a series of works that study HSAs in greater detail. First attempts were made to model the kinematics of HSAs using a piecewise constant curvature assumption [10]. More variations of the auxetic pattern were designed and shown to increase the range of forces and stiffness by orders of magnitudes [12].

2.3 Soft Robotic Perception

Prior work in soft robotic perception generally falls into two categories: sensorization via devices or sensorization via soft materials [64, 2]. For device-based sensorization, there are exogenous sensing methods with cameras and motion capture [6, 19], as well as endogenous sensing approaches, where off-the-shelf rigid sensors are strategically integrated within soft robotic bodies. Endogenous rigid devices include bendable

flex sensors [15, 38, 11], optical fibers [45, 28], Hall effect sensors [32], slide potentiometers [28], microphones [71, 53], and distributed IMUs [67, 46, 16]. Popular materials-based soft robotic sensing strategies commonly use sensors fabricated from piezoresistive composites like conductive nanoparticle-filled elastomers [22, 49, 54, 59], hydrogels [24], ionogels [61, 60], conductive liquids [35, 63], and elastomeric waveguides [55, 70, 18]. These sensing strategies have also been incorporated into deep learning pipelines for proprioceptive and tactile sensing in soft robots [59, 54, 31, 9, 50].

Each of the aforementioned sensing strategies has disadvantages. Exogenous sensing methods are not truly integrated into the robot's body, complicating its proprioceptive and tactile sensing. Endogenous sensing with rigid devices presents mechanical interfacing challenges, have limited design flexibility, and do not always provide intuitive feedback. Conversely, soft matter sensors are prone to time-varying, hysteretic feedback, environmental sensitivity, drift, and long-term issues with reliability. Finally, all of these methods present issues related to the intrinsic limitations of available methods for designing and fabricating soft sensorized robots with high spatiotemporal resolution sensing [64, 2].

Chapter 3

Sensing Pipeline

3.1 Design Overview

In this work, we sensorize HSA-based soft actuators by integrating a camera in each HSA cylinder (Fig. 1-1). The patterns of light detected by these cameras depend on the bending of the cylinder and are streamed into a neural network trained to provide accurate proprioceptive feedback.

As shown in Fig. 3-1, we use a bending HSA-based actuator design previously used in HSA-based manipulators [3, 58]. A pair of HSAs with opposite handedness are fabricated with fine constraints added along the HSA pattern diagonal that limit linear expansion to drive directional bending. At the distal end of the actuator, the HSAs are rigidly joined together with a top mount that provides reaction torques to each HSA when rotated at the base. At the base or proximal end of the actuator, the HSAs are connected to servo motors using rotating sockets, completing the setup of a so-called HSA *finger*. Servo motors are connected to a GPU-accelerated single-board computer (NVIDIA Jetson Nano) via USB and controlled with a maximum velocity in position control mode.

The HSAs' hollow cylindrical form and the actuator's structure enable sensorization with cameras (Fig. 3-1). We position fixed-focus cameras at the rigid, non-rotating adapter at the distal end of the actuator. The cameras record the interior of each HSA as shown in Fig. 3-2. However, HSAs are sparse structures; images

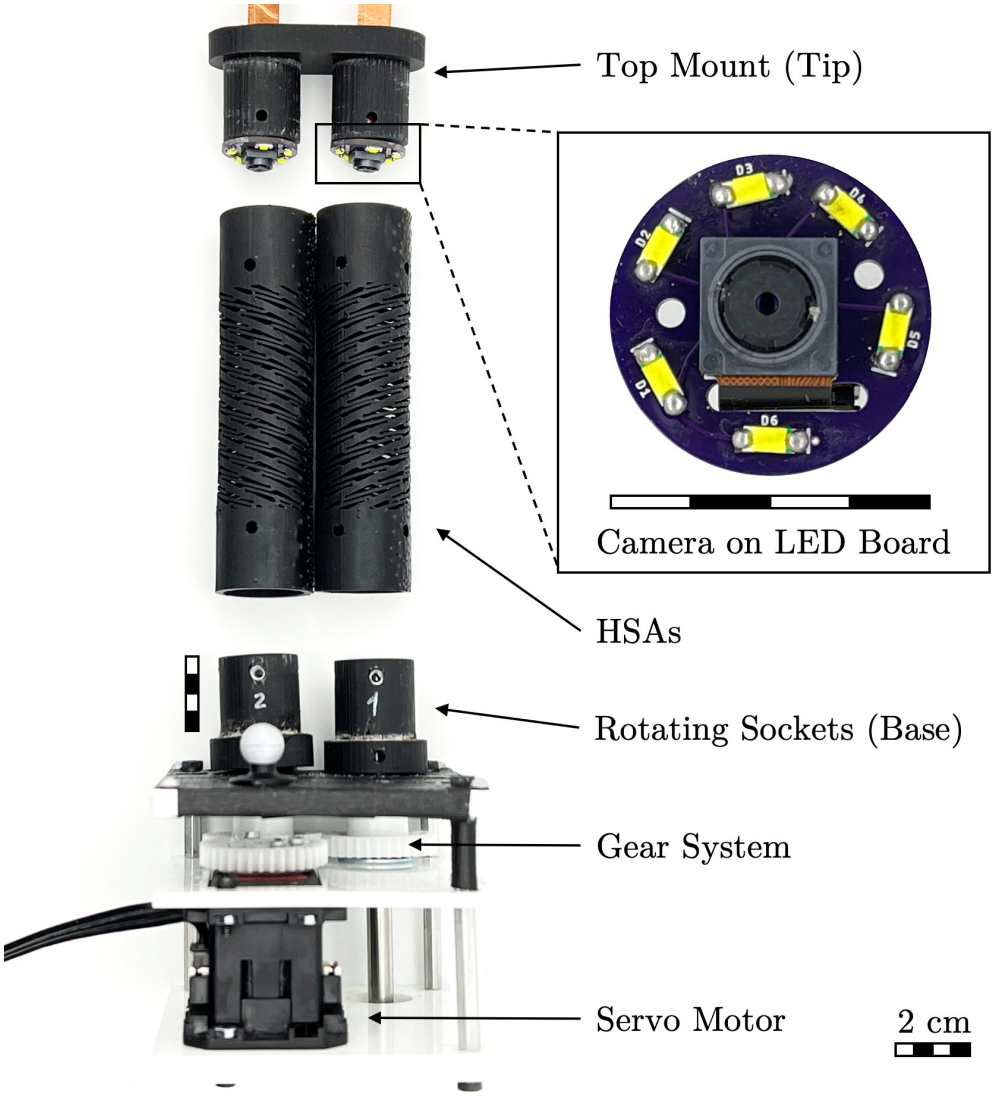

Figure 3-1: **Exploded view of hardware design.** The HSA-based soft robotic actuator is equipped with integrated cameras for perception through vision. The inset shows one HSA's camera surrounded by a ring of LEDs on our custom PCB.

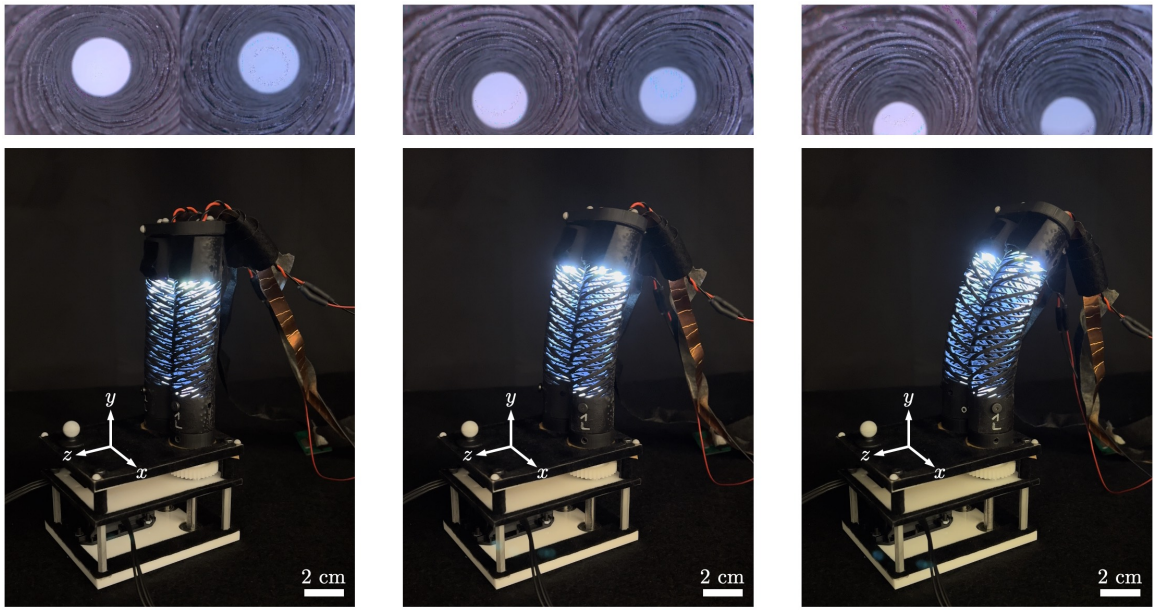

Figure 3-2: **Internal and external view during bending.** Camera images of the HSA-based actuator's interior (top row) are provided for corresponding bending states (bottom row). The columns of images from left to right show the actuator at rest, at an intermediate bending state, and in a fully bent state.

recorded by the cameras will be affected by changes in ambient light. Vision-based sensing modalities are sensitive to ambient lighting conditions, especially in low-light conditions [65]. To make our cameras robust to changes in lighting conditions in the outside environment, we integrated an array of white surface-mount LEDs around the camera. Finally, other works have shown that visually distinct markers help camera-based sensors in pose estimation tasks [66, 48]. We glue a white piece of cardboard onto the rotating sockets that has high contrast compared to the inner wall of the HSAs. This marker facilitates the cameras' visual tracking.

3.2 Actuators and Sensors

HSAs are 3D printed via digital projection lithography (Carbon M1 printer, Carbon Inc) as reported in [58] using a proprietary photopolymer resin (FPU 50, Carbon Inc). HSAs are printed with a length of 101.6 mm, an outer diameter of 25.4 mm, and a wall thickness of 2.48 mm.

Two 8-MP, fixed-focus Raspberry Pi Cameras (ArduCam IMX219) are used in the actuator. One camera is positioned into each HSA via a custom PCB (Fig. 3-1, Fig. A-2) that is fixed to the 3D printed top mount (printed from UMA 90 resin, Carbon Inc). Six white surface-mount LEDs are soldered onto each PCB. The LEDs are powered by an external circuit, connected through a JST connector that is soldered onto the bottom of the PCB. The top mount, modified from [3], enables power cables for the LEDs and flexible flat cables (FFCs) for the cameras to be fed through the non-rotating end of the actuator (Fig. A-1). The cameras are connected to the Jetson Nano with 15-pin FFCs, and the camera feed is read in with a GStreamer pipeline at 30 frames per second and a resolution of 480×360 pixels each.

3.3 Data Collection

Our data for the machine learning model consists of training and test sets, each containing input-target pairs made out of an image of the HSA interior and a 7D vector $(x, y, z, q_x, q_y, q_z, q_w)$ representing the corresponding tip pose. The tip position is given by $p := (x, y, z)$ and the quaternion components are given by $q := (q_x, q_y, q_z, q_w)$, both expressed in the base frame. Even though our pipeline takes a single image as input to output a single pose, the data is recorded in takes for practical reasons, such as hardware limitations of the Jetson Nano (e.g., RAM, hard disk size, and disk writing speed) and easier synchronization between the camera and mocap data. Furthermore, by recording whole takes instead of single images, we aim to include different actuation motifs that capture non-linear, viscoelastic material behaviors that our HSAs inherently exhibit.

The ground truth pose data is obtained with our mocap system using Motive motion capture software (OptiTrack) and six OptiTrack Flex 13 cameras that record at a frequency of 60 Hz with an accuracy of 0.2 to 0.3 mm. Our finger stands upright in the mocap workspace. Sets of infrared reflective markers are attached at the HSA top mount and base platform to define the tip and base frames, respectively. The raw mocap recordings first undergo a rigid-body transform to the reference frame before

we apply a Gaussian filter with kernel size 3 to reduce noise and down-sample them by a factor of 2. During each take, we read in each camera feed on the Jetson Nano and stack them horizontally to create a video of 960×360 pixels.

To synchronize the video recorded on the Jetson Nano with the mocap data recorded on an external PC, we make use of a bright floodlight that is detectable by the mocap system and outshines the LEDs inside the HSAs. Turning it on and off at the beginning of each recording causes rapid intensity changes in the video that can be synchronized with the appearance and disappearance of the floodlight in the mocap data. All takes for the training set are then appended together and shuffled before saving to the disk as single input-target pairs, emphasizing that our learning pipeline learns no temporal pattern. For the test take, we retain the chronological order for visualization purposes.

For force predictions, we additionally append the force component F_z to the pose, resulting in an 8D target vector, $(x, y, z, q_x, q_y, q_z, q_w, F_z)$. The ground truth force label is obtained by approximating the loading history within one take as piecewise constant (see Fig. 4-4). We obtain the force magnitude in Newtons by multiplying the weights (in grams) by $\frac{9.81}{1000}$ and the time of loading and unloading by observing the moment in the video and mocap data when the weights are put into and taken out of the bucket. We then smoothen the transitions with a Gaussian filter with kernel size 3.

3.4 Learning Architecture

As the actuator bends, the inside walls of the HSAs and the white base marking move partly or fully out of sight. Consequently, the detailed structure of the inner HSA walls is only fully observed when the actuator is in a bent state (see Fig. 3-2). This property makes traditional methods from computer vision, like detecting and tracking markers, non-straightforward to apply. Furthermore, learning-based models have proven themselves useful in vision-based sensing for soft robots [66, 48, 65]. Therefore, we use an end-to-end CNN for this work.

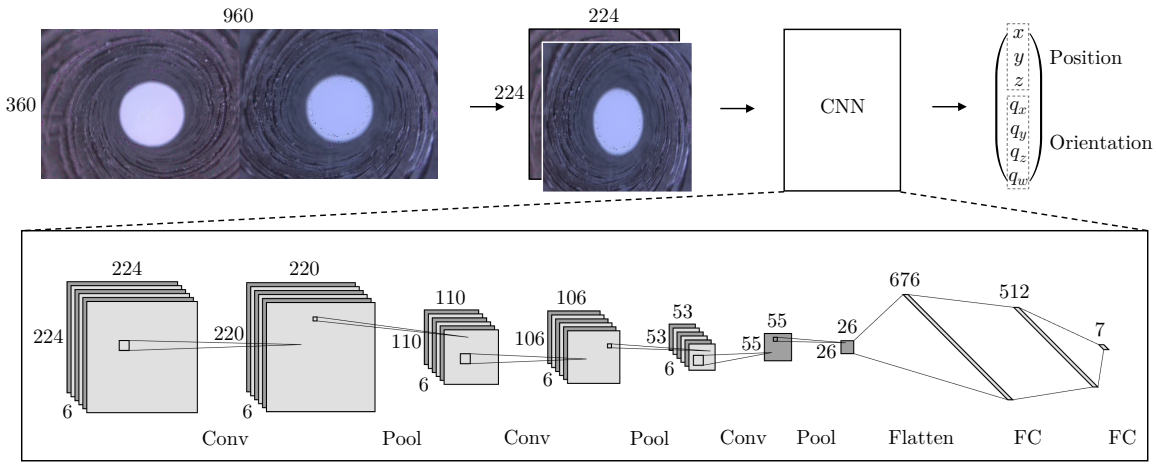

Figure 3-3: **Machine learning pipeline for tip pose estimation. Top**: Data pipeline for our sensor involves 1) reading in two 480×360 px images from the camera feed, 2) stacking them in their channel dimension and resizing them to 224×224 px, 3) normalizing each channel to zero mean and unit variance (not shown), 4) feeding them through a CNN to 5) output position (x, y, z) and orientation (q_x, q_y, q_z, q_w) of the tip. **Bottom**: Architecture of best network. Created using [27]. Numbers indicate size of feature map. All convolutional layers (Conv) use stride 1, and no padding. All pooling layers (Pool) apply max pooling with kernel size 2, stride 2, and no padding. A ReLU activation function is applied before each pooling operation (not shown). After three successive Conv-ReLU-Pool operations, the feature maps are flattened (Flatten) before sent through the fully connected layers (FC). Finally, 4 of out 7 outputs are normalized so the predicted quaternions have unit norm (not shown).

Our machine learning pipeline is depicted in Fig. 3-3 and fully implemented in PyTorch. The data loader reads a batch of 32 RGB images of size 960×360 pixels from the disk. Each image is first normalized to zero mean and unit variance in each color channel dimension. Then, the image is separated into the two images originally from each camera, stacked in their color channel dimension, and resized to 224×224 pixels. Finally, these arrays of size 224×224×6 are fed into the CNN that outputs the 7-dimensional vector defining the predicted tip pose. The ground truth tip pose is normalized to zero mean in the dimensions corresponding to the position. We choose to retain the different scales of the position dimensions to preserve the relative importance of fitting the z-coordinate, the main bending direction, in comparison to the others.

As illustrated in Fig. 3-3, we employ a conventional CNN consisting of repeated convolutional layers (Conv) followed by ReLU activation and max-pooling layers (Pool). We then flatten the feature map and add a series of fully-connected layers (FC). Finally, we apply a normalization layer to the four dimensions of the output corresponding to the orientation prediction, since orientations in 3D are described by unit quaternions and their coefficients are 4-dimensional real vectors with norm = 1.

The distance between two quaternions is given by the geodesic distance on the 4D unit sphere [33]. The Euclidean distance, which we use for position predictions, does not take the curvature of the hypersphere into account. Furthermore, quaternions with opposite signs describe the same 3D rotation, meaning one should restrict the quaternions to one-half of the sphere. Instead of the preprocessing stage, one could also alleviate this issue in the training stage by replacing the loss $\mathcal{L}(\hat{q}, q)$ between predicted $\hat{q}$ and real quaternion q with a (n.b., differentiable) loss $\min\{\mathcal{L}(\hat{q}, q), \mathcal{L}(\hat{q}, -q)\}$ that accounts for the symmetry. However, we observe in practice that for our problem, training converges even if we use the simple mean squared error (MSE) loss $||q - \hat{q}||_2^2$, which is based on the Euclidean distance. When quaternions are close, the geodesic distance can be approximated with the Euclidean distance [20]. If training is not successful, this simplification needs to be revisited. We introduce a *quaternion weight* as additional hyperparameter to balance the relative importance of fitting the position and the quaternions.

We tuned several hyperparameters: number of layers and their sizes, initial learning rate, learning rate scheduler, and quaternion weight. To select the best model without overfitting, at the beginning of each training trial we split off 10% of the training set randomly as validation set. We train each model for 100 epochs with the Adam optimizer and retain the model with the lowest validation loss.

For force predictions, the output of the last fully-connected layer has an additional dimension to predict F_z. Similar to the quaternion weight, an additional hyperparameter *force weight* balances the importance of fitting the force relative to the pose.

Chapter 4

Results

4.1 Experimental Setup

4.1.1 Tip Pose

In our first experiment, the goal is to create a vision-based sensing pipeline that outputs the current 3D pose (position and orientation) of the tip in free-bending based solely on the current image captured by the cameras. To characterize the tip pose, we first define a reference coordinate system that is fixed onto the base platform (*base* or *reference frame*), as indicated in Fig. 3-2. We then define a *tip frame* that has its origin on the top mount and its axes parallel to the base frame when the finger is in its at-rest, non-actuated configuration. The tip position is then given by the vector that connects the origin of the base frame with the origin of the tip frame. The tip orientation is given by the relative rotation of the tip frame with respect to the base frame. We choose quaternions to describe tip orientation because they offer a convenient way to represent rotations in 3D without suffering from singularities like the so-called *gimbal lock* when using Euler angles.

For the training set, we record 11 experiments in which we continuously actuate the finger back and forth between fixed configurations to sample uniformly across the range of motion. Defining the rest configuration as 0% and the fully extended configuration as 100%, our training takes consist of 11 roughly equally long recordings ac-

Type	No.	Start [%]	End [%]	v [*]	Motif
Train	1	0	67	20	move
	2	0	67	10	move
	3	0	67	40	move
	4	0	100	20	move
	5	0	100	20	move
	6	0	100	10	move
	7	0	100	40	move
	8	33	100	20	move
	9	33	100	10	move
	10	33	100	40	move
	11	33	67	20	move
Test	1	10	90	20	move & hold

Table 4.1: **Recordings for tip pose experiment.** Recordings are split up into training set and test set (Type). Start and End denote configurations of the actuator, given as a percentage of the whole range of motion from the rest configuration (0%) to the fully extended configuration (100%). The servo velocity v is given in units of 0.229 rpm. In the recordings for the training set, the actuator is moved back and forth between Start and End without interruptions (move). For the test recording, the actuator is moved between Start and End, but held in the respective extreme position for a few seconds to imitate the actuation motif of a real grasp (move & hold).

tuated between *start* and *end* configuration (in %) at servo velocity v (in 0.229 rpm), denoted as $[start, end]_v$: $[0, 67]_{20}$, $[0, 67]_{10}$, $[0, 67]_{40}$, $[0, 100]_{20}$, $[0, 100]_{20}$, $[0, 100]_{10}$, $[0, 100]_{40}$, $[33, 100]_{20}$, $[33, 100]_{10}$, $[33, 100]_{40}$, $[33, 67]_{20}$. To evaluate our trained model in a realistic scenario, we obtain a test set in which we imitate the actuation motifs used during a real grasp from a soft robotic gripper comprised of HSA fingers (as in [3]). For this, we record one take as $[10, 90]_{20}$, but hold the finger at the 10% and 90% configurations for a few seconds (see Fig. 4-2), corresponding to the robot moving around and holding the grasped item, respectively. An overview of the collected data set is given in Tab. 4.1. The training and test sets consist of 11,531 and 1,318 data points, respectively.

4.1.2 Contact Force

In our second experiment, we load the finger and modify our pipeline to predict the contact force on the tip in addition to its pose. Due to limitations of our current setup inside our mocap workspace, we only consider forces that are normal to the bending direction, i.e. parallel to the z-axis of the reference frame. Measuring the tip pose relative to the base frame makes our data independent of the absolute position of the mocap setup. This allows us to rotate the whole platform so we emulate forces in the z-direction by loading under gravity (see Fig. 4-1) without needing to make changes to the preprocessing pipeline for the pose data. We attach a small cardboard bucket to the tip into which we can put different weights, ranging from 10 to 100 grams. Since predicting contact forces and pose simultaneously is a harder problem than only predicting pose, we add vertical black lines to the originally white base markings (Fig. 4-1), providing a more informative visual feature to our sensor.

The training data for this experiment consists of 10 takes during which we load the finger with a constant weight (0 g, 50 g, 100 g, 150 g) and actuate the finger as $[0, 100]_{20}$. For each weight, we actuate the finger once without interruption (similar to training data for pose prediction) and once while holding at 0% and 100% for a few seconds (similar to test data for pose prediction). This yields eight takes. Furthermore, we record two more training takes with no load where we hold the

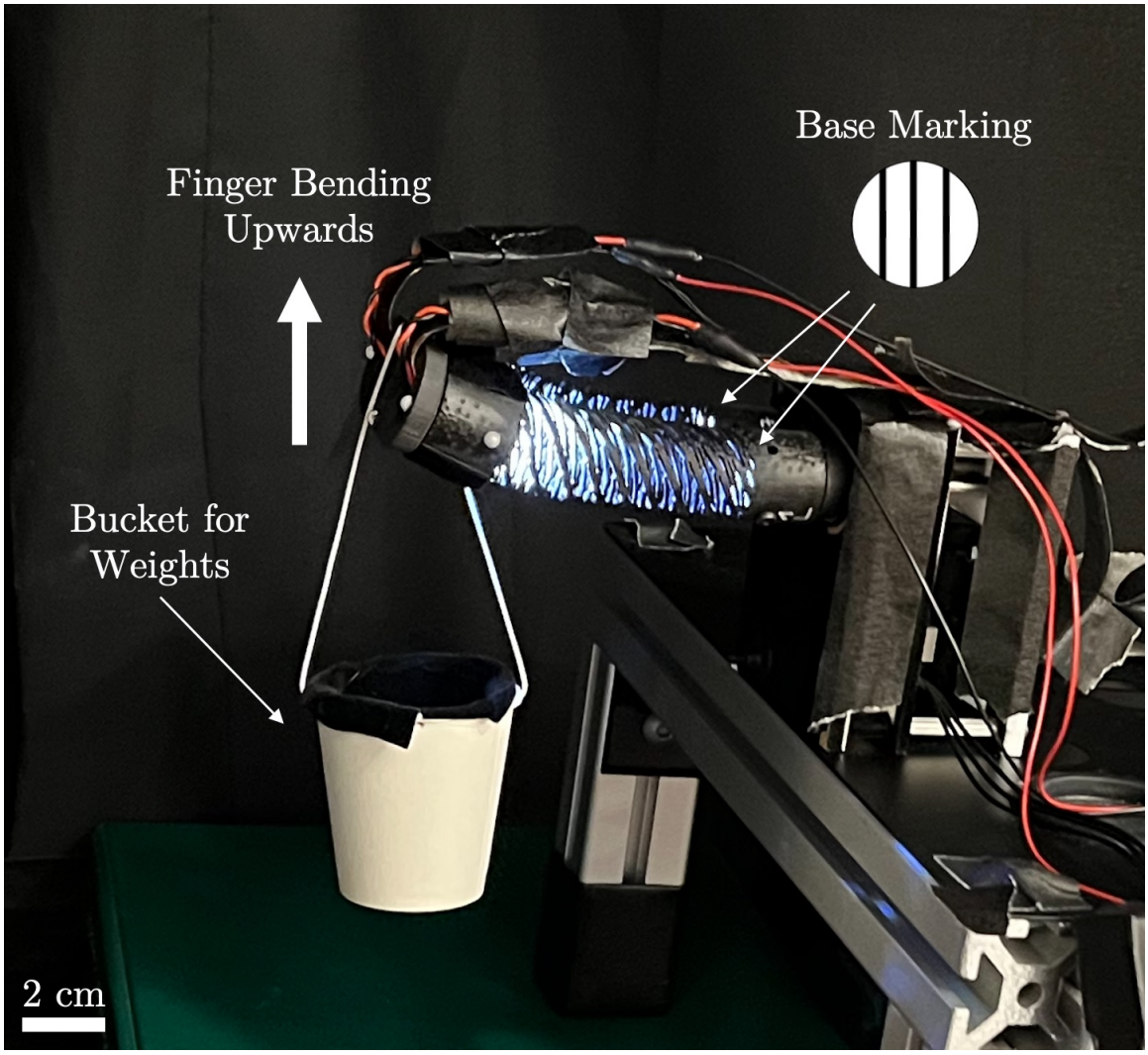

Figure 4-1: **Experimental setup for tip force prediction.** The soft robotic actuator is rotated so bending points in an upward direction. A bucket is attached to the tip of the actuator to hold any added weights. The base marking pattern added to the inside of the HSAs are indicated.

Type	No.	Start [%]	End [%]	v [*]	Motif	Weights [g]
Train	1	0	100	20	move	0
	2	0	100	20	move & hold	0
	3	0	100	20	move	50
	4	0	100	20	move & hold	50
	5	0	100	20	move	100
	6	0	100	20	move & hold	100
	7	0	100	20	move	150
	8	0	100	20	move & hold	150
	9	0	100	20	step	0
	10	0	100	20	step	0
Test	1	0	0	0	hold	{0,50,80,100,120,150}
	2	0	0	0	hold	{0,50,80,100,120,150}

Table 4.2: **Recordings for contact force experiment.** Recordings are split up into training set and test set (Type). Start and End denote configurations of the actuator, given as a percentage of the whole range of motion from the rest configuration (0%) to the fully extended configuration (100%). The servo velocity v is given in units of 0.229 rpm. In the recordings for the training set, three actuation motifs are employed: 1) The actuator is moved back and forth between Start and End without interruptions (move). 2) The actuator is moved between Start and End, but held in the respective extreme position for a few seconds to imitate the actuation motif of a real grasp (move & hold). 3) The actuator is moved to and held at different configurations between Start and End (step). In the test recording, the actuator is held still at the given configuration (hold) and loaded and unloaded successively with the listed weights. Note that weights of 80g and 120g are not included in the training set.

actuator at various positions between 0% and 100% for a few seconds. For the test data, we vary the weight within one take and hold the actuated configuration constant. To evaluate the capabilities of our model to predict previously unseen forces, we include weights that were not part of the training set (80 g, 120 g) in our two test takes. An overview of the collected data set is given in Tab. 4.2. The training and test sets consist of 10,805 and 3,188 data points, respectively.

4.2 Tip Pose Prediction

4.2.1 Model Training

The set of hyperparameters that minimize the validation loss in our trials is given by: a learning rate starting initially at 0.001 and exponentially decaying by 1% per epoch; a quaternion weight of 100; and a CNN architecture shown in Fig. 3-3. The network consists of three convolutional layers with kernel sizes $5\times5\times6$, $5\times5\times6$, and $1\times1\times1$, respectively. Two fully-connected layers with output and input size 512 connect the flattened feature maps with the CNN output. The network has roughly 300,000 trainable parameters, and the training for 100 epochs takes $\approx$ 100 min on a single GPU (NVIDIA GeForce RTX 3090).

4.2.2 Model Evaluation

To obtain our final model for deployment, we retrain with the best hyperparameters on all the available training data (i.e., without splitting off 10% for the validation set) for 300 epochs. To get an unbiased estimate of the model performance in a real-life setting, we evaluate our final model on the test set, which was neither used to train any model nor used to inform the model selection process. The predictions on the test set are shown in Fig. 4-2. Our final model achieves a mean absolute error (MAE) on position predictions of 0.10 mm and therefore achieves the same order of accuracy with which the mocap system measures the real tip position (0.2-0.3 mm). On the scale of the whole system, the tip pose predictions are virtually indistinguishable from

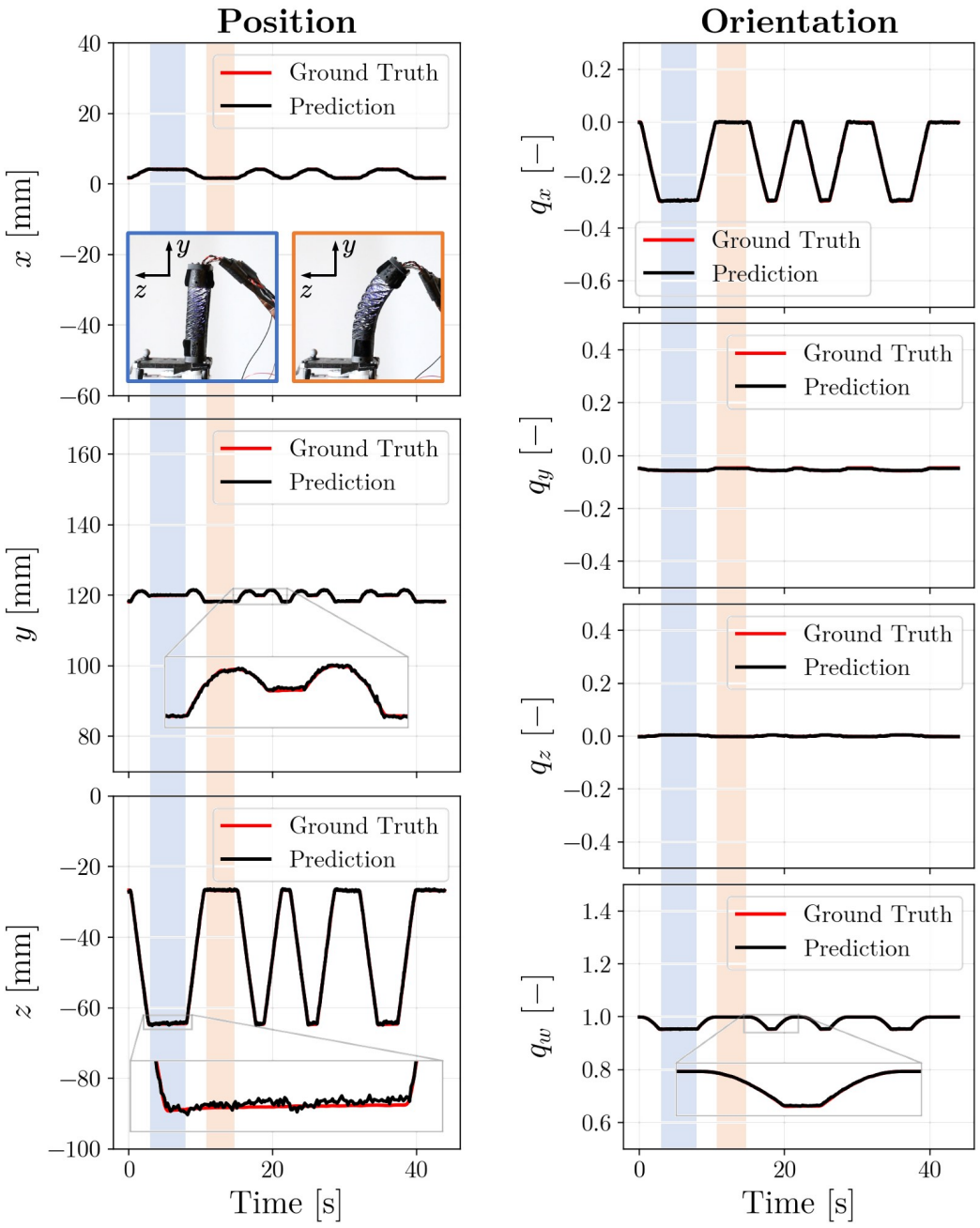

Figure 4-2: **Predictions of tip pose.** Ground truth and model predictions for the tip pose - defined by the position (left column) and orientation (right column) - are provided over time as the red and black plots, respectively. The data corresponds to the actuation sequence from our test set, where the actuator is cyclically bent and held in at-rest or bent configurations for variable periods of time. The photograph insets for x versus time show the actuator in the at-rest (blue outlined image at left) and fully-bent (orange outlined image at right) states. The first instance the actuator is in these states is indicated across all plots by the shaded blue and orange bars.

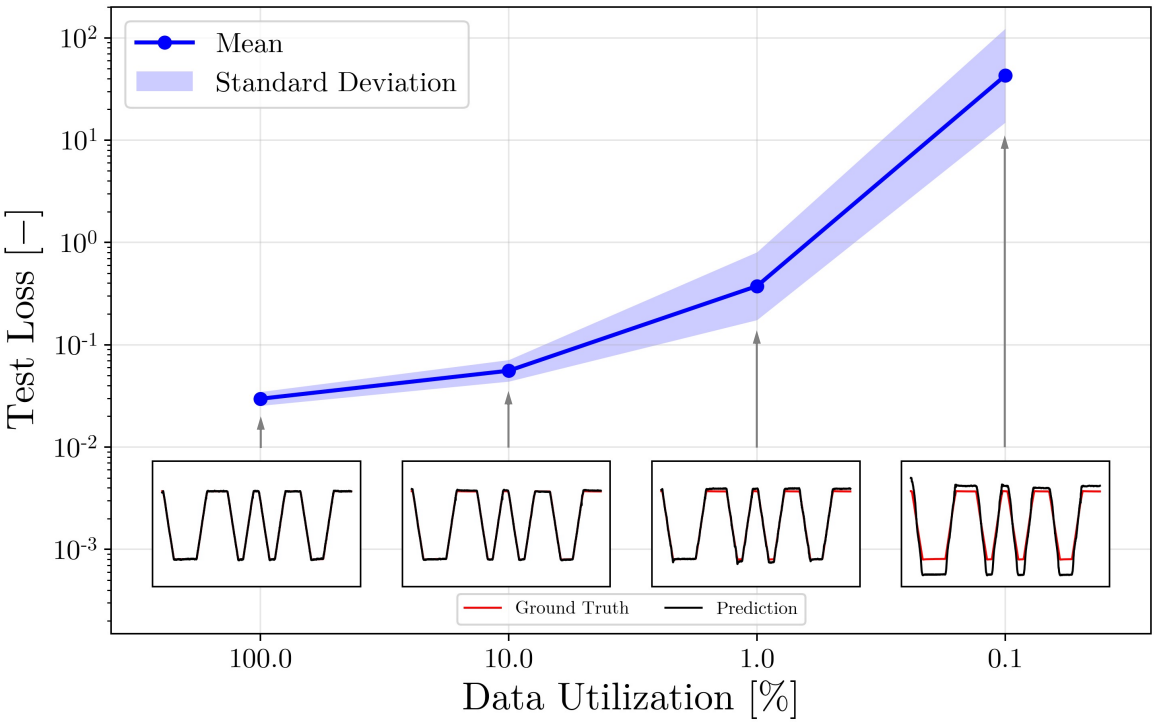

Figure 4-3: **Data utilization analysis.** Training results using various fractions of the available training data. Test loss is plotted as a function of percent data utilization. Inset images show the z-coordinate predictions of tip pose for the same representative actuation sequence for the corresponding percentages of data utilization. Logarithmic error bars are determined following [51].

the ground truth. The average inference time per frame of our model deployed on the Jetson Nano is 18 milliseconds.

4.2.3 Data Utilization

During the training trials for hyperparameter tuning, we notice that the loss converges quickly. This motivates us to systematically study the model performance when trained on a randomly sampled fraction of the available data. The average performance of totally five repetitions for data utilization rates of 100%, 10%, 1%, and 0.1% each is shown in Fig. 4-3. As we decrease the number of samples, the probability of sampling a set of data points that covers the whole range of motion decreases. We therefore expect the error variance to increase, which is consistent with our observations. However, on average we are still able to achieve sub-millimeter accuracy on

position predictions (MAE of 0.16 mm for 10% and 0.42 mm for 1%). This comes at the benefit of significantly reduced training times, from 100 minutes for 100 epochs when using all the data, to 11 minutes when using 10%, and 1.5 minutes when using 1%. We postulate that in free-bending, it is conceivable that our CNN finds a one-to-one mapping between camera images and tip pose easily, since different degrees of bending (and therefore differently looking camera images) correspond to different tip poses. However, when the finger is interacting with the environment, there could be ambiguous mappings and the problem becomes generally much harder. For example, a finger actuated to an intermediate bend (similar to Fig. 3-2, center) could have an additional force pushing the tip back. For the right amount of force, the finger could be almost perfectly vertical and the camera image could look very similar to an unactuated finger with no load.

4.3 Contact Force Prediction

Reusing the CNN architecture for the pose prediction task, we tune the remaining hyperparameters and find that a quaternion weight of 1, a force weight of 10, and a constant learning rate of 0.001 minimize the validation loss. The performance on one test take is shown in Fig. 4-4. Our model achieves an MAE of 0.03 N on force predictions and an MAE of 0.27 mm on position predictions. We note that our model successfully learned to predict forces not included in the training set (see out-of-distribution data in Fig. 4-4). Results on the other test take are similar, with force MAE of 0.04 N, and position MAE of 0.27 mm. Even though a strictly harder problem, our model is still able to match the accuracy of the mocap system for the position predictions.

We note the quasistatic nature of our CNN. Displaying the input-target pairs in the test set in chronological order for convenience could create the impression that our CNN learns the pattern of increasing weights. Since our CNN makes a prediction of current pose and force based solely on the current image, it does not learn any history or time-varying patterns as an RNN would. However, our quasistatic CNN predicts

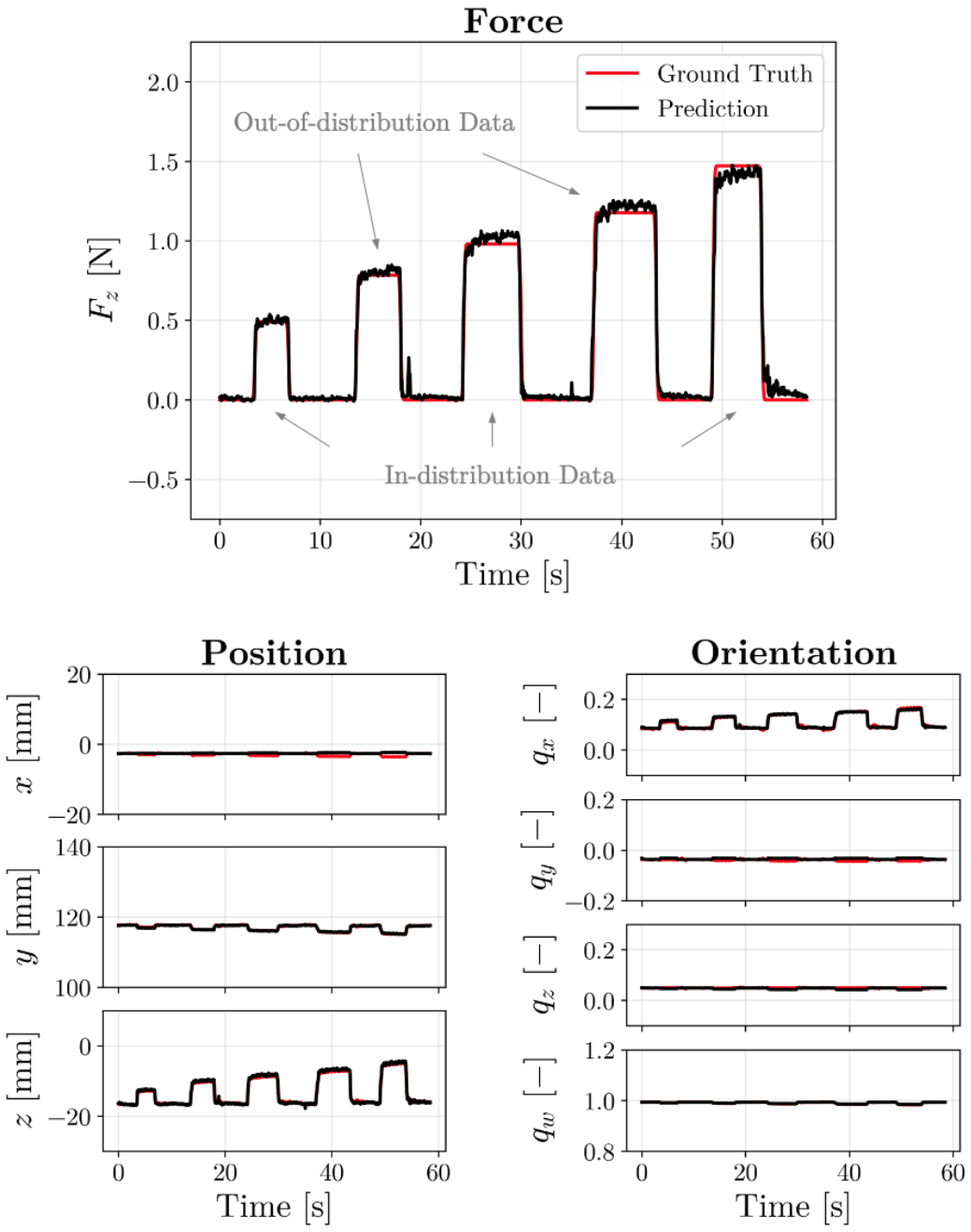

Figure 4-4: **Force predictions.** Force (top) and tip pose predictions, with tip position (bottom left) and orientation (bottom right), are provided over time for one actuation sequence from the test set. Importantly, the model is never trained with 80 g and 120 g weights (corresponding to the second and forth peak, respectively, in the Force vs. Time plot at the top) but still provides a satisfactory prediction. Ground truth and prediction are indicated as the red and black plots, respectively, in all plots (with labels removed in center and right plots for legibility). The mean absolute error (MAE) of the force is 0.03 N, and the MAE of the position is 0.27 mm.

values that are representative of time-dependent, viscoelastic material behaviors. For example, as seen in Fig. 4-4, even though we load our finger with a constant force through the addition and removal of weights, the finger keeps deforming slightly due to viscoelastic creep of the HSA material. This results in our model predicting a force that follows a creep-like response. This issue cannot be alleviated with better model training, but it is inherent to vision-based sensing with soft robots constructed from viscoelastic, architected materials.

Chapter 5

Discussion

5.1 Conclusions

In this work, we have demonstrated a strategy for sensorizing a soft robot through camera-based vision and electrically-driven HSA actuators. We have presented an HSA-based soft finger design that incorporates two miniature cameras and internal lighting control hardware to eliminate the influence of ambient lighting conditions. We developed and trained a CNN using ≈ 6 min of data that provides accurate finger tip pose and contact forces for proprioceptive and tactile sensing capabilities. The experimental data shows that our model's prediction accuracy matches the fidelity with which our ground truth data is captured. In addition, our model predicts the soft finger's tip forces with an absolute error of 0.04 N on forces that range from 0 to 1.5 N.

A key advantage of our strategy is the robustness from both a perception and actuation perspective. The proposed vision-based strategy allows us to introduce a reliable sensorization scheme that is not susceptible to the challenges that material-based approaches usually introduce. This has important implications on the learning side of our work: robust vision-based sensing enables shorter training times, highly lightweight models, and seamless integration with machine vision platforms and GPU-accelerated single-board computers. On the actuation side, the electrically-driven HSAs provide a more reliable actuation approach compared to fluidic actuation.

Though we only required several minutes of training data in the current work, this robust actuation strategy will be essential as we begin training our soft actuators in more dynamic, contact-rich scenarios, an endeavor that will require long operating times to collect the satisfactorily large data sets required.

5.2 Limitations and Future Work

On the hardware side, future improvements to this work include making HSAs from polymer resins that are less susceptible to creep, improving hardware integration to enable somatosensitive manipulation in contact-rich applications (e.g., rummaging through piles of objects), and fusing our vision-based perception methods with other integrated sensors for richer sensory feedback. Furthermore, the hardware design could be improved to minimize obstructions by the camera and LED cables (see Fig. 1-1). Thanks to the durability of the 3D printed HSAs, all of the data was collected on one set of actuators. Investigating the robustness of the trained models to changing the HSA cylinders would show the limitations of our current approach.

On the learning side, more sophisticated methods could be used to model time-dependent effects like long-term material degradation. Even though actuator tip pose is widely used, other state representations for downstream control could include reconstructing the shape of the HSA fingers, extracting and tracking key points along the actuator walls, and modeling the neutral axis of the HSA actuators as spline and learning the spline parameters. Finally, it might even be possible to bypass picking an explicit state representation by simultaneously learning control policies and implicit latent state representations.

Even though best efforts were made to collect data that resembles a real actuation sequence and to shield the inside environment from ambient light, the performance in deployment can only be evaluated by putting the sensor onto a robot arm to perform a real task.

5.3 Lessons Learned

The technical knowledge I have acquired during this project includes dealing with hardware and electronics, designing printed circuit boards, soldering surface-mount components, working with motion capture systems, and writing efficient data loaders for PyTorch. Furthermore, I gained appreciation for readily available and well documented open-source software tools from hardware-level C code to Python libraries for machine learning.

Finally, I would like to elaborate on two key takeaways on the meta level that working on this project resulted in:

Do not overthink. Probably because of my previous research experience in computational mechanics, where it is common to deal with computations that take hours, days, or sometimes weeks, I internalized the importance to get things right at the very beginning. However, during this project, especially working with hardware, I realized the impossibility to account for all kinds of outcomes beforehand and the difficulty to nail the design at the first iteration. In most cases, I found the best approach was to *just do it*: Prototype, identify weaknesses, fix them, iterate.

Most useful things are not taught in classes. Many of the things that greatly improve my day-to-day workflow were not taught in any of the courses I took: efficient data handling, version control, shell scripting, debugging, proper use of the command line. To that end, I can really recommend the course "The Missing Semester of Your CS Education", taught during IAP at MIT and available at full length online[1].

[1] `https://missing.csail.mit.edu/`

Appendix A

Figures

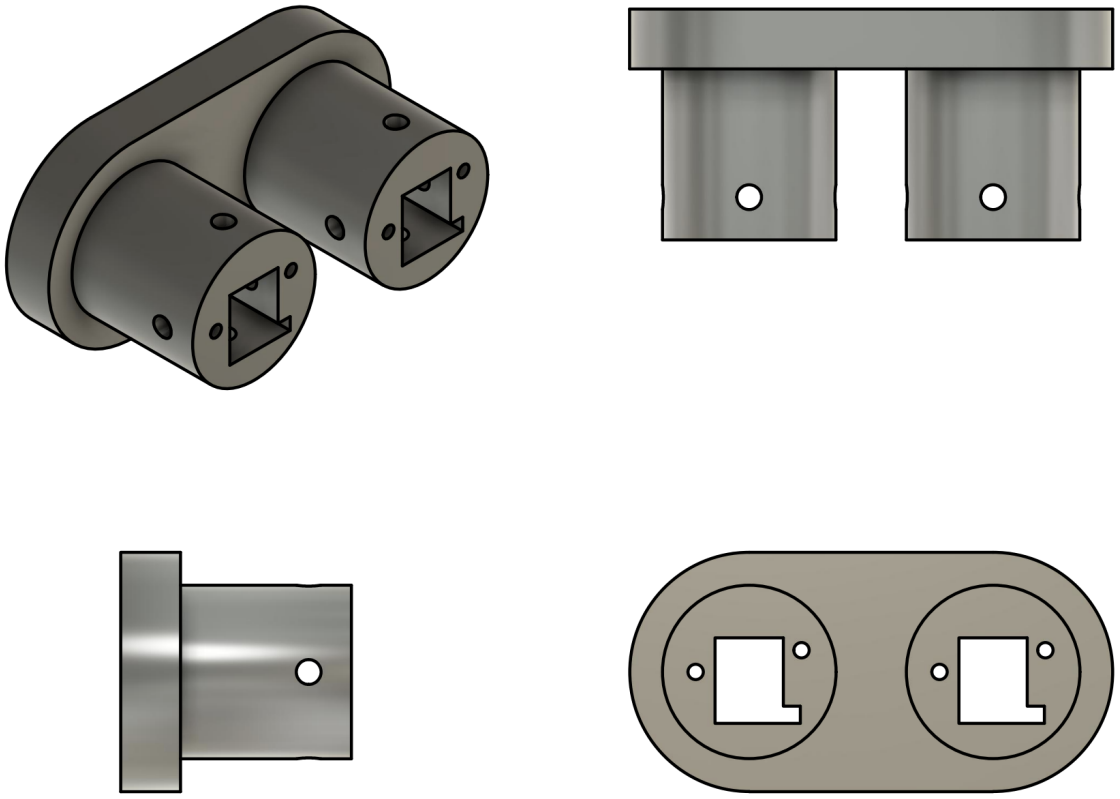

Figure A-1: **Top mount design.** Modified design of the top mount from [3], with additional cutouts to accommodate for LED power connectors (square), camera cables (slit), and PCB fixation (small holes).

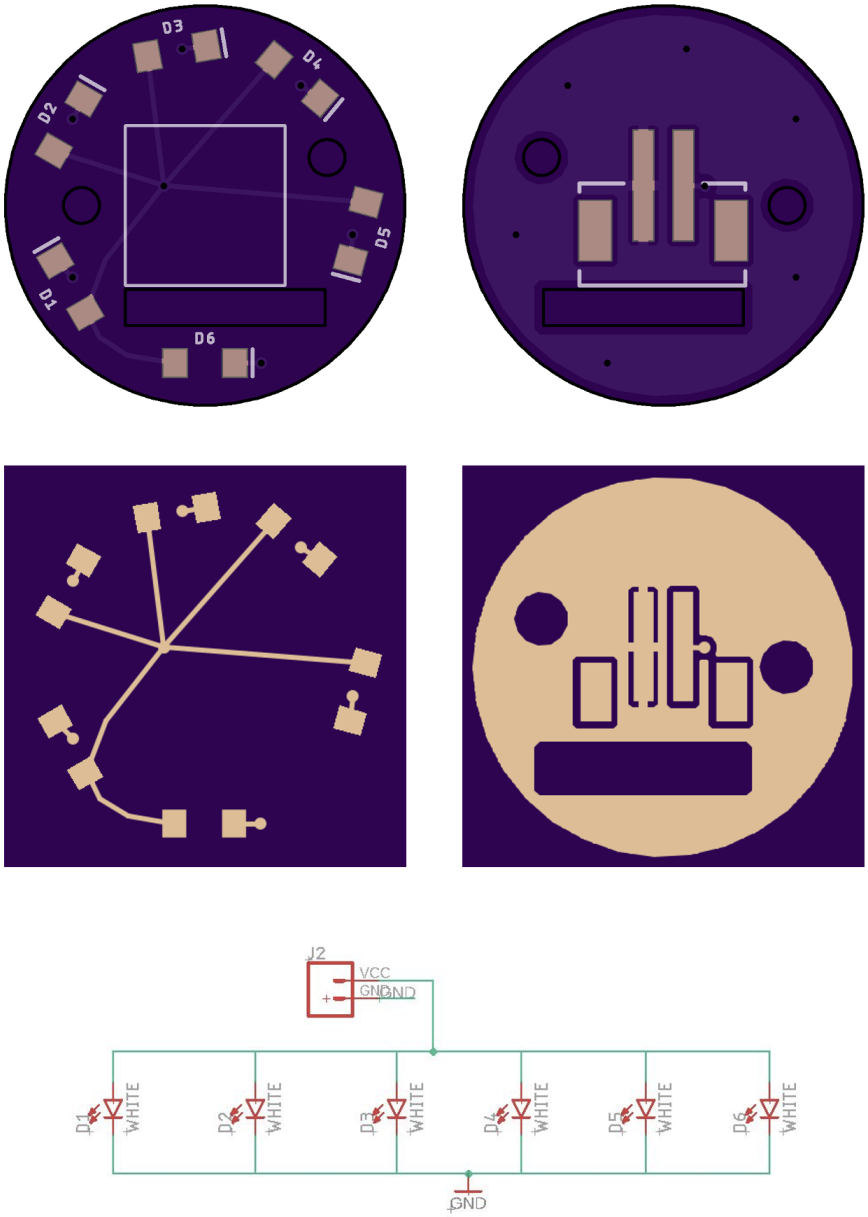

Figure A-2: **PCB design for camera and LEDs. Top**: Front view (left) of the PCB, with the exposed copper to connect 6 surface-mount LEDs. Back view (right) with the exposed copper to fixate (outer two) and to connect (inner two) the JST connector. **Middle**: Full top (left) and bottom (right) copper layers. **Bottom**: The underlying circuit diagram of the PCB. We note that we use an external circuit to control the current supplied to this board. PCB drawings are provided by the manufacturer, OSH Park. Schematic is created using Autodesk EAGLE.

Bibliography

[1] Eric Brown, Nicholas Rodenberg, John Amend, Annan Mozeika, Erik Steltz, Mitchell R Zakin, Hod Lipson, and Heinrich M Jaeger. Universal robotic gripper based on the jamming of granular material. *Proceedings of the National Academy of Sciences*, 107(44):18809–18814, 2010.

[2] Keene Chin, Tess Hellebrekers, and Carmel Majidi. Machine learning for soft robotic sensing and control. *Advanced Intelligent Systems*, 2(6):1900171, 2020.

[3] Lillian Chin, Jeffrey Lipton, Robert MacCurdy, John Romanishin, Chetan Sharma, and Daniela Rus. Compliant electric actuators based on handed shearing auxetics. In *2018 IEEE International Conference on Soft Robotics (RoboSoft)*, pages 100–107. IEEE, 2018.

[4] Raphael Deimel and Oliver Brock. A novel type of compliant and underactuated robotic hand for dexterous grasping. *The International Journal of Robotics Research*, 35(1-3):161–185, 2016.

[5] Cosimo Della Santina, Manuel G Catalano, and Antonio Bicchi. Soft robots. *Encyclopedia of Robotics*, 489, 2020.

[6] Cosimo Della Santina, Robert K. Katzschmann, Antonio Biechi, and Daniela Rus. Dynamic control of soft robots interacting with the environment. In *2018 IEEE International Conference on Soft Robotics (RoboSoft)*, pages 46–53, 2018.

[7] Cosimo Della Santina, Cristina Piazza, Gian Maria Gasparri, Manuel Bonilla, Manuel Giuseppe Catalano, Giorgio Grioli, Manolo Garabini, and Antonio Bicchi. The quest for natural machine motion: An open platform to fast-prototyping articulated soft robots. *IEEE Robotics & Automation Magazine*, 24(1):48–56, 2017.

[8] Pascal Auf der Maur, Betim Djambazi, Yves Haberthür, Patricia Hörmann, Alexander Kübler, Michael Lustenberger, Samuel Sigrist, Oda Vigen, Julian Förster, Florian Achermann, et al. Roboa: Construction and evaluation of a steerable vine robot for search and rescue applications. In *2021 IEEE 4th International Conference on Soft Robotics (RoboSoft)*, pages 15–20. IEEE, 2021.

[9] Ze Yang Ding, Junn Yong Loo, Vishnu Monn Baskaran, Surya Girinatha Nurzaman, and Chee Pin Tan. Predictive Uncertainty Estimation Using Deep Learning for Soft Robot Multimodal Sensing. *IEEE Robotics and Automation Letters*, 6(2):951–957, 2021.

[10] Aman Garg, Ian Good, Daniel Revier, Kevin Airis, and Jeffrey Lipton. Kinematic modeling of handed shearing auxetics via piecewise constant curvature. *arXiv preprint arXiv:2112.04706*, 2021.

[11] Giada Gerboni, Alessandro Diodato, Gastone Ciuti, Matteo Cianchetti, and Arianna Menciassi. Feedback Control of Soft Robot Actuators via Commercial Flex Bend Sensors. *IEEE/ASME Transactions on Mechatronics*, 22(4):1881–1888, 2017.

[12] Ian Good, Tosh Brown-Moore, Aditya Patil, Daniel Revier, and Jeffrey Ian Lipton. Expanding the design space for electrically-driven soft robots through handed shearing auxetics. *arXiv preprint arXiv:2110.00669*, 2021.

[13] Markus Grebenstein, Alin Albu-Schäffer, Thomas Bahls, Maxime Chalon, Oliver Eiberger, Werner Friedl, Robin Gruber, Sami Haddadin, Ulrich Hagn, Robert Haslinger, et al. The dlr hand arm system. In *2011 IEEE International Conference on Robotics and Automation*, pages 3175–3182. IEEE, 2011.

[14] Elliot W Hawkes, Laura H Blumenschein, Joseph D Greer, and Allison M Okamura. A soft robot that navigates its environment through growth. *Science Robotics*, 2(8), 2017.

[15] Bianca S Homberg, Robert K Katzschmann, Mehmet R Dogar, and Daniela Rus. Haptic identification of objects using a modular soft robotic gripper. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1698–1705, 2015.

[16] Josie Hughes, Francesco Stella, Cosimo Della Santina, and Daniela Rus. Sensing Soft Robot Shape Using IMUs: An Experimental Investigation. In *International Symposium on Experimental Robotics*, pages 543–552. Springer, 2020.

[17] Filip Ilievski, Aaron D Mazzeo, Robert F Shepherd, Xin Chen, and George M Whitesides. Soft robotics for chemists. *Angewandte Chemie*, 123(8):1930–1935, 2011.

[18] Jaewoong Jung, Myungsun Park, DongWook Kim, and Yong-Lae Park. Optically Sensorized Elastomer Air Chamber for Proprioceptive Sensing of Soft Pneumatic Actuators. *IEEE Robotics and Automation Letters*, 5(2):2333–2340, 2020.

[19] Robert K. Katzschmann, Cosimo Della Santina, Yasunori Toshimitsu, Antonio Bicchi, and Daniela Rus. Dynamic motion control of multi-segment soft robots using piecewise constant curvature matched with an augmented rigid body model. In *2019 2nd IEEE International Conference on Soft Robotics (RoboSoft)*, pages 454–461, 2019.

[20] Alex Kendall and Roberto Cipolla. Geometric loss functions for camera pose regression with deep learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5974–5983, 2017.

[21] Sangbae Kim, Cecilia Laschi, and Barry Trimmer. Soft robotics: a bioinspired evolution in robotics. *Trends in biotechnology*, 31(5):287–294, 2013.

[22] Kenji Kure, Takefumi Kanda, Koichi Suzumori, and Shuichi Wakimoto. Intelligent FMA using flexible displacement sensor with paste injection. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1012–1017, 2006.

[23] Jessica Lanini, Hamed Razavi, Julen Urain, and Auke Ijspeert. Human intention detection as a multiclass classification problem: Application in physical human–robot interaction while walking. *IEEE Robotics and Automation Letters*, 3(4):4171–4178, 2018.

[24] C. Larson, B. Peele, S. Li, S. Robinson, M. Totaro, L. Beccai, B. Mazzolai, and R. Shepherd. Highly stretchable electroluminescent skin for optical signaling and tactile sensing. *Science*, 351(6277):1071–1074, 2016.

[25] Cecilia Laschi, Matteo Cianchetti, Barbara Mazzolai, Laura Margheri, Maurizio Follador, and Paolo Dario. Soft robot arm inspired by the octopus. *Advanced robotics*, 26(7):709–727, 2012.

[26] Cecilia Laschi, Barbara Mazzolai, and Matteo Cianchetti. Soft robotics: Technologies and systems pushing the boundaries of robot abilities. *Science Robotics*, 1(1):eaah3690, 2016.

[27] Alexander LeNail. Nn-svg: Publication-ready neural network architecture schematics. *Journal of Open Source Software*, 4(33):747, 2019.

[28] Shuguang Li, Samer A Awale, Katharine E Bacher, Thomas J Buchner, Cosimo Della Santina, Robert J Wood, and Daniela Rus. Scaling Up Soft Robotics: A Meter-Scale, Modular, and Reconfigurable Soft Robotic System. *Soft Robotics*, 2021.

[29] Shuguang Li, John J Stampfli, Helen J Xu, Elian Malkin, Evelin Villegas Diaz, Daniela Rus, and Robert J Wood. A vacuum-driven origami "magic-ball" soft gripper. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 7401–7408. IEEE, 2019.

[30] Jeffrey Ian Lipton, Robert MacCurdy, Zachary Manchester, Lillian Chin, Daniel Cellucci, and Daniela Rus. Handedness in shearing auxetics creates rigid and compliant structures. *Science*, 360(6389):632–635, 2018.

[31] Junn Yong Loo, Ze Yang Ding, Vishnu Monn Baskaran, Surya Girinatha Nurzaman, and Chee Pin Tan. Robust Multimodal Indirect Sensing for Soft Robots Via Neural Network-Aided Filter-Based Estimation. *Soft Robotics*, 2021.

[32] Ming Luo, Erik H Skorina, Weijia Tao, Fuchen Chen, Selim Ozel, Yinan Sun, and Cagdas D Onal. Toward modular soft robotics: Proprioceptive curvature sensing and sliding-mode control of soft bidirectional bending modules. *Soft Robotics*, 4(2):117–125, 2017.

[33] Siddharth Mahendran, Haider Ali, and René Vidal. 3d pose regression using convolutional neural networks. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 2174–2182, 2017.

[34] William McMahan, V Chitrakaran, M Csencsits, D Dawson, Ian D Walker, Bryan A Jones, M Pritts, D Dienno, M Grissom, and Christopher D Rahn. Field trials and testing of the octarm continuum manipulator. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, pages 2336–2341. IEEE, 2006.

[35] John Morrow, Hee Sup Shin, Calder Phillips-Grafflin, Sung Hwan Jang, Jacob Torrey, Riley Larkins, Steven Dang, Yong Lae Park, and Dmitry Berenson. Improving Soft Pneumatic Actuator Fingers through Integration of Soft Sensors, Position and Force Control, and Rigid Fingernails. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 5024–5031, 2016.

[36] Bobak Mosadegh, Panagiotis Polygerinos, Christoph Keplinger, Sophia Wennstedt, Robert F Shepherd, Unmukt Gupta, Jongmin Shim, Katia Bertoldi, Conor J Walsh, and George M Whitesides. Pneumatic networks for soft robotics that actuate rapidly. *Advanced functional materials*, 24(15):2163–2170, 2014.

[37] Francesca Negrello, Alessandro Settimi, Danilo Caporale, Gianluca Lentini, Mattia Poggiani, Dimitrios Kanoulas, Luca Muratore, Emanuele Luberto, Gaspare Santaera, Luca Ciarleglio, et al. Humanoids at work: The walk-man robot in a postearthquake scenario. *IEEE Robotics & Automation Magazine*, 25(3):8–22, 2018.

[38] Selim Ozel, Erik H. Skorina, Ming Luo, Weijia Tao, Fuchen Chen, Yixiao Pan, and Cagdas D. Onal. A composite soft bending actuation module with integrated curvature sensing. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4963–4968, 2016.

[39] Panagiotis Polygerinos, Nikolaus Correll, Stephen A Morin, Bobak Mosadegh, Cagdas D Onal, Kirstin Petersen, Matteo Cianchetti, Michael T Tolley, and Robert F Shepherd. Soft robotics: Review of fluid-driven intrinsically soft devices; manufacturing, sensing, control, and applications in human-robot interaction. *Advanced Engineering Materials*, 19(12):1700016, 2017.

[40] Panagiotis Polygerinos, Zheng Wang, Kevin C Galloway, Robert J Wood, and Conor J Walsh. Soft robotic glove for combined assistance and at-home rehabilitation. *Robotics and Autonomous Systems*, 73:135–143, 2015.

[41] Tommaso Ranzani, Giada Gerboni, Matteo Cianchetti, and A Menciassi. A bioinspired soft manipulator for minimally invasive surgery. *Bioinspiration & biomimetics*, 10(3):035008, 2015.

[42] Ellen T Roche, Markus A Horvath, Isaac Wamala, Ali Alazmani, Sang-Eun Song, William Whyte, Zurab Machaidze, Christopher J Payne, James C Weaver, Gregory Fishbein, et al. Soft robotic sleeve supports heart function. *Science translational medicine*, 9(373):eaaf3925, 2017.

[43] Daniela Rus and Michael T Tolley. Design, fabrication and control of soft robots. *Nature*, 521(7553):467–475, 2015.

[44] Siddharth Sanan, Peter S Lynn, and Saul T Griffith. Pneumatic torsional actuators for inflatable robots. *Journal of Mechanisms and Robotics*, 6(3):031003, 2014.

[45] Sina Sareh, Yohan Noh, Min Li, Tommaso Ranzani, Hongbin Liu, and Kaspar Althoefer. Macrobend optical sensing for pose measurement in soft robot arms. *Smart Materials and Structures*, 24(12):125024, 2015.

[46] Arthur Seibel and Lars Schiller. Integrated curvature sensing of soft bending actuators using inertial measurement units. In *ASME International Mechanical Engineering Congress and Exposition*, volume 52149, page V009T12A034, 2018.

[47] Sangok Seok, Cagdas D Onal, Robert Wood, Daniela Rus, and Sangbae Kim. Peristaltic locomotion with antagonistic actuators in soft robotics. In *2010 IEEE international conference on robotics and automation*, pages 1228–1233. IEEE, 2010.

[48] Yu She, Sandra Q Liu, Peiyu Yu, and Edward Adelson. Exoskeleton-covered soft finger with vision-based proprioception and tactile sensing. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10075–10081. IEEE, 2020.

[49] Benjamin Shih, Dylan Drotman, Caleb Christianson, Zhaoyuan Huo, Ruffin White, Henrik I. Christensen, and Michael T. Tolley. Custom soft robotic gripper sensor skins for haptic object visualization. In *IEEE International Conference on Intelligent Robots and Systems*, volume 2017-Septe, pages 494–501, 2017.

[50] Gabor Soter, Andrew Conn, Helmut Hauser, and Jonathan Rossiter. Bodily aware soft robots: integration of proprioceptive and exteroceptive sensors. In *2018 IEEE international conference on robotics and automation (ICRA)*, pages 2448–2453. IEEE, 2018.

[51] Eric M Stuve. Estimating and plotting logarithmic error bars, 2004.

[52] Koichi Suzumori, Shoichi Iikura, and Hiroshisa Tanaka. Applying a flexible microactuator to robotic mechanisms. *IEEE Control systems magazine*, 12(1):21–27, 1992.

[53] Ken Takaki, Yoshitaka Taguchi, Satoshi Nishikawa, Ryuma Niiyama, and Yoshihiro Kawahara. Acoustic length sensor for soft extensible pneumatic actuators with a frequency characteristics model. *IEEE Robotics and Automation Letters*, 4(4):4292–4297, 2019.

[54] Thomas George Thuruthel, Benjamin Shih, Cecilia Laschi, and Michael Thomas Tolley. Soft robot perception using embedded soft sensors and recurrent neural networks. *Science Robotics*, 4(26), 2019.

[55] Celeste To, Tess Lee Hellebrekers, and Yong-Lae Park. Highly stretchable optical sensors for pressure, strain, and curvature measurement. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5898–5903, 2015.

[56] Michael T Tolley, Robert F Shepherd, Bobak Mosadegh, Kevin C Galloway, Michael Wehner, Michael Karpelson, Robert J Wood, and George M Whitesides. A resilient, untethered soft robot. *Soft Robotics*, 1(3):213–223, 2014.

[57] Deepak Trivedi, Christopher D Rahn, William M Kier, and Ian D Walker. Soft robotics: Biological inspiration, state of the art, and future research. *Applied bionics and biomechanics*, 5(3):99–117, 2008.

[58] Ryan L Truby, Lillian Chin, and Daniela Rus. A Recipe for Electrically-Driven Soft Robots via 3D Printed Handed Shearing Auxetics. *IEEE Robotics and Automation Letters*, 6(2):795–802, 2021.

[59] Ryan L Truby, Cosimo Della Santina, and Daniela Rus. Distributed proprioception of 3D configuration in soft, sensorized robots via deep learning. *IEEE Robotics and Automation Letters*, 5(2):3299–3306, 2020.

[60] Ryan L Truby, Robert K Katzschmann, Jennifer A Lewis, and Daniela Rus. Soft robotic fingers with embedded ionogel sensors and discrete actuation modes for somatosensitive manipulation. In *2019 2nd IEEE International Conference on Soft Robotics (RoboSoft)*, pages 322–329, 2019.

[61] Ryan L Truby, Michael Wehner, Abigail K Grosskopf, Daniel M Vogt, Sebastien GM Uzel, Robert J Wood, and Jennifer A Lewis. Soft somatosensitive actuators via embedded 3D printing. *Advanced Materials*, 30(15):1706383, 2018.

[62] Hideyuki Tsukagoshi, Ato Kitagawa, and Mitsuru Segawa. Active hose: An artificial elephant's nose with maneuverability for rescue operation. In *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No. 01CH37164)*, volume 3, pages 2454–2459. IEEE, 2001.

[63] Vincent Wall, Gabriel Zoller, and Oliver Brock. A method for sensorizing soft actuators and its application to the RBO hand 2. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4965–4970, 2017.

[64] Hongbo Wang, Massimo Totaro, and Lucia Beccai. Toward Perceptive Soft Robots: Progress and Challenges. *Advanced Science*, 5:1800541, 2018.

[65] Ruoyu Wang, Shiheng Wang, Songyu Du, Erdong Xiao, Wenzhen Yuan, and Chen Feng. Real-Time Soft Body 3D Proprioception via Deep Vision-Based Sensing. *IEEE Robotics and Automation Letters*, 5(2):3382–3389, 2020.

[66] Peter Werner, Matthias Hofer, Carmelo Sferrazza, and Raffaello D'Andrea. Vision-based proprioceptive sensing: tip position estimation for a soft inflatable bellow actuator. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8889–8896. IEEE, 2020.

[67] Osman Dogan Yirmibesoglu and Yigit Menguc. Hybrid soft sensor with embedded IMUs to measure motion. In *2016 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 798–804, 2016.

[68] Wenzhen Yuan, Rui Li, Mandayam A. Srinivasan, and Edward H. Adelson. Measurement of shear and slip with a GelSight tactile sensor. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 304–311, 2015.

[69] Annan Zhang, Ryan L Truby, Lillian Chin, Shuguang Li, and Daniela Rus. Perception through vision for soft robotic fingers based on handed shearing auxetics. 2022.

[70] Huichan Zhao, Kevin O'Brien, Shuo Li, and Robert F Shepherd. Optoelectronically innervated soft prosthetic hand via stretchable optical waveguides. *Science Robotics*, 1(1), 2016.

[71] Gabriel Zöller, Vincent Wall, and Oliver Brock. Acoustic Sensing for Soft Pneumatic Actuators. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6986–6991, 2018.