A Vision-Based Sensing Approach for a Spherical Soft Robotic Arm

Matthias Hofer*, Carmelo Sferrazza and Raffaello D'Andrea

Institute for Dynamic Systems and Control, ETH Zurich, Zurich, Switzerland

Sensory feedback is essential for the control of soft robotic systems and to enable deployment in a variety of different tasks. Proprioception refers to sensing the robot's own state and is of crucial importance in order to deploy soft robotic systems outside of laboratory environments, i.e. where no external sensing, such as motion capture systems, is available. A vision-based sensing approach for a soft robotic arm made from fabric is presented, leveraging the high-resolution sensory feedback provided by cameras. No mechanical interaction between the sensor and the soft structure is required and consequently the compliance of the soft system is preserved. The integration of a camera into an inflatable, fabric-based bellow actuator is discussed. Three actuators, each featuring an integrated camera, are used to control the spherical robotic arm and simultaneously provide sensory feedback of the two rotational degrees of freedom. A convolutional neural network architecture predicts the two angles describing the robot's orientation from the camera images. Ground truth data is provided by a motion capture system during the training phase of the supervised learning approach and its evaluation thereafter. The camera-based sensing approach is able to provide estimates of the orientation in real-time with an accuracy of about one degree. The reliability of the sensing approach is demonstrated by using the sensory feedback to control the orientation of the robotic arm in closed-loop.

Keywords: soft robotics, proprioception, vision-based sensing, computer vision, supervised machine learning, pneumatic actuation, fabric bellows

INTRODUCTION

Soft robots show promise to overcome challenges encountered with rigid robots due to the versatility resulting from the soft materials employed (Polygerinos et al., 2017). Their intrinsic mechanical properties are beneficial in terms of safety, allowing for close human-robot collaboration (Abidi and Cianchetti, 2017). The academic relevance of the field is reflected by an increasing number of publications and growing attention within the field of robotics in general (Bao et al., 2018). However, the potential benefits of soft robots come with several challenges, such as the complex dynamics that are difficult to model and limit the application of open-loop control (Rus and Tolley, 2015). Therefore, sensory feedback is indispensable for accurate control and deployment in real-word applications (Wang et al., 2018).

A wide range of sensing principles are explored to provide proprioceptive feedback, i.e. feedback of the robot's own state. Vision-based approaches relying on internal cameras to observe the deformation of soft materials are promising because the sensor provides a high resolution and is not required to mechanically interact with the soft material that is observed.
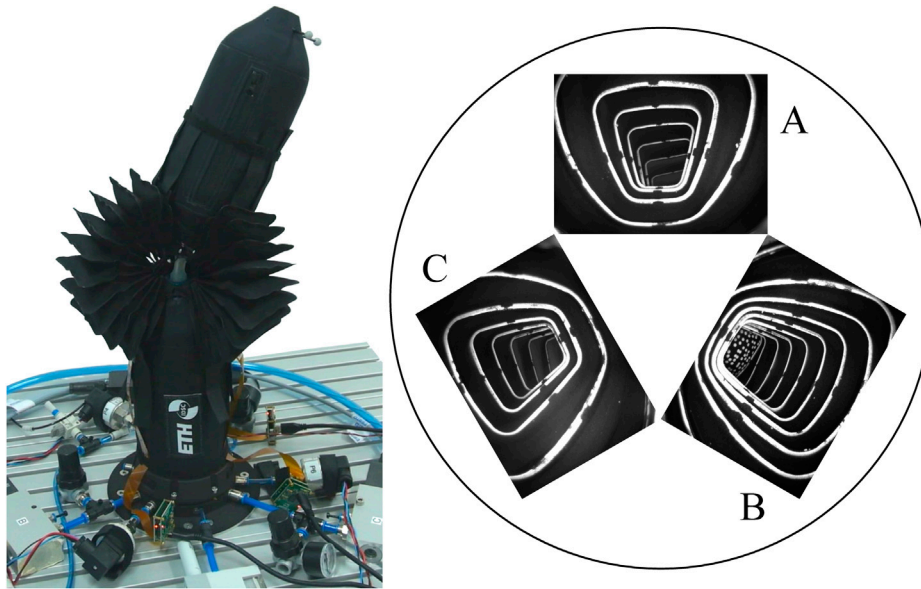
FIGURE 1 | The figure on the left shows the spherical robotic arm used for evaluation of the vision-based sensing approach proposed in this work. The figure on the right shows images from the cameras placed inside three inflatable bellow actuators. The arrangement of the camera images matches a view from the bottom looking upwards. The orientation of the movable link can be observed in certain actuator elongations and deformations that are observed by the internal cameras.

Our method originates from the work documented in Werner et al. (2020). The principle idea is to integrate a camera into a fabric-based bellow actuator and use three of these actuators to control a spherical robotic arm (see **Figure 1**). The two rotational degrees of freedom of the robotic arm are estimated from the actuator expansion and deformation observed by the three internal cameras.

The mapping from camera images to the orientation of the movable link is identified by relying on a supervised learning approach with ground truth information provided by a motion capture system. A convolutional neural network architecture maps the camera images to the orientation of the robot arm. The sensing pipeline developed is deployed in real-time and used for closed-loop control of the robotic arm.

Related Work

A number of different approaches are investigated to retrieve the shape of a soft robot based on internal sensors only (Wang et al., 2018). A common approach is to combine sensing algorithms with machine learning techniques to retrieve the quantity of interest from the raw sensor output. An overview of such applications for sensing and control is provided in Chin et al. (2020). Any sensor relies on the change in a physical property induced by the soft structure undergoing a deformation or expansion. Different sensing principles and the applications thereof are summarized below.

Resistive and piezoresistive strain sensors detect a change in resistance caused by material deformation (Stassi et al., 2014). The advantage of resistive and piezoresistive sensors is their relatively easy fabrication and integration (Wang et al., 2018). An application of a resistive sensor is presented in Thuruthel et al.

(2019), in combination with a recurrent neural network that maps the raw sensor output to the bending state of a soft finger. Piezoresistive sensors are employed in Truby et al. (2020) to form a soft, proprioceptive sensor skin, which can be attached to a soft robotic system. A recurrent neural network predicts the robot configuration based on the sensor measurements.

A capacitive strain sensor is presented in Shintake et al. (2018) and deployed for an intelligent glove application. An approach based on the change of inductance is presented in Felt (2019) for a bellows-driven continuum robot and used in closed-loop for feedback control. A sensing approach relying on a magnet and a Hall sensor integrated into a soft bending actuator is documented in Luo et al. (2017). The relative orientation between the magnet and sensor varies as the soft structure deforms, causing the observed magnetic field to change. The sensing type is simple to integrate and can be used to control the bending angle of the actuator. The method presented in Takaki et al. (2019) leverages an acoustic sensing principle. A speaker and microphone are integrated into a soft extensible pneumatic actuator and used to detect the changing resonance characteristics as the elongation of the actuator varies. The sensing approach is used in closed-loop to track the desired length of the actuator. The work presented in Grzesiak et al. (2011) relies on commercially readily available Bowden cable potentiometers to retrieve and control the shape of a continuum robot arm.

Optical sensors detect changes in the light transmission of a soft medium when deformed. A common approach is to measure the varying light intensity. The integration of macrobend stretch sensors into a soft arm is documented in Sareh et al. (2015). The bending of the light transmitting fiber causes the intensity of the

light transmission to change. The use of stretchable optical waveguides is reported in Zhao et al. (2016a) to provide sensing capabilities for shape and force for a prosthetic hand. Closed-loop control is also demonstrated in Zhao et al. (2016b). A fabric-based bellow actuator, similar to ours, is used in Yang et al. (2019) as the light reflecting surface. A photo transistor is attached to one end of the linear bellow actuator and measures the light intensity from a light emitting diode (LED) attached to the opposing end of the actuator. The intensity decreases as a function of the actuator elongation. The advantage of optical sensors is their high level of sensitivity and repeatability (Kappassov et al., 2015) and the fact that the electronics can be placed outside the sensing area (Wang et al., 2018).

Another optical sensing principle relies on fiber Bragg grating. The use of a distributed fiber Bragg sensor network is demonstrated in Wang et al. (2016) for a cone-shaped soft manipulator made from silicone.

The discussion of camera-based sensing approaches is limited here to examples relying on internal cameras. Methods purely based on external cameras, including motion capture systems, are not discussed. Cameras visually observe material deformation through the movement of visual features located in or attached to the soft material. Camera-based sensing is actively explored in the field of tactile sensing with an overview provided in Shimonomura (2019).

Compared to placing the cameras externally, an advantage of integrating cameras into the soft system and pointing them to the interior of the structure is the possibility to design the area observed by the camera for best performance without external influences. The application of a pattern to the interior surface of the structure allows for the provision of rich information about the deformation state and control of the lighting conditions. Consequently, the sensing approach does not depend on the visual features present in the environment or the existing external lighting conditions.

A vision-based tactile sensor including pneumatic actuation is presented in McInroe et al. (2018). A combination of blob detection and optical flow is used to track a number of markers and infer contact conditions and membrane shear. Increasing the internal pressure allows for inflation of the membrane and thereby control of the interaction force. A tactile sensor named TacEA combines vision-based tactile sensing, pneumatic actuation and electroadhesive grasping capabilities and is presented in Xiang et al. (2019). The sensing principle relies on the TacTip family as presented in Ward-Cherrier et al. (2018). After an object is gripped using the electroadhesion, releasing the object can take a considerable amount of time. Pneumatic actuation, i.e. inflation of the soft membrane, allows the object to be released quickly. Other camera-based tactile sensors are presented in Yuan et al. (2017) and Sferrazza and D'Andrea (2019).

A method to sense the three-dimensional shape of a soft robot relying on a self-observing camera is documented in Wang et al. (2020). External depth cameras provide ground truth to train a neural network, which predicts the shape of the object only from images of the self-observing cameras. The approach is executed on a graphics processing unit (GPU) and provides the three-dimensional deformation of soft objects in real-time. A vision-based sensing approach providing both proprioceptive and exteroceptive sensing is demonstrated in She et al. (2020) for an exoskeleton-covered soft finger. The sensing method relies on a convolutional neural network architecture being executed on a GPU that is able to predict the shape of a single finger in real-time and to classify objects which are grasped with a gripper made from two fingers.
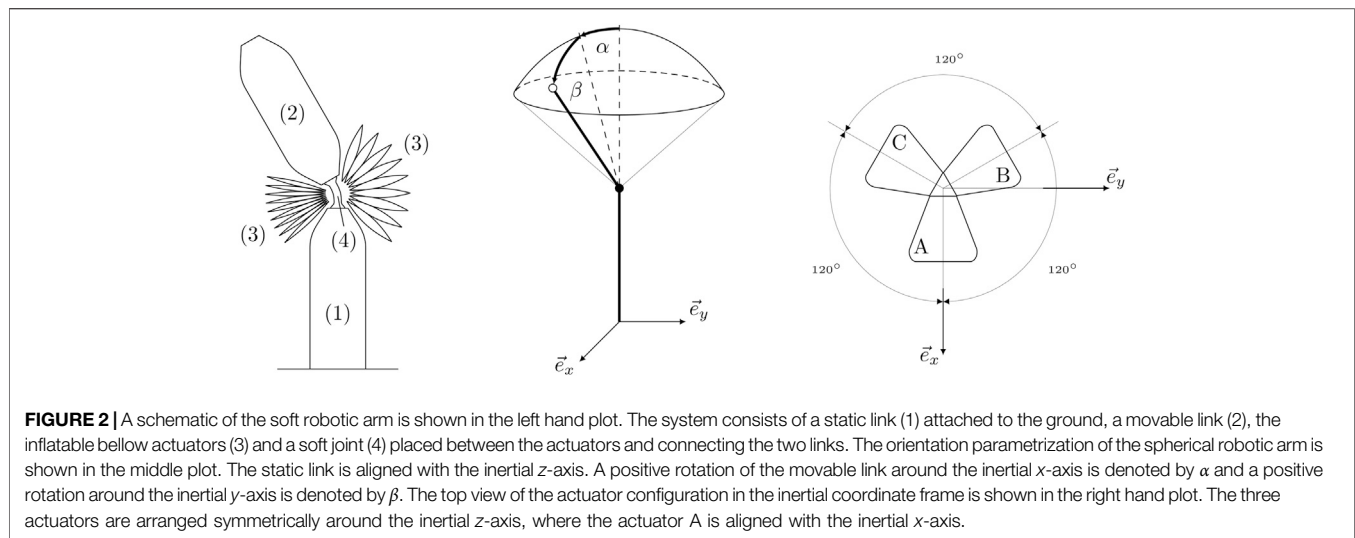
In Oliveira et al. (2020), a sensing method is presented to measure the bending deformation of a soft link and detect interactions with the environment. A camera is mounted inside an inflatable and compliant link. A blob detection algorithm relates the two-dimensional tip position displacement to the location of a center blob and changes in the relative positions of lateral blobs are interpreted as a contact with the environment. The compliant link is actuated by two electric motors and a filtering approach is employed to the input signals to reduce an excitation of the lowest natural frequency of the link. The internal pressure is increased and thereby demonstrated to compensate for a shift in the lowest natural frequency when a payload is attached to the link.

Sensing approaches relying on a camera are promising because the sensor (i.e. the camera) is not required to mechanically interact with the soft material being observed. Therefore, the compliance of the sensor and the soft material are not required to match, simplifying material selection and avoiding stress concentrations at the interface between the sensor and the soft material, which otherwise can limit the maximum number of load cycles of the soft robotic system. Furthermore, cameras provide a high resolution, they are not affected by environmental influences such as temperature or electromagnetic noise and their low cost enables the deployment of multiple sensors in soft robotic systems. Finally, the images recorded by the internal cameras allow us to detect aging phenomena or damage to the observed structure. The challenge with cameras is to integrate the rigid sensor into the soft structure. The size of the camera itself can impose design constraints and the material deformation of interest is required to lie in the visible area of the camera, which can further complicate integration. Additionally, the high-dimensional sensor output needs to be processed in real-time, which requires computational capacity (Kappassov et al., 2015).

Contribution

While camera-based sensing approaches have been demonstrated for a soft plush robot, soft fingers and a compliant link, we demonstrate a vision-based sensing approach for a fabric-based bellow actuator used in a soft robotic arm. Our approach relies on the integration of a small camera with a footprint of 7 mm × 7 mm and a distinctive white pattern which is applied to the interior surface of the actuator. Multiple LEDs are integrated to control the illumination.

A convolutional neural network architecture is trained and used to map the raw camera images to the rotational degrees of freedom of the robotic arm. We show that a lightweight network architecture, which can be deployed on a regular laptop computer

FIGURE 2 | A schematic of the soft robotic arm is shown in the left hand plot. The system consists of a static link (1) attached to the ground, a movable link (2), the inflatable bellow actuators (3) and a soft joint (4) placed between the actuators and connecting the two links. The orientation parametrization of the spherical robotic arm is shown in the middle plot. The static link is aligned with the inertial z-axis. A positive rotation of the movable link around the inertial x-axis is denoted by α and a positive rotation around the inertial y-axis is denoted by β. The top view of the actuator configuration in the inertial coordinate frame is shown in the right hand plot. The three actuators are arranged symmetrically around the inertial z-axis, where the actuator A is aligned with the inertial x-axis.

without GPU support, can predict the orientation of the robot arm at 30 Hz and achieves a root-mean-square accuracy of about one degree.

While camera-based interaction force control is demonstrated in McInroe et al. (2018) and feed-forward vibration control in Oliveira et al. (2020), no closed-loop position control relying on feedback from cameras has been demonstrated for a soft robotic system. We extend the results presented in Werner et al. (2020) for a single, linear actuator, to the control of a spherical robotic arm using three actuators each including an internal camera.

Outline

The remainder of this paper is organized as follows: **Section 2.1** presents the design of the soft bellow actuator and the integration of the camera and the peripherals required. The machine learning pipeline to retrieve the orientation from the camera images is discussed in **Section 2.2** and the control approach employed in **Section 2.3**. Results showing the real-time prediction capability of the sensing approach are presented in **Section 3**, along with a validation of the sensing approach to provide feedback for closed-loop control experiments. Finally, a conclusion is drawn in **Section 4**.

MATERIAL AND METHODS

The hardware used for realizing the camera-based sensing approach is discussed in the first part of this section. In a second part, a supervised machine learning approach is presented that maps the camera images to the angles describing the orientation of the robotic arm. The section is concluded with a brief description of the controller employed on the robotic arm.

Hardware

We start with an overview of the spherical robotic arm used for evaluation of the camera-based sensing method. The following sections outline principle design considerations regarding the vision-based actuator, the manufacturing of the soft actuators and the required camera peripherals employed. The section is concluded by a discussion of the integration of the camera into the actuator.

Spherical Robotic Arm

The spherical robotic arm is closely related to the system presented in Zughaibi et al. (2020) and consists of two inflatable links and three fabric-based bellow actuators that are arranged symmetrically around a soft silicone joint connecting the two links. The robotic arm has two rotational degrees of freedom, which are described by the extrinsic Euler angles α and β (see **Figure 2**). The orientation of the movable link can be adjusted by inflating the actuators A, B and C to different pressures p_A, p_B and p_C to control the elongation of each actuator. Therefore, the three actuator pressures form the control inputs to the system. Note that each bellow actuator can not only expand longitudinally (when pressurized), but allows also for lateral deformation when the other actuators expand.

Since we have three control inputs for only two rotational degrees of freedom, it is also possible to control the stiffness of the joint. An intuitive way to understand this property is the fact that a certain orientation of the movable link can be attained by multiple pressure combinations, where the sole difference lies in the resulting joint stiffness. In this work, the capability of adjusting the joint stiffness is not explored and the reader is referred to Hofer and D'Andrea (2020) and Zughaibi et al. (2020) for more details.
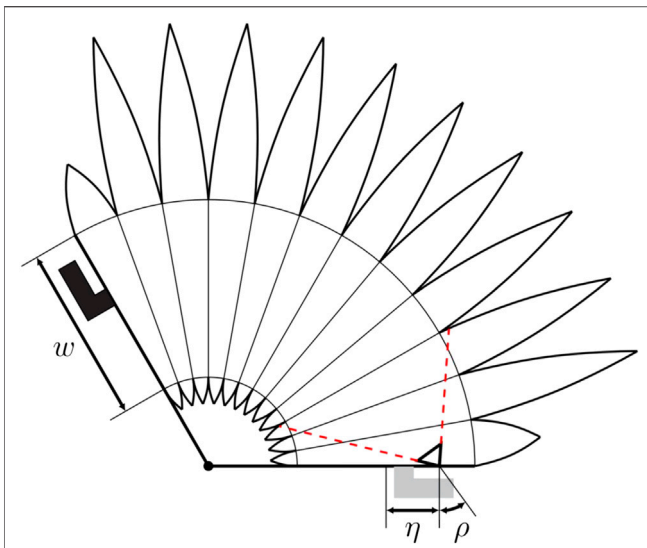
FIGURE 3 | The figure shows a simplified sketch of the cross section of an actuator with the visible area of the camera indicated in red (dashed). Angle connectors are attached to both the top (shown in black on the left of the figure) and bottom of the actuator (shown in gray in the bottom of the figure). The bottom connector is used to pressurize the actuator and the top connector to align it with the movable link. The inner opening which connects neighboring cushions has a width denoted by w and plays a crucial role in the resulting visible area of the camera. If the opening is sufficiently wide, the majority of the cushions are within the visible area of the camera. The area of the actuator deformation covered by the camera is increased if the camera is placed with an offset η with respect to the center of the inner opening and tilted by an angle ρ with respect to the normal direction.

Vision-Based Actuator

Principle design considerations for the vision-based actuator are discussed in this section. The fabric-based actuator consists of individual cushions, which are combined at seams around an inner opening, and forming a bellow-type actuator. A simplified sketch of the actuator is shown in **Figure 3**. The actuator combines soft actuation when pressurized and sensing of the actuator's angular elongation and lateral deformation through an integrated camera. The combined actuation and sensing system needs to address several requirements for successful deployment. These requirements are discussed below.

The camera field of view should cover a large range of the actuator expansion to provide sensory feedback over a large range of the movable link. The sensitivity of the sensing approach is maximized if the actuator expansion and deformation cause large variations in the camera images observed. Increasing the width of the inner opening clearly improves the visible area of the camera. Placing the camera with an offset, η, toward the outer edge of the actuator and tilting the camera by an angle, ρ, toward the center of the actuator increases the visible area of the actuator deformation covered by the camera (compare **Figure 3**).

The angular expansion of the entire actuator should be maximized such that the angular range of the movable link is maximized. Therefore, the angular expansion of a single cushion should be maximized by either increasing the radial width of the

actuator, which is done for all cushions except the top and bottom cushions, or reducing the width of the inner opening which violates the previously discussed design requirement. The angular expansion of the actuator can further be improved by increasing the number of cushions employed, which however also leads to a longer production time.

Finally, the actuator needs to be compatible with the links of the robotic arm. Therefore, the ratio of linear and angular expansion of the actuator need to approximately fit the robotic arm. The ratio of angular and linear expansion mainly depend on the ratio between the inner opening width to the radial width of the actuators. Additionally, the location of the inner opening also affects the ratio between angular and linear expansion, where a central positioning yields a linear actuator and a placement of the opening off-center primarily leads to an angular expansion of the actuator.

The final actuator geometry addressing all the design requirements mentioned is detailed in the supplementary files provided.
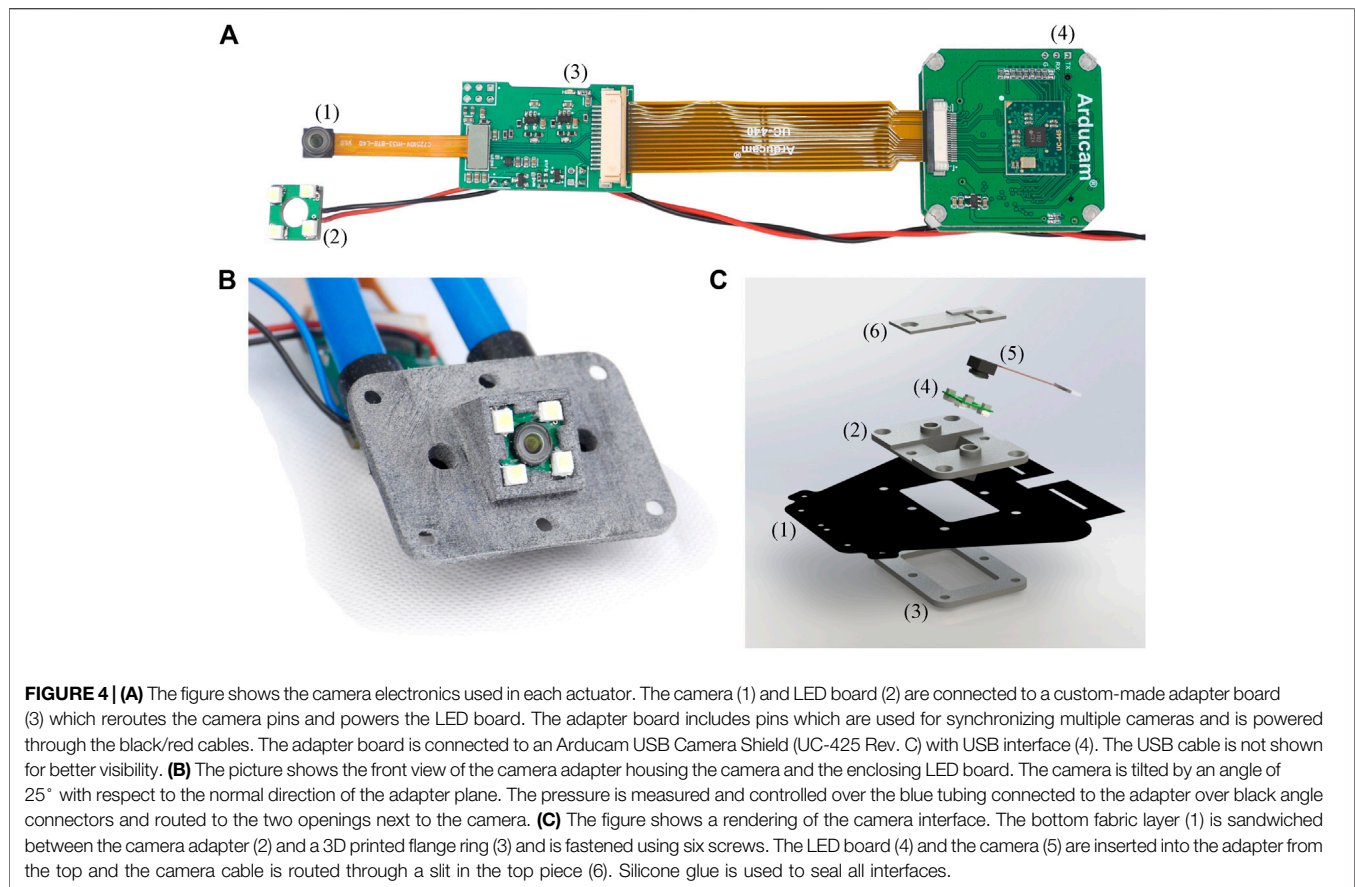
Manufacturing of the Soft Actuator

After defining the requirements of the vision-based actuator in the previous section, the manufacture of the inflatable bellow actuators is discussed here. The manufacture of the rotary actuator is similar to the design presented in Werner et al. (2020) for a linear actuator. The fabrication method as presented in Yang and Asbeck (2018) is applied. The actuators are made from fabric sheets consisting of a sandwich structure. A layer of thermoplastic polyurethane (TPU) film (HM65-PA, 0.1 mm by perfectex) is used inbetween two layers of poplin fabric (polyester cotton blend 65/35 by extremtextil) fused in a heat press (TS7 swingaway heat press by Secabo). The resulting fabric material is inextensible, airtight and sturdier than the single layers of poplin fabric.

The actuator is composed of twelve single cushions, with each cushion being constructed from two pieces of fabric. The pieces have a cutout in the middle (except for the top and bottom parts) where the individual cushions are connected to form a bellow actuator. As mentioned in the previous section, placing the cutout off-center results in a rotary expansion type. Additional TPU ring-shaped seam pieces are prepared to combine the fabric pieces. The actuator is built by stacking the fabric and TPU pieces and fusing them sequentially in a bottom-up process. All fabric and TPU pieces are prepared with a laser cutter and a detailed description of the fabrication procedure can be found in Yang and Asbeck (2018) (Layered Manufacturing-Type I).

Before the fabric layers are fused together, a white pattern is applied to the layers facing the camera to provide visual features with a high contrast to the black fabric. The pattern is cut from adhesive stencil film (S380 by ASLAN) and applied with textile spray paint (319921 textile spray paint by DupliColor) in consecutively applied thin layers. The pattern includes dots with a diameter of 2 mm on the top layer and rings around the opening of the middle cushions.

Since the fabric material is opaque, a light source is required to illuminate the interior of the actuator and make the white pattern visible to the camera. The camera electronics, including its

FIGURE 4 | (A) The figure shows the camera electronics used in each actuator. The camera (1) and LED board (2) are connected to a custom-made adapter board (3) which reroutes the camera pins and powers the LED board. The adapter board includes pins which are used for synchronizing multiple cameras and is powered through the black/red cables. The adapter board is connected to an Arducam USB Camera Shield (UC-425 Rev. C) with USB interface (4). The USB cable is not shown for better visibility. **(B)** The picture shows the front view of the camera adapter housing the camera and the enclosing LED board. The camera is tilted by an angle of 25° with respect to the normal direction of the adapter plane. The pressure is measured and controlled over the blue tubing connected to the adapter over black angle connectors and routed to the two openings next to the camera. **(C)** The figure shows a rendering of the camera interface. The bottom fabric layer (1) is sandwiched between the camera adapter (2) and a 3D printed flange ring (3) and is fastened using six screws. The LED board (4) and the camera (5) are inserted into the adapter from the top and the camera cable is routed through a slit in the top piece (6). Silicone glue is used to seal all interfaces.

peripherals, are discussed the next section. The files of all fabric parts are provided in the supplementary files of this publication.

Camera Electronics Setup

The camera electronics setup used in each actuator is depicted in **Figure 4A**. The camera (OV7251 by OmniVision) houses a CMOS VGA sensor with a maximum resolution of 640 × 480 pixels and a corresponding frame rate of 100 frames per second. The camera has a footprint of 7 mm x 7 mm and is significantly smaller than the camera employed in Werner et al. (2020), therefore simplifying integration. The camera is connected to an adapter board which reroutes the pins to an Arducam USB Camera Shield (UC-425 Rev. C). A custom LED board is powered and controlled via the adapter board. The light intensity is adjusted by setting the duty cycle of a pulse width-modulated signal. A constant duty cycle of 0.22 is used throughout this work. The camera employed allows synchronization of the cameras over pins routed to the adapter board and connected between the three cameras. The camera of actuator A takes the role of the leading camera that triggers the picture and the other two cameras take the role of followers. The Arducam USB Camera Shield SDK library is used for software integration. The schematics of the LED board and the adapter board are provided in the supplementary files to this publication.

Camera Integration

In this section we discuss the integration of the camera and its peripherals into the soft actuator. Only the camera and the LED board are mounted inside the actuator. The other peripherals shown in **Figure 4A** are placed outside the actuator. The camera and the LED board should point in the same direction irrespective of the actuator expansion. Therefore, both the camera and the LED board are glued (Silicone multi-purpose sealant 732 by Dow Corning) to the 3D printed adapter piece (made from PA12). The camera offset η and angle ρ, as discussed in *Camera-Based Sensing*, can be addressed in the design of the 3D printed adapter. The adapter is fixed to an opening in the bottom layer of the actuator over a flange-like structure (see **Figure 4C**). The CAD files of the camera adapter are provided in the supplementary files of this publication.

The resulting actuator is shown in **Figure 5**, when inflated to different elongations with the image from the internal camera alongside. Although the camera adapter is made from rigid material, the adapter is enclosed by the actuator and the static link parts shielding the rigid part toward the surroundings. The camera electronics could be routed internally at the cost of a higher design complexity. Equipping the actuators with a camera and the required peripherals for the vision-based sensing approach adds additional weight to the lightweight soft robotic system. A single camera including all its peripherals (boards,
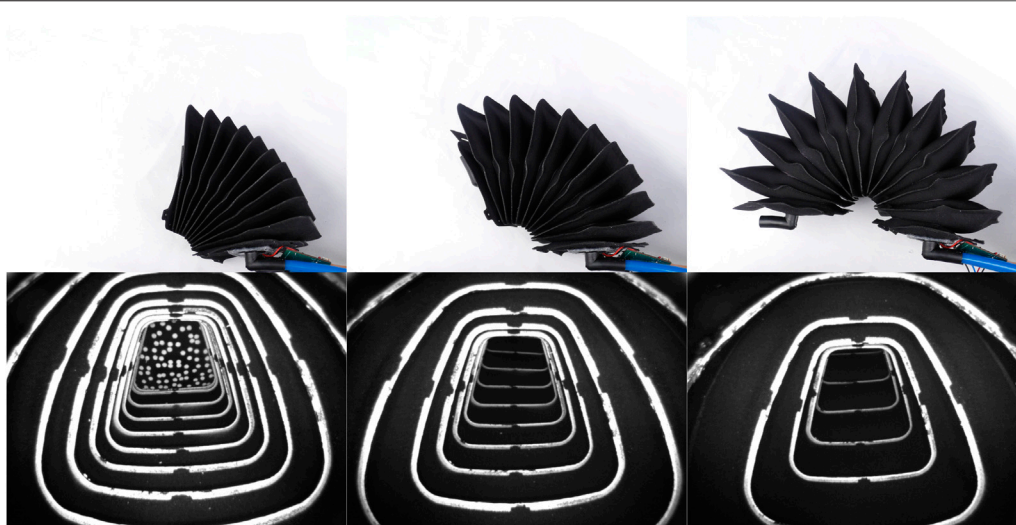
FIGURE 5 | The picture shows a single actuator inflated to different expansions and the corresponding image from the internal camera. The number of cushion rings visible to the internal camera decreases as the actuator expands. The light intensity is set such that the white pattern is visible over the full range of the actuator expansion.

adapter, etc.) adds 14 g for a single actuator, whereas the actuator without camera and peripherals weights 77 g. Hence, the vision-based sensing approach leads to a relative increase in weight of 18% for the actuators.

Camera-Based Sensing

The method to predict the orientation of the robotic arm from the internal cameras is discussed in this section. The identification of the mapping from camera images to the angles describing the orientation of the robotic arm is posed as a supervised learning problem with ground truth data available by means of a motion capture system. The approach presented in Werner et al. (2020) relies on a feature engineering step, followed by a support vector regression to predict the actuator deformation. The key advantage is the limited training complexity of the support vector regression, which came at the cost of the required feature engineering step. In this work, a lightweight convolutional network is used to predict the orientation of the robot arm from the camera images. The end-to-end learning approach bypasses any feature engineering step, but requires more training data, resulting in longer training times. The network architecture is outlined in the first part of this section. The data collection and model learning are discussed thereafter.

Data Collection

Ground truth data is provided by an infrared motion capture system running at 200 Hz and providing sub-millimeter accuracy of the x-y-z-position of the tip of the movable link. The extrinsic Euler angles α and β are calculated by applying the following formulas,

$$\alpha = \arcsin\left(-\left(y - y_0\right)/R\right) \qquad (1)$$
$$\beta = \arcsin\left(\left(x - x_0\right)/\left(R \cdot \cos\left(\alpha\right)\right)\right). \qquad (2)$$

Thereby, (x_0, y_0) denotes the coordinates of the pivot point and R the radius of the movable link, which are both determined in a calibration process.

The data collection includes storing the images from the three internal cameras of the current link orientation and the corresponding ground truth labels α and β. In order to cover the α-β plane uniformly, a position controller as discussed in *Control* is used to track a regular grid of α and β setpoints in the range of $[-30, 30^\circ]$. Although the camera images and the corresponding ground truth labels are collected at a rate of 10 Hz, the sensing approach is deployed at a rate of 30 Hz for the validation and control experiments described in *Results*. Note that the maximum angular range of the robotic arm is slightly larger than the quadratic area covered during data collection. However, the same procedure could be applied to the full angular range but would require more time for data collection and training of the network.

The data is preprocessed by first sub-sampling each image using linear interpolation to a resolution of 120×160 pixels. All pixel values are converted to floating point format and normalized to the interval $[-1, 1]$. A training data set of approximately 54,000 images and labels is collected (corresponding to 90 min of data) and a validation data set of approximately 15,000 images and labels is recorded (corresponding to 25 min of data).

Network Architecture

The network architecture used in this work is related to LeNet as documented in Lecun et al. (1998). The main building block is a convolutional layer followed by a nonlinear activation (i.e., ReLU) and a max pooling step. This building block is repeated three times, before the output of the last pooling step is fed into two fully connected layers predicting the two dimensional output. The resulting network is required to provide inference in real-time on
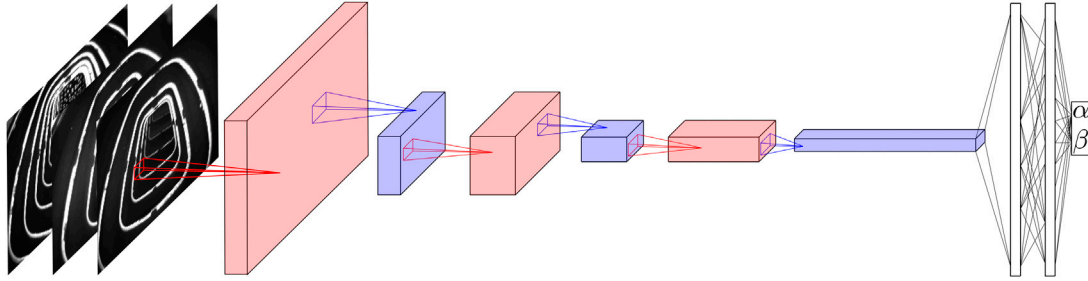
FIGURE 6 | The figure depicts the architecture of the neural network employed. The input consists of the three camera images sub-sampled to a resolution of 120 × 160 pixels. First a convolutional layer (in red) is applied, with four output channels followed by a nonlinear activation (not shown) and a max pooling step (in blue) reducing the size of the image to 24 × 32. The procedure is repeated twice more, while the number of channels is increased to eight and 16, respectively. The pooling steps reduce the size to 6 × 8 after the second and to 3 × 4 after the third max pooling step. Finally, two fully connected layers are applied which output the angles α and β. The pooling layers reduce the number of parameters and consequently the computational complexity significantly, while retaining the most important features.
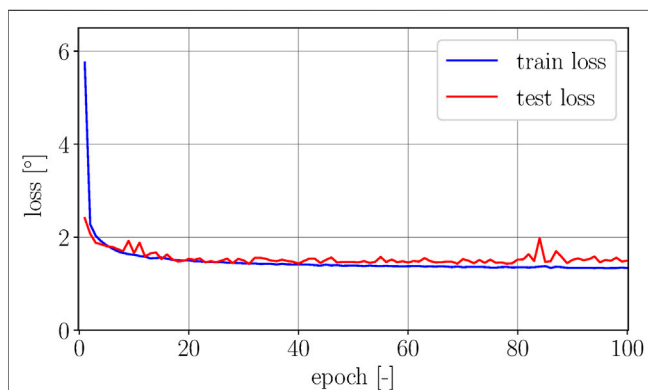
FIGURE 7 | The figure shows the square root of the training and test loss over 100 epochs. The gap between train and test loss is relatively small, showing slight indications of overfitting for higher epochs. The lowest test error is achieved in epoch 31 with a corresponding error of 1.43°.

a standard laptop computer without GPU support. Therefore, the maximum size of the network is limited. The following network exhibits a good trade-off between prediction accuracy and computational complexity.

All convolutional layers have a kernel of size three, a stride of one and a padding of one. The max pooling kernel sizes (and the corresponding strides) are chosen as (5, 4, 2) in the first, second and third layers, respectively. No padding is used for the pooling step. The output of the last pooling layer is fed into two fully connected layers with 40 neurons each and ReLU activation functions. The network architecture is depicted in **Figure 6**. The network has a total number of 9,378 parameters, which is about half of LeNet-4 with roughly 17,000 parameters.

Model Learning

The PyTorch framework (Paszke et al., 2019) is used for model training. The AdamW optimizer is used to minimize the mean squared error. The model is trained for 100 epochs with a learning rate of 1e-3 and a batch size of 128. The data is shuffled before training and a GPU (Nvidia Titan X Pascal) is used for training

the network (not used during the deployment of the network). The evolution of the train and test loss over all epochs is shown in **Figure 7**.

Variations of the parameters defining the model architecture, namely the number of channels in each convolution, the convolution and pooling kernel sizes and the number of linear units were also considered, with no significant improvement in prediction accuracy for a network of similar size.

Control

The control approach for the spherical robotic arm is introduced in this section. The focus of this work lies on the sensing method and therefore only a simple control strategy is employed. A more elaborate, i.e., model-based approach is documented in Zughaibi et al. (2020).

The control approach relies on a cascaded control architecture, similar to the one presented in Hofer and D'Andrea (2020), with an outer control loop for the slower motion dynamics of the robotic arm and three independent, inner control loops for the faster pressure dynamics. Based on the sensory feedback of α and β, the position controller computes the pressure setpoints, which are the inputs to the inner control loops. The sensor feedback is either provided by the motion capture system or by the vision-based sensing approach presented in *Camera-Based Sensing*. The control inputs required to track the setpoints α_{SP} and β_{SP} are computed based on two decoupled proportional-integral controllers,

$$u_\alpha = K_{\mathrm{P}}\left(\alpha_{\mathrm{SP}} - \alpha\right) + K_{\mathrm{I}}\int \left(\alpha_{\mathrm{SP}} - \alpha\right)dt, \qquad (3)$$

$$u_\beta = K_{\mathrm{P}}\left(\beta_{\mathrm{SP}} - \beta\right) + K_{\mathrm{I}}\int \left(\beta_{\mathrm{SP}} - \beta\right)dt, \qquad (4)$$

with K_{P} and K_{I} denoting the proportional and integral gains. The two control inputs are mapped to the three actuator pressure setpoints by applying the following two-part procedure. First, the control inputs u_α, u_β are aligned with the actuators by applying the following linear transformation, which is based on the actuator geometry (compare **Figure 2**),

$$p_{\text{AB,SP}} = -\frac{1}{\sqrt{3}}u_\alpha - u_\beta \tag{5}$$

$$p_{\text{BC,SP}} = \frac{2}{\sqrt{3}}u_\alpha, \tag{6}$$

where $p_{\text{AB,SP}}$ corresponds to the pressure setpoint difference between actuators A and B and $p_{\text{BC,SP}}$ to the difference between actuators B and C, respectively. Secondly, the relative pressure setpoint differences between two actuators are allocated to the absolute pressure setpoints by the following set of equations originating from Zughaibi et al. (2020),

$$p_{\text{A,SP}} = \max\left\{\overline{p}, \overline{p} + p_{\text{AB,SP}}, \overline{p} + p_{\text{AB,SP}} + p_{\text{BC,SP}}\right\} \tag{7}$$

$$p_{\text{B,SP}} = \max\left\{\overline{p}, \overline{p} + p_{\text{BC,SP}}, \overline{p} - p_{\text{AB,SP}}\right\} \tag{8}$$

$$p_{\text{C,SP}} = \max\left\{\overline{p}, \overline{p} - p_{\text{BC,SP}}, \overline{p} - p_{\text{AB,SP}} - p_{\text{BC,SP}}\right\}. \tag{9}$$

Thereby, $\overline{p}$ is defined as a lower pressure level of all three actuators,

$$\overline{p} = \min\left\{p_{\text{A,SP}}, p_{\text{B,SP}}, p_{\text{C,SP}}\right\}. \tag{10}$$

The validity of the second step can be verified by computing the pressure differences between actuator A and B and similarly for B and C and performing the required case distinctions. The lower pressure level $\overline{p}$ can be interpreted as a mean to adjust the unidirectional stiffness of the robotic arm (see (Hofer and D'Andrea, 2020) for more details).

The pressure setpoints for actuators A, B and C are tracked by three independent proportional-integral controllers,

$$u_\text{A} = \tilde{K}_\text{P}\left(p_{\text{A,SP}} - p_\text{A}\right) + \tilde{K}_\text{I}\int\left(p_{\text{A,SP}} - p_\text{A}\right)dt \tag{11}$$

$$u_\text{B} = \tilde{K}_\text{P}\left(p_{\text{B,SP}} - p_\text{B}\right) + \tilde{K}_\text{I}\int\left(p_{\text{B,SP}} - p_\text{B}\right)dt \tag{12}$$

$$u_\text{C} = \tilde{K}_\text{P}\left(p_{\text{C,SP}} - p_\text{C}\right) + \tilde{K}_\text{I}\int\left(p_{\text{C,SP}} - p_\text{C}\right)dt, \tag{13}$$

with $\tilde{K}_\text{P}$ and $\tilde{K}_\text{I}$ denoting the proportional and integral gains of the pressure controllers. The three controllers are executed on an embedded hardware at a higher rate than the position controller.

RESULTS

The results of the experimental evaluation of the method proposed are presented in this section. The results of the real-time prediction of the two angles are presented in the first part. The closed-loop experiments relying on the feedback from the camera-based sensing approach are presented in the second part.

The network is deployed on a standard laptop computer (Intel Core i7 CPU, 2.8 GHz). The ONNX Runtime framework[1] is used to reduce inference time of the neural network and provide a prediction of α and β at a rate of 30 Hz. The frame rate of the cameras is set accordingly to 30 Hz during the deployment of the sensing approach for both experiments reported in *Real-Time*

---

[1]https://onnxruntime.ai

Prediction and Vision-Based Control. The multithreaded software application includes, besides model inference, a graphical user interface and a position controller running at 50 Hz, where the previous prediction of the angles is used for intermediate executions of the controller. The pressure controllers are executed at 1,000 Hz on an embedded platform (STM32 Nucleo-144 development board with STM32F413ZH MCU from STMicroelectronics). The gains of the position controller are set to $K_\text{P} = 0.16$ bar $\cdot$ rad^{-1} and $K_\text{I} = 0.19$ bar $\cdot$ rad$^{-1} \cdot$ s^{-1} and the gains of the pressure controllers are set to $\tilde{K}_\text{P} = 0.05$ bar^{-1} and $\tilde{K}_\text{I} = 0.03$ bar$^{-1} \cdot$ s^{-1}. All actuator pressures are measured by pressure sensors (8230 from Burkert) and the outputs of the pressure controllers are applied to proportional valves (MPYE-5–1/8-HF-010-B from Festo) adjusting the air pressures. The lower pressure level is set to $\overline{p} = 1.02$ bar. Communication between the main application and the embedded platform is realized by means of serial communication. The interested reader is referred to the video attachment to gain an impression of the experiments conducted (https://youtu.be/CldCKhukqqQ).

Real-Time Prediction

The robotic arm is commanded to track a series of steps and ramps in the α and β-directions, relying on feedback from the motion capture system. The range of frequencies considered corresponds to the range of frequencies in which the system typically operates. The trajectory is repeated five times to investigate repeatability of the sensing approach. The camera-based sensing approach is executed in real-time and the results of two repetitions of the trajectory are depicted in **Figure 8** showing the network prediction along with ground truth.

In order to evaluate the consistency of the angle predictions over the five iterations, the root-mean-square error between prediction and ground truth is calculated for both angles and for each of the five iterations separately. The mean of the individual root-mean-square errors in α-direction is $\mu_\alpha = 1.03°$ and the standard deviation is $\sigma_\alpha = 0.06°$. Accordingly, the mean in β-direction is $\mu_\beta = 1.39°$ and the standard deviation is $\sigma_\beta = 0.10°$. The relatively low standard deviations of both prediction errors emphasize the repetitive nature of the deviation, indicating that there is a systematic trend which the network employed fails to capture. However, the similarity between the two realizations implies that a hardware limitation is not the cause of the deviations in the predictions. In fact, these repeatable behaviors are ascribed to the limited representation power of the compact network architecture mentioned in *Model Learning*. In this regard, variations of the training parameters result in a trade-off in accuracy across different regions of the α-β plane. As a result, with the current architecture, improvements in one region are only possible at the cost of an accuracy degradation in another region.

Finally, a worst case estimation for the delay of the sensing approach is presented. Considering a frame rate of 30 Hz and a measured average inference time of the neural network of approximately 2 ms, the resulting total worst case delay is 35 ms. Thereby, inter thread communication delay and communication delays between the laptop computer and the
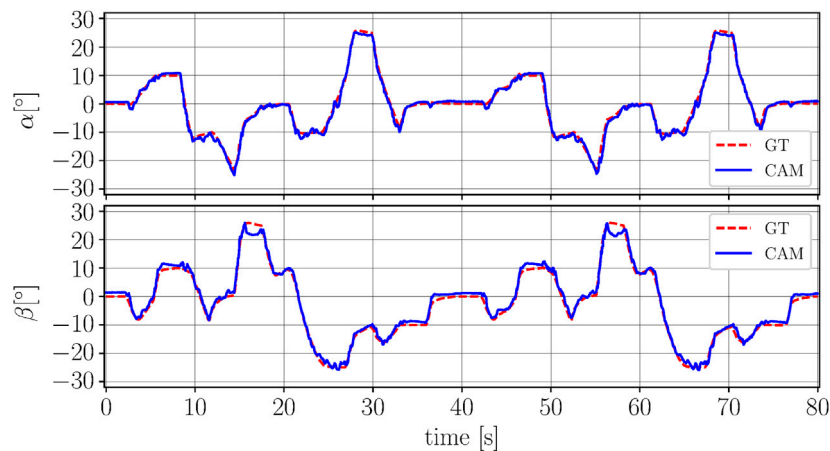
FIGURE 8 | The figure shows the rotational degrees of freedom α **(top plot)** and β **(bottom plot)** over time. The camera-based prediction (CAM) is shown in blue (solid line) and the ground truth (GT) from the motion capture is depicted in red (dashed line). No filtering is applied to the output from the network. Slight deviations can be seen for large values of β. The deviations between the two repetitions of the trajectory are comparable.
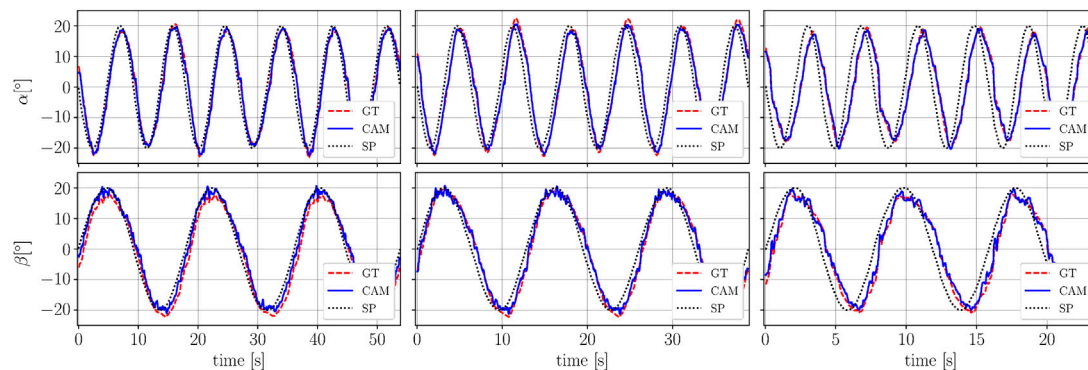

FIGURE 9 | The figure shows the results of the tracking experiments with the camera-based sensing approach used as sensory feedback. The angle α is shown in the top plots and β is shown in the bottom plots. The tracking experiment is performed for three different frequencies. The left hand plot shows the lowest frequency (0.12 Hz in α and 0.06 Hz in β), the middle plot corresponds to an intermediate frequency (0.16 Hz in α and 0.08 Hz in β) and the right hand plot shows the highest frequency (0.25 Hz in α and 0.12 Hz in β). The camera-based predictions are shown in blue (solid line) and the ground truth from the motion capture in red (dashed line). The commanded setpoint is shown in black (dotted line). Slight oscillations can be observed for large absolute values of β.

embedded hardware are not considered since they are not inherent to the sensing approach.

Vision-Based Control

The results of the closed-loop control experiment relying on feedback from the vision-based sensing approach are discussed in this section. The robotic arm is commanded to track sinusoidal setpoint trajectories in α and β-directions. The tracking experiment is repeated for three different frequencies. The results are depicted in **Figure 9** showing the network predictions (used as feedback) along with ground truth and the setpoints. It can be observed that the delay between the setpoint and the measured angle is increasing for higher frequencies for both degrees of freedom. This is a consequence of the robot's dynamics exhibiting a higher phase lag for higher frequencies.

The accuracy of the vision-based sensing approach is evaluated by comparing it to ground truth. The root-mean-square errors are calculated for both angles and then averaged. The combined errors are summarized in column three of **Table 1** (RMSE GT-VBS). The resulting errors for the low, intermediate and high frequency correspond to the left, middle and right hand plots of **Figure 9**, respectively.

The accuracy of the sensing approach and the setpoint tracking performance achievable, are compared to the case when ground truth sensory feedback is available. Therefore, the same three tracking experiments as shown in **Figure 9** are repeated with ground truth from the motion capture system used as feedback. The results are summarized in **Table 1**. The root-mean-square error between different signals and averaged between the two angles is calculated when using either ground truth or the vision-based sensing approach as sensory feedback.

TABLE 1 | Overview of the different root-mean-square errors (averaged between the two angles) when using ground truth (GT) as feedback (column two) or when using the vision-based sensing approach (VBS) as feedback (columns three to five). The evaluation is done for trajectories of different frequencies shown in rows three to five. The setpoint is abbreviated with SP. The signals used to calculate the root-mean-square errors (RMSE) are indicated by the two abbreviations following the RMSE.

Trajectory frequency	GT as feedback	VBS as feedback		
	RMSE GT-SP	RMSE GT-VBS	RMSE VBS-SP	RMSE GT-SP
Low	2.98°	1.56°	3.04°	3.68°
Intermediate	4.41°	1.15°	4.48°	4.87°
High	7.18°	1.26°	7.06°	7.30°

The second column shows the root-mean-square error between ground truth and the setpoint (RMSE GT-SP) when relying on ground truth as sensory feedback. These tracking errors show the control performance achievable with the soft robotic arm when ground truth is available as feedback. The experimental results shown in columns three to five rely on the vision-based sensing approach as feedback. Column three reports the accuracy of the vision-based sensing approach when compared to ground truth (RMSE GT-VBS). The tracking error between the vision-based sensing approach and the setpoint, as used in the controller, is reported in column four (RMSE VBS-SP). The reported errors are similar for all three frequencies when compared to the results shown in column two. Finally, the true tracking errors between ground truth and the setpoint, when relying on feedback from the vision-based sensing approach, are shown in column five (RMSE GT-SP). The results indicate that the actual tracking errors achievable with the vision-based sensing approach, are slightly higher compared to the case when relying on ground truth sensory feedback. However, the additional deviation induced by the vision-based sensing approach is relatively small, emphasizing the suitability of the sensing approach proposed for closed-loop control of the robotic arm.

DISCUSSION

This paper presents a vision-based sensing approach for a soft robotic arm made from fabric. The camera integration into inflatable bellow actuators has been discussed, with three actuators being used to control a spherical robotic arm. An end-to-end deep learning approach relying on a shallow convolutional network is employed and trained with ground truth data from a motion capture system to map the camera images to the two rotational degrees of freedom of the robotic arm. Note that formulating a model-based sensing approach would require us to explicitly account for material behaviors, e.g., deformation, that are challenging to model accurately. In addition, the use of convolutional filters results in an efficient processing of the information at all pixels of the images, contributing to the high accuracy obtained. The resulting method is computationally lightweight and can be deployed in real-time on a standard laptop computer providing predictions of the two angles at a rate of 30 Hz with an accuracy of about one degree. The reliability of the vision-based sensing approach has been demonstrated by closed-loop

control experiments relying on the sensory feedback from the camera-based sensing approach.

The proposed sensing approach, relying on a relatively small network architecture, can be deployed on a standard laptop computer without GPU support at 30 Hz. Note that the neural network architecture enables the simultaneous prediction of the two angles, efficiently exploiting the interconnections across the three synchronous images and the two outputs with a single architecture. This differs from Werner et al. (2020), where a single-output support vector regression was employed to predict the elongation of a linear actuator with a single camera at 40 Hz. A multi-output angle prediction with such a method would require the use of separate single-output regressors, with a consequent increase in the computational cost.

In principle, the internal cameras would provide images at a frame rate of up to 100 Hz. The computational resources provided by the computer employed is currently the primary limitation in terms of model size and update rate. Leveraging specialized computational hardware that is able to process the acquired image stream at full rate, would allow for the full exploitation of the sensory feedback provided by the cameras.

In order to further improve the prediction accuracy, larger network architectures might be required. The repeatability of the current deviations indicates that the physical limitation of the sensing approach is not yet reached and better prediction accuracy is possible at the cost of larger networks and correspondingly higher computational costs for training and inference. Furthermore, we only investigated algebraic mappings from camera images to output angles without any previous state dependency. The use of e.g. recurrent neural network architectures to rely on past predictions to capture time dependent effects in the actuator deformation might be another means to improve the performance of the sensing approach presented.

There are several possible extensions to the approach presented: The sensing pipeline is identified for a fixed lower pressure level $\bar{p}$, which is related to joint stiffness. Therefore, the lower pressure level is likely to affect the actuator deformation and hence the images observed by the internal cameras. An extension of the current work would be to feed $\bar{p}$ as an input to the network and train it for different values of $\bar{p}$. Secondly, the robotic arm is commanded to track positions during the data collection and evaluation experiments. A subject for future work is an investigation of the method presented for deployment during interactive applications with disturbances acting on the movable link. Thirdly, the information stream provided by the cameras is currently only used for sensing the

rotational degrees of freedom. An interesting extension would be to use the camera images to identify aging phenomena of the bellow actuators or to detect damages in the actuators, both changing the actuator deformation observed by the cameras. Finally, the sensing pipeline presented relies on three cameras (one in each actuator) to predict the rotational degrees of freedom. However, a change in both angles, in the α and β-directions causes a variation of each of the camera images. Hence, it is likely that the orientation of the robotic arm would be observable with only one or two cameras. This would simplify the hardware setup, but might come at the cost of reduced prediction accuracy.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

Abidi, H., and Cianchetti, M. (2017). On intrinsic safety of soft robots. *Front. Robot. AI* 4, 5. doi:10.3389/frobt.2017.00005

Bao, G., Fang, H., Chen, L., Wan, Y., Xu, F., Yang, Q., et al. (2018). Soft robotics: academic insights and perspectives through bibliometric analysis. *Soft Robot.* 5, 229–241. doi:10.1089/soro.2017.0135

Chin, K., Hellebrekers, T., and Majidi, C. (2020). Machine learning for soft robotic sensing and control. *Adv. Intell. Syst.* 2, 1900171. doi:10.1002/aisy.201900171

Felt, W, Remy, C. D., Telleria, M. J., Allen, T. F., Hein, G., Pompa, J. B., et al. (2019). An inductance-based sensing system for bellows-driven continuum joints in soft robots. *Auton. Robot.* 43, 435–448. doi:10.1007/s10514-018-9769-7

Grzesiak, A., Becker, R., and Verl, A. (2011). The bionic handling assistant: a success story of additive manufacturing. *Assemb. Autom.* 31, 329–333. doi:10.1108/01445151111172907

Hofer, M., and D'Andrea, R. (2020). Design, fabrication, modeling and control of a fabric-based spherical robotic arm. *Mechatronics* 68, 102369. doi:10.1016/j.mechatronics.2020.102369

Kappassov, Z., Corrales, J.-A., and Perdereau, V. (2015). Tactile sensing in dexterous robot hands - Review. *Robot. Autonom. Syst.* 74, 195–220. doi:10.1016/j.robot.2015.07.015

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE.* 86, 2278–2324. doi:10.1109/5.726791

Luo, M., Skorina, E. H., Tao, W., Chen, F., Ozel, S., Sun, Y., et al. (2017). Toward modular soft robotics: proprioceptive curvature sensing and sliding-mode control of soft bidirectional bending modules. *Soft Robot.* 4, 117–125. doi:10.1089/soro.2016.0041

McInroe, B. W., Chen, C. L., Goldberg, K. Y., Bajcsy, R., and Fearing, R. S. (2018). Towards a soft fingertip with integrated sensing and actuation," in 2018 IEEE/RSJ international conference on intelligent robots and systems (IROS), Madrid, Spain, October 1–5, 2018, 6437–6444.

Oliveira, J., Ferreira, A., and Reis, J. C. P. (2020). Design and experiments on an inflatable link robot with a built-in vision sensor. *Mechatronics* 65, 102305. doi:10.1016/j.mechatronics.2019.102305

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). "Pytorch: an imperative style, high-performance deep learning library," in

Advances in neural information processing systems (Curran Associates, Inc.), 32, 8026–8037.

Polygerinos, P., Correll, N., Morin, S. A., Mosadegh, B., Onal, C. D., Petersen, K., et al. (2017). Soft robotics: review of fluid-driven intrinsically soft devices; manufacturing, sensing, control, and applications in human-robot interaction. *Adv. Eng. Mater.* 19, 1700016. doi:10.1002/adem.201700016

Rus, D., and Tolley, M. (2015). Design, fabrication and control of soft robots. *Nature* 521, 467–475. doi:10.1038/nature14543

Sareh, S., Noh, Y., Li, M., Ranzani, T., Liu, H., and Althoefer, K. (2015). Macrobend optical sensing for pose measurement in soft robot arms. *Smart Mater. Struct.* 24, 125024. doi:10.1088/0964-1726/24/12/125024

Sferrazza, C., and D'Andrea, R. (2019). Design, motivation and evaluation of a full-resolution optical tactile sensor. *Sensors* 19, 928. doi:10.3390/s19040928

She, Y., Liu, S. Q., Yu, P., and Adelson, E. (2020). "Exoskeleton-covered soft finger with vision-based proprioception and tactile sensing," in 2020 IEEE international conference on robotics and automation (ICRA), Paris, France, May 31, 2020, 10075–10081.

Shimonomura, K. (2019). Tactile image sensors employing camera: a review. *Sensors* 19, 3933. doi:10.3390/s19183933

Shintake, J., Piskarev, E., Jeong, S. H., and Floreano, D. (2018). Ultrastretchable strain sensors using carbon black-filled elastomer composites and comparison of capacitive versus resistive sensors. *Adv. Mater. Technol.* 3, 1700284. doi:10.1002/admt.201700284

Stassi, S., Cauda, V., Canavese, G., and Pirri, C. F. (2014). Flexible tactile sensing based on piezoresistive composites: a review. *Sensors* 14, 5296–5332. doi:10.3390/s140305296

Takaki, K., Taguchi, Y., Nishikawa, S., Niiyama, R., and Kawahara, Y. (2019). Acoustic length sensor for soft extensible pneumatic actuators with a frequency characteristics model. *IEEE Robot. Autom. Lett.* 4, 4292–4297. doi:10.1109/lra.2019.2931273

Thuruthel, T. G., Shih, B., Laschi, C., and Tolley, M. T. (2019). Soft robot perception using embedded soft sensors and recurrent neural networks. *Sci. Robot.* 4, eaav1488. doi:10.1126/scirobotics.aav1488

Truby, R. L., Santina, C. D., and Rus, D. (2020). Distributed proprioception of 3d configuration in soft, sensorized robots via deep learning. *IEEE Robot. Autom. Lett.* 5, 3299–3306. doi:10.1109/lra.2020.2976320

Wang, H., Totaro, M., and Beccai, L. (2018). Toward perceptive soft robots: progress and challenges. *Adv. Sci.* 5, 1800541. doi:10.1002/advs.201800541

Wang, H., Zhang, R., Chen, W., Liang, X., and Pfeifer, R. (2016). Shape detection algorithm for soft manipulator based on fiber Bragg gratings. *IEEE ASME Trans. Mechatron.* 21, 2977–2982. doi:10.1109/TMECH.2016.2606491

Wang, R., Wang, S., Du, S., Xiao, E., Yuan, W., and Feng, C. (2020). Real-time soft body 3d proprioception via deep vision-based sensing. *IEEE Robot. Autom. Lett.* 5, 3382–3389. doi:10.1109/LRA.2020.2975709

Ward-Cherrier, B., Pestell, N., Cramphorn, L., Winstone, B., Giannaccini, M. E., Rossiter, J., et al. (2018). The tactip family: soft optical tactile sensors with 3d-printed biomimetic morphologies. *Soft Robot.* 5, 216–227. doi:10.1089/soro.2017.0052.PMID:29297773

Werner, P., Hofer, M., Sferrazza, C., and D'Andrea, R. (2020). "Vision-based proprioceptive sensing: tip position estimation for a soft inflatable bellow actuator," in 2020 IEEE/RSJ international conference on intelligent robots and systems, Las Vegas, United States, October 25–29, 2020 (IROS), 8889–8896.

Xiang, C., Guo, J., and Rossiter, J. (2019). Soft-smart robotic end effectors with sensing, actuation, and gripping capabilities. *Smart Mater. Struct.* 28, 055034. doi:10.1088/1361-665x/ab1176

Yang, H. D., Greczek, B. T., and Asbeck, A. T. (2018). Modeling and analysis of a high-displacement pneumatic artificial muscle with integrated sensing. *Front. Robot. AI.* 5, 136. doi:10.3389/frobt.2018.00136

Yang, H. D., and Asbeck, A. T. (2018). "A new manufacturing process for soft robots and soft/rigid hybrid robots," in 2018 IEEE/RSJ international conference on intelligent robots and systems (IROS), Madrid, Spain, October 1–5, 2018, 8039–8046. doi:10.1109/IROS.2018.8593688

Yuan, W., Dong, S., and Adelson, E. H. (2017). Gelsight: high-resolution robot tactile sensors for estimating geometry and force. *Sensors* 17, 2762. doi:10.3390/s17122762

Zhao, H., O'Brien, K., Li, S., and Shepherd, R. F. (2016a). Optoelectronically innervated soft prosthetic hand via stretchable optical waveguides. *Sci. Robot.* 1, eaai7529. doi:10.1126/scirobotics.aai7529

Zhao, H., Jalving, J., Huang, R., Knepper, R., Ruina, A., and Shepherd, R. (2016b). A helping hand: soft orthosis with integrated optical strain sensors and emg control. *IEEE Robot. Autom. Mag.* 23, 55–64. doi:10.1109/MRA.2016.2582216

Zughaibi, J., Hofer, M., and D'Andrea, R. (2020). A fast and reliable pick-and-place application with a spherical soft robotic arm. arXiv preprint. Available at: https://arxiv.org/abs/2011.04624 (Accessed November 9, 2020).