2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Predicting Interaction Shape of Soft Continuum Robots using Deep Visual Models

Yunqi Huang[1], Abdulaziz Y. Alkayas[1,3], Jialei Shi[2,5], Federico Renda[3,4],
Helge Wurdemann[2] and Thomas George Thuruthel[1]

***Abstract*— Soft continuum robots, characterized by their inherent compliance and dexterity, are increasingly pivotal in applications requiring delicate interactions with the environment such as the medical field. Despite their advantages, challenges persist in accurately modeling and controlling their shape during interactions with surrounding objects. This is because of the difficulty in modeling the large degrees of freedom in soft-bodied objects that become more active during interactions. In this study, we present a deep visual model to predict the interaction shapes of a soft continuum robot in contact with surrounding objects. By formulating this task as a forward-statics problem, the model uses the initial state images containing the object configuration and future actuation values to predict interactive state images of the robot under this actuation condition. We developed and tested the model in both simulated and physical environments, explored the model's predictive capabilities using monocular and binocular views, and tested the model's generalization ability on different datasets. Our results show that deep learning methods are a promising tool for solving the complex problem of predicting the shape of a soft continuum robot interacting with the environment, requiring no prior knowledge about the system dynamics and explicit mapping of the environment. This study paves the way for future explorations in robot-environment interaction modeling and the development of more adaptable interaction shape control strategies.**

I. INTRODUCTION

Soft continuum robots have bodies that are made of soft and extensible materials [1], enabling them to form curvilinear shapes and exhibit high degrees of freedoms (DOFs) [2]. Unlike their rigid-bodied counterparts, soft continuum robots with inherent compliance and dexterity are suitable for complex and interactive tasks. In minimally invasive surgery (MIS), these robots navigate through tight spaces within the human body, enabling less invasive procedures and reduced patient trauma [3]. In rehabilitation, they can conform around human limbs to provide customized assistance or resistance [4]. Despite these advantages, the shape modeling of soft continuum robots during interactive tasks remains a significant challenge [5]. In surgical applications, a soft robotic tool must delicately maneuver around sensitive organs and tissues to avoid inadvertent harm. In rehabilitation applications, the robot must adapt its movements to match the needs of patients and understand its shape changes in response to external forces. Thus, accurately predicting and modeling the shape of a soft continuum robot, as well as its interactions with surrounding objects, become not only beneficial but essential [6], [7].

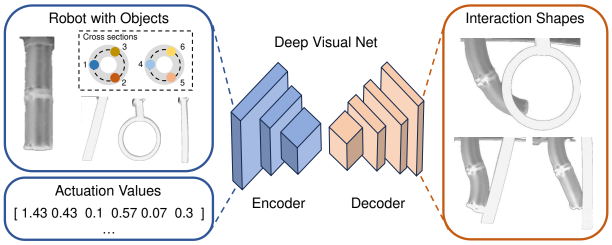


Fig. 1. An overview of the autoencoder-based network for predicting the visual state of a soft robot given the actuation signal and environmental conditions.

The interaction problem in soft continuum robots exemplifies a multi-physics challenge. This problem encompasses the physics of the soft body itself, including its material properties and the actuation dynamics. Additionally, it involves the physics of the external environment, such as fluid dynamics or rigid body interactions. The coupling between these external forces and the soft robot's structure further adds to the complexity of the problem [6]. Mathematically, the interaction problem can be broken down into modeling the internal mechanics and external interactions, and some analytical methods have been developed for these. Finite Element Method (FEM) has been widely used for modeling internal mechanics. It divides the robot into computational meshes, and applies continuum solid mechanics to simulate the robot's behavior [8] [9]. Rod models, such as Cosserat rod models, represent the soft robot as a one-dimensional rod to simplify the complexity of modeling [10] [11]. For external interactions, solving the contact force between the actuated robot structures and the surrounding medium is crucial. This involves applying fluid mechanics to calculate fluid interaction force [12], or employing contact mechanics and tribology for solid-robot interactions [13]. Although these analytical models provide foundational insights into the complex multi-physics problem, they encounter limitations because accurate discretization is computationally intensive, the environment is changing and unknown, and due to inherent uncertainties of the soft robot in material properties and actuator behaviors [14] [15]. The advent of machine learning techniques inspires researchers to implement data-

---

[1]Department of Computer Science and [2]Department of Mechanical Engineering, University College London, London, UK.

[3]Department of Mechanical and Nuclear Engineering, Khalifa University, Abu Dhabi, UAE.

[4]Khalifa University Center for Autonomous Robotic Systems (KUCARS), Khalifa University, Abu Dhabi, UAE.

[5]Mechatronics in Medicine, The Hamlyn Centre for Robotic Surgery, Department of Mechanical Engineering, Imperial College London, UK.

979-8-3503-7769-9/24/$31.00 ©2024 IEEE

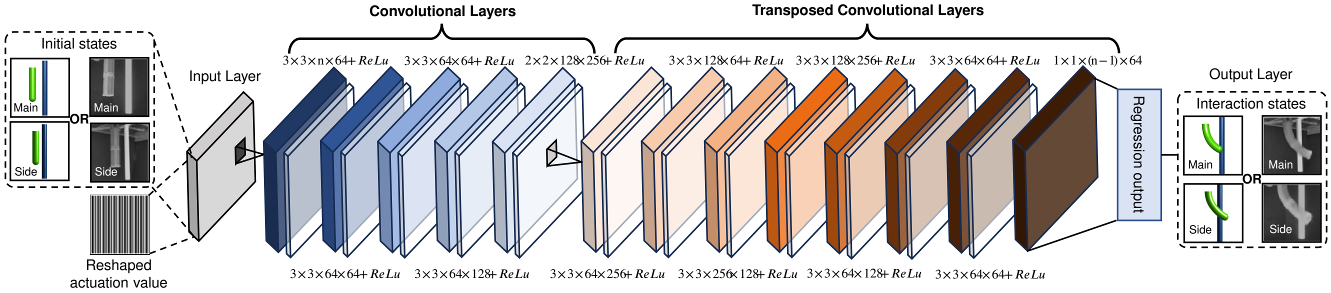


Fig. 2. Network architecture: The input consists of initial state images and actuation values. An encoder extracts input features while downsampling spatial dimensions. A decoder upsamples the feature maps to reconstruct the predicted interactive state images under given actuation conditions. The format of the input and output images remains consistent, allowing for full color or grayscale, monocular or binocular formats. The value of n in the first convolution layer can be 2, 3, 4, or 7, depending on the specific input configuration.

driven methods to understand the dynamics of soft continuum robots, and learn the transformation from actuation space to the deformation space of the robot's body [16] [17] [18]. These methods adaptively learn from data, capturing the complex, non-linear relationships inherent in soft robotics without relying on pre-defined assumptions or analytical models. To effectively implement data-driven approaches for soft robot shape modeling, data collection is critical. Embedding flexible sensors within the robot is a prevalent strategy to get real-time, responsive data regarding the robot's movements [19]. However, most embedded sensors do not incorporate well with soft materials and only capture one DOF movement of robots. An alternative is the use of camera systems, which can offer multi-DOF movement tracking without physical integration [20] [21]. To learn the mapping from actuation space to the deformation space, neural networks have emerged as the predominant method for approximating this mapping [22]. Additionally, alternative regression techniques like locally weighted projection regression [23] and Gaussian process regression [24] are efficient in addressing learning issues in soft continuum robots.

In this paper, we contributed to addressing the challenges of modeling interactions between soft continuum robots and the environments using an autoencoder-based convolutional neural network (CNN), as shown in Fig. 1. We collected data from both simulated and physical environments, developed and trained the CNN model to predict the interactive state images given actuation conditions and initial state images. This data-driven method can learn the complex input-output relationship without explicit physics-based modeling. Through extensive experiments on both simulated and real-world datasets, we demonstrated the effectiveness of our proposed approach in predicting interaction shapes of soft continuum robots under various conditions and highlighted its potential in terms of prediction accuracy and generalization ability.

II. THEORY

A. Forward Problem

In Cosserat rod theory, a soft continuum robot can be represented by a continuous stack of rigid cross-sections along its center curvature [10]. The position of each section can be defined by a function of $X \in [0, L]$, where L is the length of the rod. It can be represented by the homogeneous transformation matrix:

$$g(X) = \begin{pmatrix} \mathbf{R}(X) & \mathbf{t}(X) \\ 0^T & 1 \end{pmatrix} \tag{1}$$

where $\mathbf{t}(X) \in \mathbb{R}^3$ is the vector from the origin of the inertial frame to the section at X, and $\mathbf{R}(X) \in SO(3)$ is the rotation matrix representing the orientation of this section. Then the strains of the rod at X can be denoted as:

$$\xi(X) = [\mathbf{P}^T \ \mathbf{Q}^T]^T \tag{2}$$

where $\mathbf{P}(X) \in \mathbb{R}^3$ and $\mathbf{Q}(X) \in \mathbb{R}^3$ represent the angular and linear strains, respectively. The derivative of g with respect to X can be derived as [25]:

$$g' = g\hat{\xi} \tag{3}$$

where $\hat{\xi} = \begin{pmatrix} \hat{\mathbf{P}} & \mathbf{Q} \\ 0^T & 1 \end{pmatrix}$, connecting the strain field ξ with the robot configuration g and its space derivative. The strain field can be discretized to a finite basis of n vector functions:

$$\xi(X) = \Phi_\xi(X)\mathbf{q} + \xi^*(X) \tag{4}$$

where $\Phi_\xi(X) \in \mathbb{R}^{6\times n}$ is a matrix function and its columns are the basis of the strain field. The vector $\mathbf{q} \in \mathbb{R}^n$ is the generalized coordinates expressed in the chosen basis, and $\xi^*(X)$ is the reference strain, which is computed at the stress-free reference configuration $g^*(X)$ [26].

According to [27], the quasi-static equilibrium can be expressed as :

$$K\mathbf{q} = B(\mathbf{q})\mathbf{u} + \mathbf{F}_e \tag{5}$$

where K is the generalized stiffness matrix calculated from the strain field. $\mathbf{q}$ denotes the generalized coordinates. Matrix $B(\mathbf{q})$ is the generalized actuation matrix that translates actuation inputs, such as tendon forces or pneumatic pressures, into movements or deformations. The vector $\mathbf{u}$ represents these actuation parameters. $\mathbf{F}_e$ signifies the external forces acting on the robot, including those arising from interactions with the environment.

B. Learning of Forward Model

From the quasi-static equilibrium equation (5), the robot's generalized coordinates $\mathbf{q}$, influenced by actuation and external forces, is thus given by:

$$\mathbf{q} = K^{-1}(B(\mathbf{q})\mathbf{u} + \mathbf{F}_e) \tag{6}$$

this generalized coordinates $\mathbf{q}$ describes the robot shape in its configuration space, which is in turn observable through visual images $\mathbf{I}$. These static images offer a visual, higher-dimensional representation of the robot's shape.

The external force $\mathbf{F}_e$ is further modeled as a function of the actuation parameters $\mathbf{u}$ and the initial state images $\mathbf{I}^\eta$, where η is the environment condition:

$$\mathbf{F}_e = f(\mathbf{u}, \mathbf{I}^\eta) \tag{7}$$

So the generalized coordinates $\mathbf{q}$ can be represented as a mapping from the specific actuation value $\mathbf{u}$ and the initial state images $\mathbf{I}^\eta$, and this relationship is captured by the function h. The coordinates $\mathbf{q}$ can then be observed in the visual task-space $\mathbf{I}^t$:

$$\mathbf{q} = h(\mathbf{u}, \mathbf{I}^\eta, \mathbf{F}_e) \quad \mathbf{I}^t = j(\mathbf{q}) \tag{8}$$

To learn this mapping function $j(h())$, a convolutional neural network in Fig. 2 is used. The proposed network comprises an encoder that compresses image data and actuation value data, with a decoder designed to reconstruct the images depicting the interactive states of the robot.

1) Encoder: The encoder takes initial state image data alongside the actuation value data as inputs. The image channels provide visual information on the robot in an unactuated state, the object's geometry, and their relative positions. The initial state conditions can be mono-vision or stereo-vision, and the images can be processed in full color (RGB) or grayscale. The actuation value, which is reshaped to match the dimension of the input images, provides the robot's movement command. The network then applies a series of convolutional operations to extract essential features from both the robot's initial state images and the actuation parameters into a latent space.

2) Decoder: The decoder of the network serves as an image reconstruction part. It learns the intricate relationships among the encoded features that correspond to the robot's behavior and internal and external interactions. Through a sequence of transposed convolutional operations, the decoder up-samples the condensed feature data to reconstruct images. Then a regression output layer refines these reconstructions to align with the dimensions of the original input images.

III. SIMULATION AND RESULTS

A. Simulation Setup

We developed a simulated soft continuum robot in MATLAB using the SoRoSim toolbox [28], which is based on the Cosserat rod model mentioned in Section II. The simulated robot has the same parameters as the STIFF-FLOP robot used in physical experiments, which are detailed in Section IV. The robot consists of two segments, each with three thread-like actuators placed equidistantly along the perimeter of the robot. As such systems exhibit minimal torsion, elongation, and shear, we parameterized only the two directions of bending, each with a 4^{th} order Legendre polynomial per section. This parametrization results in a total of 20 degrees of freedom for the entire robot.

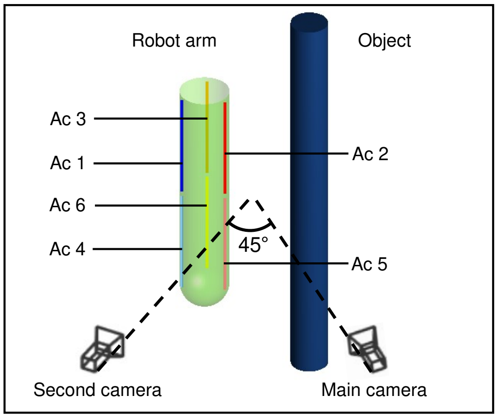


Fig. 3. Simulated experimental setup: The soft robot arm has 6 actuators (Ac 1 to Ac 6) depicted as lines. The main camera is positioned facing the central plane of the robot and object. The second camera is on the same plane at a 45° angle, providing binocular views of the interaction.

To simulate the robot's interaction with the environment, we implemented a simple contact model between the soft robot and a rigid object, as presented in [29]. The contact is considered to be a concentrated force from the object on the robot at the point of contact, represented by the $\mathbf{F}_e$ term in equation (5). Readers are referred to [30], [31] for more details on how concentrated forces are projected onto the generalized coordinates space. Once the simulation environment is set up, equation (5) is solved for the system's static configuration for different values of actuation, using root-finding methods implemented in the SoRoSim. A diagram of the simulation setup is depicted in Fig. 3.

For the training, a mini-batch size of 256 and the Adam optimizer were employed. The training data were split into 80% and 20% for training and validation, respectively.

B. Prediction Results

The prediction results presented in Fig. 4 demonstrate the effectiveness of the deep visual model in predicting simulated interaction shapes. The yellow mask highlights pixels with significant differences, exceeding a threshold of 10 (on a scale of 0-255) in all RGB channels, between the predicted and ground truth images. It provides a clear visual assessment of the model's performance and uncertainty regions.

When trained only on monocular images from the main camera (Fig. 4(a)), the model performs well in scenarios where the deformations are fully visible. However, its performance degrades in occlusion scenarios where the object or robot is partially obscured. The results present two specific cases where the model struggles with relative position prediction. The uncertainty regions, indicated by the yellow masks, are primarily located near the robot body, suggesting that the model struggles to accurately predict the robot's deformation and relative position when trained only on monocular data.

To address this limitation, binocular image data was introduced into the training process. The results in Fig. 4(b) show a notable improvement in the model's ability to understand spatial relationships and predict robot deformations, even in more complex occlusion scenarios. The reduced yellow

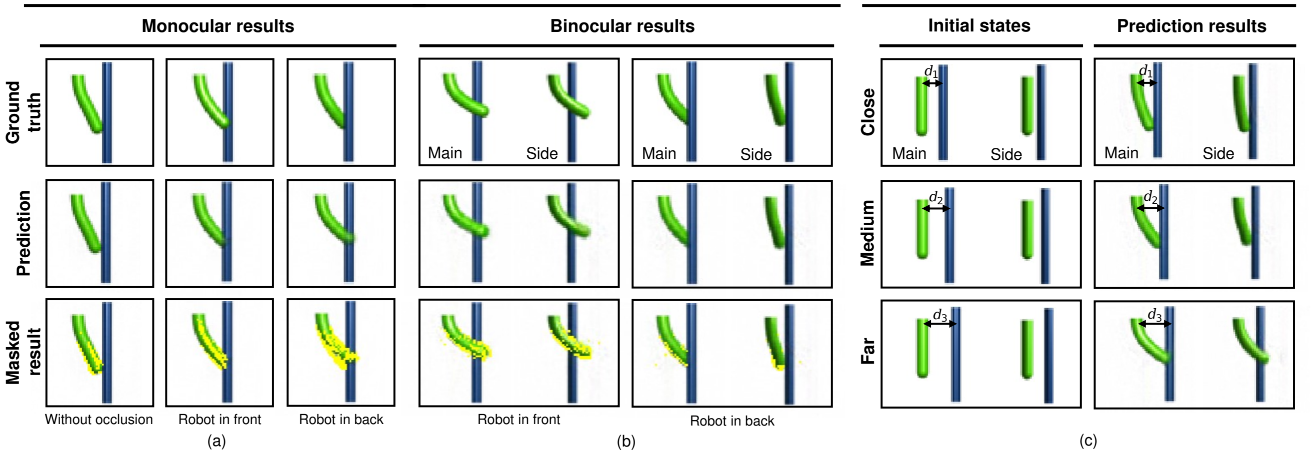


Fig. 4. Prediction results on simulated data. (a) shows predictions from the model trained on monocular image data. (b) shows predictions from the model trained on binocular image data, including two occlusion scenarios. The yellow mask highlights the pixels with a large difference between the predicted and ground truth images, indicating prediction uncertainty regions. (c) shows predictions when initial state images change with varying object-robot distance, but using the same actuation values.

masking in the binocular results compared to the monocular results demonstrates the enhanced predictive capability of the model when trained on binocular data.

Furthermore, Fig. 4(c) presents an analysis of the model's performance when the initial state images change, specifically when the distance between the object and the robot varies while maintaining the same actuation values. The results indicate that the model can effectively capture the change in the spatial relationships from initial state images and predict the interaction shapes accordingly.

Overall, the results highlight the importance of binocular vision in improving the model's understanding of spatial relationships and its ability to predict interaction shapes accurately in occlusion scenarios. This is because binocular vision provides more accurate information about the initial condition which is vital for predicting interaction dynamics.

C. Quantitative Analysis

We use the root mean square error (RMSE) to quantitatively evaluate the prediction performance of the model. Fig. 5 illustrates the relationship between the size of the training dataset and the RMSE. As the dataset size increases from 100 to 1000 samples, the RMSE values consistently decrease for both monocular and binocular datasets. Comparing the monocular and binocular results, the model achieves only a modest reduction in RMSE. This suggests that the additional spatial information provided by the side view has little impact on the model's numerical prediction accuracy. However, with the introduction of more diverse objects and environments, the benefits of stereo vision would become vital to localize the object with respect to the robot.

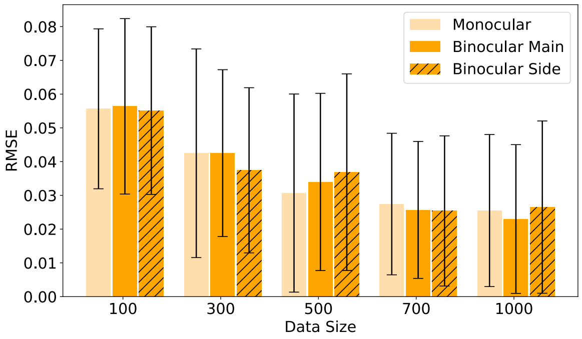
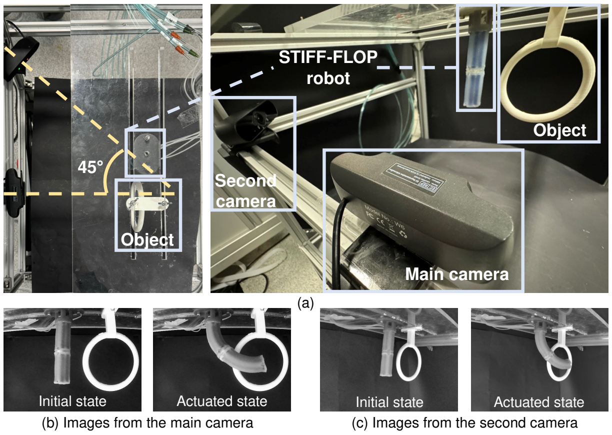


Fig. 5. Effect of dataset size on prediction performance for monocular and binocular data in simulated environment.



(b) Images from the main camera

(c) Images from the second camera

Fig. 6. Physical experimental setup: (a) The STIFF-FLOP robot with an object. The main camera faces the robot and the object, while the second camera is positioned at a 45° angle. (b) Example grayscale images captured by the main camera, showing the robot in its initial state (left) and actuated state (right). (c) Corresponding images captured by the second camera.

IV. PHYSICAL EXPERIMENT AND RESULTS

A. Experiment Setup

To evaluate the model with real-world data, we conducted a series of experiments using the STIFF-FLOP robot, a silicone-based, pneumatic-driven soft robotic arm [32] designed for applications in MIS [33]. In our experiment, we used a fiber-reinforced STIFF-FLOP robot comprising two segments connected in series. Each robotic segment is made of highly deformable silicone material (Ecoflex 00-

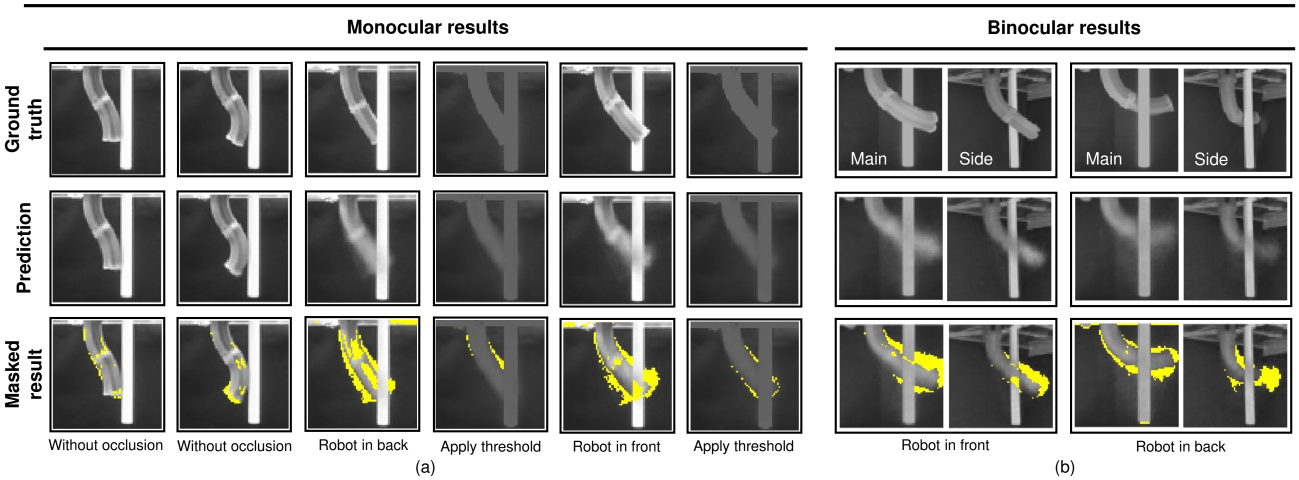


Fig. 7. Prediction results on real-world data. (a) shows predictions from the model trained on monocular image data. By applying a threshold to the gray value, the robot and the object are considered as an entity to mitigate the occlusion problem caused by their relative positions. (b) shows predictions from the model trained on binocular image data.

50, SmoothOn), featuring a diameter of 25 mm and a total length of 46 mm. The six actuation chambers of the robot are reinforced using inextensible threads to constrain the radial expansion. The design and fabrication details of the robot are thoroughly presented in a previous work [34]. The actuation system of the robotic platform consists of a pressurized air compressor (HYUNDAI HY5508), and six proportional pressure regulators (Camozzi K8P). An Arduino Due board controls the STIFF-FLOP robot's motion by regulating analogue signals to the pressure regulators while monitoring the real-time pressure.

The configuration of the experimental platform [35] and the soft continuum robot is shown in Fig. 6. The robot is placed next to a stationary object at a fixed distance. Two RGB cameras with a resolution of 1920×1080 are positioned at specific locations to capture the interactive state images from different view angles. When the robot is actuated, the two cameras capture the images together, and the pressure values read from the pressure regulators are recorded simultaneously.

B. Data Collection and Preprocess

During the data collection session, the actuation signals are randomly generated and sent to the robot at a sending rate of 2 Hz. For each signal, the robot remains actuated, the pressure values from the regulators are recorded, and a snapshot trigger simultaneously prompts the cameras to capture images. The latency between the images captured by the two cameras is less than 10 milliseconds, ensuring near-simultaneous data capture. To enrich the dataset, various interaction objects (cylinder, circular ring, and oblique prism) were placed at different distances from the robot during the data collection process. These object shapes represent a diverse range of geometric constraints that the robot might encounter in real-world applications.

Following data collection, the captured images are converted to grayscale, and the regions containing the robot and object are cropped and resized to a dimension of 64×64. Each set of image data is paired with its corresponding pressure data, represented by one-dimensional arrays with six elements. They are normalized and resized to 64×64, and then concatenated with the image data for training. The training options, such as batch size, optimizer, and data splitting ratio, are kept the same as those used for the simulated data.

C. Prediction Results

The prediction results presented in Fig. 7 demonstrate the capabilities of the deep visual model in predicting real-world interaction shapes. Fig. 7(a) shows the prediction results of the model trained using monocular image data. For the first two simple interaction scenarios, where the entire deformation of the robot can be observed from the camera view, the network's predictions show only minor differences compared to the ground truth. This indicates the network's effective performance in predicting simple 2D interaction shapes. However, when the interaction scenarios become more complex, involving occlusion, the network faces difficulties in accurately predicting the interaction shape. The performance is notably worse than the results obtained on simulated data. The uncertainty region, highlighted in the masked results, increases along the robot from base to tip, with the largest uncertainty near the tip.

Several factors may lead to the network's poor performance in occlusion cases, which include inherent challenges such as hysteresis, lack of information in the initial image, or network-related issues, such as an improper loss function. To investigate these factors and mitigate the influence of occlusion, we applied a grayscale threshold to the images. This thresholding approach treats the robot and the object as a single entity within the images. The network was retrained using the same dataset with this thresholding applied. The predictions after threshold adjustment show improvements compared to the raw images, demonstrating the network's

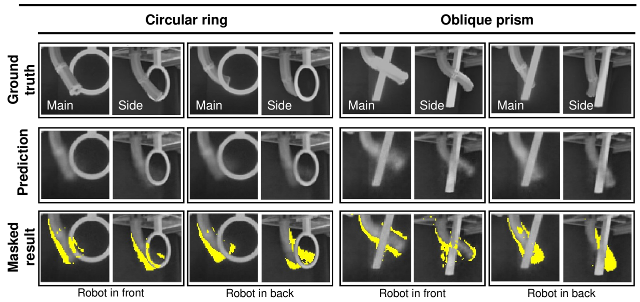


Fig. 8. Prediction results on interactions with various objects.

capability to predict robot deformations more accurately when the effects of occlusion are reduced.

To further investigate whether the errors are happening because of unobservable initial conditions, an additional view is integrated to provide the network with more spatial information. The results in Fig. 7(b) indicate that the model generally predicts the robot's deformation well. However, the uncertainty region, highlighted in yellow, still increases from the robot's base towards its tip. This observation can be attributed to that hysteresis in this soft robot affects the tip more significantly, and the real robot also exhibits elongated motion, leading to increased prediction uncertainty in this part. Additionally, noise in the training data and the inherent stochasticity of the robot-environment interaction problem itself make it more challenging for the model to accurately capture the behavior of the physical robot.

Comparing the binocular results across different objects (Fig. 8), the model demonstrates a good ability to predict the robot's interaction deformations. However, it struggles to accurately distinguish the relative positions between the robot and the object, especially when the robot is positioned in front of the object. This effect is particularly noticeable with complex object geometries, such as the circular ring.

V. CONCLUSION

Fig. 9 presents the RMSE of the model's predictions for monocular and binocular data from both simulated and physical experiments. The performance on simulated data is consistently better than on real-world data for both monocular and binocular datasets. This difference can be attributed to several factors, including synchronization issues during the real-world data collection process and the presence of uncertainties in the physical experiments. In real-world experiments, the data collection process involved coordinating multiple sensors, including regulators, cameras, and pressure sensors, which can introduce synchronization errors. These errors can lead to misalignments between the visual data and the corresponding actuation values, negatively impacting the model's learning ability. The real robot exhibits elongated motion, which is not considered in the simulation, further contributing to the discrepancy between the simulated and real-world results. Additionally, real-world experiments are subject to various sources of uncertainty, such as camera

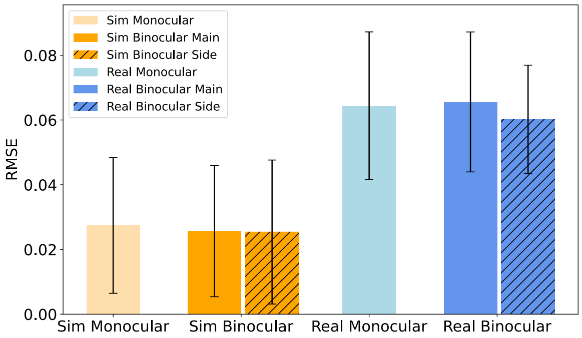


Fig. 9. Comparison of prediction performance on simulated and real-world data.

positions, and external disturbances like environmental light, which can further degrade the model's performance.

In conclusion, the comparison result confirms the observations from the previous sections. The analysis of the RMSE performance across different dataset sizes highlights the importance of having a sufficient amount of diverse and representative training data for accurate interaction shape prediction. The model's performance is better on simulated data than on real-world data, highlighting the impact of real-world uncertainties and data collection challenges on the model's prediction accuracy. Furthermore, the incorporation of binocular vision provides only a modest improvement in prediction accuracy compared to monocular vision. The promising results of this work can be extended to various applications, such as MIS. Given the CT scan of an internal cavity (initial condition), this model can be used to predict the shape and location of the MIS tool, enabling surgeons to better plan and execute surgical procedures.

Future research could focus on several key areas to further enhance the model's performance and applicability. Firstly, integrating more spatial information from multiple viewpoints to improve the model's accuracy in challenging occlusion scenarios. Secondly, integrating data from embedded sensors, such as force or tactile sensors, could provide additional sensing modalities and help the model better estimate the interaction forces. Thirdly, conducting a deeper analysis of the outputs from different layers of the network to get insights into the learned features and their relevance to the interaction shape prediction task, exploring techniques such as saliency maps on the output to indirectly estimate robot-environment contact locations. Furthermore, leveraging the trained model in a model predictive controller could enable more adaptable and robust interaction shape control strategies.

ACKNOWLEDGMENTS

This work was supported by the US Office of Naval Research Global under Grant N62909-21-1-2033, in part by Khalifa University under Awards No. RIG-2023-048, RC1-2018-KUCARS and the Royal Society research grant RGS/R1/231472, in part by the Engineering and Physical Sciences Research Council under Grant EP/R037795/1, Grant EP/S014039/1, and Grant EP/V01062X/1, in part by the Royal Academy of Engineering under Grant IAPP18-19/264, and in part by the UCL Dean's Prize.

REFERENCES

[1] D. Rus and M. Tolley, "Design, fabrication and control of soft robots," *Nature*, vol. 521, no. 7553, pp. 467–475, 2015.

[2] J. Burgner-Kahrs, D. Rucker, and H. Choset, "Continuum Robots for Medical Applications: A Survey," *IEEE Transactions on Robotics*, vol. 31, no. 6, pp. 1261–1280, 2015.

[3] K.-W. Kwok, H. Wurdemann, A. Arezzo, A. Menciassi, and K. Althoefer, "Soft Robot-Assisted Minimally Invasive Surgery and Interventions: Advances and Outlook," *Proceedings of the IEEE*, vol. 110, no. 7, pp. 871–892, 2022.

[4] M. Cianchetti, C. Laschi, A. Menciassi, and P. Dario, "Biomedical applications of soft robotics," *Nature Reviews Materials*, vol. 3, no. 6, pp. 143–153, 2018.

[5] E. Almanzor, F. Ye, J. Shi, T. Thuruthel, H. Wurdemann, and F. Iida, "Static Shape Control of Soft Continuum Robots Using Deep Visual Inverse Kinematic Models," *IEEE Transactions on Robotics*, vol. 39, no. 4, pp. 2973–2988, 2023.

[6] G. Mengaldo, F. Renda, S. Brunton, M. Bächer, M. Calisti, C. Duriez, G. Chirikjian, and C. Laschi, "A concise guide to modelling the physics of embodied intelligence in soft robotics," *Nature Reviews Physics*, 2022.

[7] J. Shi, A. Shariati, S.-A. Abad, Y. Liu, J. S. Dai, and H. A. Wurdemann, "Stiffness modelling and analysis of soft fluidic-driven robots using Lie theory," *The International Journal of Robotics Research*, vol. 43, no. 3, pp. 354–384, 2024.

[8] C. Duriez and T. Bieze, "Soft robot modeling, simulation and control in real-time," in *Soft Robotics: Trends, Applications and Challenges*. Cham: Springer International Publishing, 2017, pp. 103–109.

[9] Y. Hu, J. Liu, A. Spielberg, J. B. Tenenbaum, W. T. Freeman, J. Wu, D. Rus, and W. Matusik, "ChainQueen: A Real-Time Differentiable Physical Simulator for Soft Robotics," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 5 2019, pp. 6265–6271.

[10] J. Till, V. Aloi, and C. Rucker, "Real-time dynamics of soft and continuum robots based on Cosserat rod models," *International Journal of Robotics Research*, vol. 38, no. 6, pp. 723–746, 5 2019.

[11] F. Boyer, V. Lebastard, F. Candelier, and F. Renda, "Dynamics of Continuum and Soft Robots: A Strain Parameterization Based Approach," *IEEE Transactions on Robotics*, vol. 37, no. 3, pp. 847–863, 6 2021.

[12] E. H. Dowell and K. C. Hall, "Modeling of Fluid-Structure Interaction," *Annual Review of Fluid Mechanics*, vol. 33, no. 1, pp. 445–490, 1 2001.

[13] E. Coevoet, A. Escande, and C. Duriez, "Optimization-Based Inverse Model of Soft Robots With Contact Handling," *IEEE Robotics and Automation Letters*, vol. 2, no. 3, pp. 1413–1419, 7 2017.

[14] C. Della Santina, C. Duriez, and D. Rus, "Model-Based Control of Soft Robots: A Survey of the State of the Art and Open Challenges," *IEEE Control Systems*, vol. 43, no. 3, pp. 30–65, 6 2023.

[15] C. Della Santina, R. K. Katzschmann, A. Biechi, and D. Rus, "Dynamic control of soft robots interacting with the environment," in *2018 IEEE International Conference on Soft Robotics (RoboSoft)*. IEEE, 4 2018, pp. 46–53.

[16] D. Kim, S. H. Kim, T. Kim, B. B. Kang, M. Lee, W. Park, S. Ku, D. W. Kim, J. Kwon, H. Lee, J. Bae, Y. L. Park, K. J. Cho, and S. Jo, "Review of machine learning methods in soft robotics," *PLoS ONE*, vol. 16, no. 2 February, 2 2021.

[17] Y. Wang, Y. Fan, J. Wang, and W. Chen, "Long-term navigation for autonomous robots based on spatio-temporal map prediction," *Robotics and Autonomous Systems*, vol. 179, p. 104724, 2024.

[18] X. Wang, Y. Li, and K.-W. Kwok, "A Survey for Machine Learning-Based Control of Continuum Robots," *Frontiers in Robotics and AI*, vol. 8, 9 2021.

[19] K. Elgeneidy, N. Lohse, and M. Jackson, "Bending angle prediction and control of soft pneumatic actuators with embedded flex sensors – A data-driven approach," *Mechatronics*, vol. 50, pp. 234–247, 4 2018.

[20] M. Hofer, C. Sferrazza, and R. D'Andrea, "A Vision-Based Sensing Approach for a Spherical Soft Robotic Arm," *Frontiers in Robotics and AI*, vol. 8, 2 2021.

[21] J. Lu, F. Liu, C. Girerd, and M. C. Yip, "Image-based Pose Estimation and Shape Reconstruction for Robot Manipulators and Soft, Continuum Robots via Differentiable Rendering," in *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2023-May. Institute of Electrical and Electronics Engineers Inc., 2023, pp. 560–566.

[22] T. George Thuruthel, B. Shih, C. Laschi, and M. Thomas Tolley, "Soft robot perception using embedded soft sensors and recurrent neural networks," Tech. Rep., 2019. [Online]. Available: https://www.science.org

[23] J. D. L. Ho, K.-H. Lee, W. L. Tang, K.-M. Hui, K. Althoefer, J. Lam, and K.-W. Kwok, "Localized online learning-based control of a soft redundant manipulator under variable loading," *Advanced Robotics*, vol. 32, no. 21, pp. 1168–1183, 11 2018.

[24] G. Fang, X. Wang, K. Wang, K. H. Lee, J. D. Ho, H. C. Fu, D. K. C. Fu, and K. W. Kwok, "Vision-Based Online Learning Kinematic Control for Soft Robots Using Local Gaussian Process Regression," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1194–1201, 4 2019.

[25] H. Li, L. Xun, and G. Zheng, "Piecewise Linear Strain Cosserat Model for Soft Slender Manipulator," *IEEE Transactions on Robotics*, 2023.

[26] F. Renda, C. Armanini, A. Mathew, and F. Boyer, "Geometrically-Exact Inverse Kinematic Control of Soft Manipulators with General Threadlike Actuators' Routing," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7311–7318, 7 2022.

[27] F. Renda, C. Armanini, V. Lebastard, F. Candelier, and F. Boyer, "A Geometric Variable-Strain Approach for Static Modeling of Soft Manipulators with Tendon and Fluidic Actuation," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4006–4013, 7 2020.

[28] A. T. Mathew, I. B. Hmida, C. Armanini, F. Boyer, and F. Renda, "SoRoSim: A MATLAB Toolbox for Hybrid Rigid–Soft Robots Based on the Geometric Variable-Strain Approach," *IEEE Robotics & Automation Magazine*, vol. 30, no. 3, pp. 106–122, 9 2023.

[29] A. T. Mathew, C. Armanini, A. A. S. A. Alshehhi, I. M. B. Hmida, and F. Renda, "Multifunctional Underwater Soft Robots: A Simulation Essay," *IOP Conference Series: Materials Science and Engineering*, vol. 1261, no. 1, p. 012008, 10 2022.

[30] A. T. Mathew, D. Feliu-Talegon, A. Y. Alkayas, F. Boyer, and F. Renda, "Reduced order modeling of hybrid soft-rigid robots using global, local, and state-dependent strain parameterization," *The International Journal of Robotics Research*, vol. 0, no. 0, p. 02783649241262333, 2024. [Online]. Available: https://doi.org/10.1177/02783649241262333

[31] C. Armanini, M. Farman, M. Calisti, F. Giorgio-Serchi, C. Stefanini, and F. Renda, "Flagellate Underwater Robotics at Macroscale: Design, Modeling, and Characterization," *IEEE Transactions on Robotics*, vol. 38, no. 2, pp. 731–747, 4 2022.

[32] H. A. Wurdemann, A. Stilli, and K. Althoefer, "Lecture Notes in Computer Science: An Antagonistic Actuation Technique for Simultaneous Stiffness and Position Control," 2015, pp. 164–174.

[33] M. Cianchetti, T. Ranzani, G. Gerboni, T. Nanayakkara, K. Althoefer, P. Dasgupta, and A. Menciassi, "Soft Robotics Technologies to Address Shortcomings in Today's Minimally Invasive Surgery: The STIFF-FLOP Approach," *Soft Robotics*, vol. 1, no. 2, pp. 122–131, 6 2014.

[34] J. Shi, S.-A. Abad, J. S. Dai, and H. A. Wurdemann, "Position and orientation control for hyperelastic multisegment continuum robots," *IEEE/ASME Transactions on Mechatronics*, vol. 29, no. 2, pp. 995–1006, 2024.

[35] J. Shi, W. Gaozhang, H. Jin, G. Shi, and H. A. Wurdemann, "Characterisation and control platform for pneumatically driven soft robots: Design and applications," in *2023 IEEE International Conference on Soft Robotics (RoboSoft)*, 2023, pp. 1–8.