This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.
Smart e-skins and machine learning for soft robot perception

Delin Hu

THE UNIVERSITY OF EDINBURGH

A thesis submitted for the degree of Doctor of Philosophy
THE UNIVERSITY OF EDINBURGH

Feb 2023
Abstract

Perception is the foundation for intelligent robots to effectively explore environments and interact with users. Bio-inspired soft robots can exploit the compliance of their bodies, consequently demonstrating advantages in terms of safety during human-robot interaction and operability in unstructured scenarios. However, this very feature dramatically increases the complexity of perception, which prevents soft robots from wide adoption in practical applications. This thesis aims at this challenge to implement a comprehensive study in coupling field simulation, sensor design and fabrication and learning-based perception algorithms. The major contributions of this thesis can be summarised as follows.

First, coupling field simulation (CFS) was developed to integrate sensory systems into mechanical structures, which is critical for developing perceptive soft robots but neglected by most existing studies. The proposed simulation method was demonstrated through a 16-electrode capacitive sensor array deployed on a soft robot manipulator. The understanding of the sensor behaviours was built by simulating the sensor responses to a range of deformations using CFS. One of the most important applications of CFS is to generate annotated samples which can be used to train learning-based perception algorithms. The trained learning models could be transferred to practical scenarios without the need for tremendous annotated real-world data, thus reducing time and costs for data acquisition. Two case studies for applied force estimation and deformation classification were performed with annotated data generated by CFS to demonstrate its potential in developing learning-based perception methods.

On soft robot proprioception, previous studies only achieved low proprioceptive geometry resolution (PGR), thus suffering from loss of geometric details (e.g., local deformation and surface information) and limited applicability (e.g., only applicable to prescribed simple deformations). This thesis proposed a high PGR soft robot proprioception system, which encapsulates an intrinsically stretchable capacitive e-skin (SCAS) and a capacitance-to-deformation transformer (C2DT), to endow full-geometry, millimetre-level bodily awareness to soft robots. The SCAS has a redundant planar electrode layout that forms a sequence of capacitors sensitive to deformation across the entire soft robot. The C2DT based on transformer architecture can explore the dependency
over the SCAS signals and recover deformation from the signals. The proposed proprioception framework synergistically combining the SCAS and the C2DT can achieve real-time (30 fps), high PGR (3,900) full-geometry deformation reconstruction with high accuracy (2.322 ± 0.687 mm CD error) under a range of complex deformations on a 20 × 20 × 200 mm soft manipulator. This high PGR was not demonstrated previously and is one or two orders of magnitude improvement over previous methods.

Regarding tactile sensing (one of the most important exteroceptions), previous studies ignored the impact of deformation on the sensing data, making them a mismatch to practical scenarios where the geometric and tactile variations are coupled in the sensor signals. The thesis performed a preliminary exploration of tactile sensing with severe interference, which is inevitable in some practical applications. A simplified SCAS was proposed to achieve touch recognition with the interference induced by deformation while reducing the complexity of fabrication, deployment and wiring. The proposed method was validated on a pneumatic robotic platform. Contact location estimation was successfully achieved at low spatial resolution in various inflation conditions with 99.88% of classification accuracy. Moreover, deformation sensing with the interference induced by physical contact was also demonstrated by the coordinates estimation of 5 visual markers deployed on the soft robot platform. The C2DT was employed to perform marker tracking with the information of the first frame in a trajectory as prior knowledge and achieved 2.905±2.207 mm AD error, which shows the potential of the simplified SCAS to simultaneously detect internal and external stimuli.

In summary, this thesis presented a framework for soft robot perception, which comprises stretchable capacitive e-skins to translate geometric and tactile variations to capacitance data and neural architectures to interpret the capacitance data to desired parameters. High PGR morphological reconstruction and tactile sensing with interference were demonstrated, paving the way towards the autonomy of soft robots.
In nature, perception refers to the ability of living organisms to sense internal (e.g., body movement and deformation) and external stimuli (e.g., physical contact), which constitutes the foundation of dexterous body manipulation. Intelligent robots are expected to possess similar capabilities, which enable them to realise lifelike functionality. Soft robots are made of flexible materials such as silicone and can exploit the compliance of their bodies. This feature improves the safety of human users and operability in unstructured environments, making soft robots preferable to various applications such as human-robot interaction and surgical robots. However, the deformable nature also brings new challenges to soft robot perception.

This thesis aims to develop a perception system that has the potential to achieve autonomy on emerging soft robotic platforms. This requires the perception system can provide real-time high-resolution morphological feedback and tactile information under inevitable interference induced by the deformation of the soft robots. Therefore, the thesis proposes a novel perception framework empowered by intrinsically stretchable capacitive e-skins and deep neural architectures. The results of this thesis progress soft robot perception a step forward, paving the way beyond open-loop control.
Acknowledgements

First of all, I would like to express my sincere gratitude to my principal supervisor Dr Yunjie Yang, who leads the SMART group at the University of Edinburgh, for his guidance, support, and encouragement throughout my PhD study. His insightful comments, valuable feedback and patient guidance play a critical role in completing the thesis. His passion for research and pursuit of tackling the most challenging issues in the field inspire me a lot and have a long-term impact on my academic career. Our collaboration during the 3.5-year PhD program is incredibly enjoyable.

I would also like to thank my assistant supervisor Dr Francesco Giorgio-Serchi for his stimulating discussions and constructive comments in our weekly meetings, which enables me to smoothly progress my research. I have learned a lot from his profound knowledge of soft robotics. Meanwhile, I want to thank my assistant supervisor Dr Mehrdad Yaghoobi Vaighan for his valuable suggestions for my research and career development.

My appreciation also goes to my colleagues in the SMART Group and the Agile Tomography Group. They are Dr Shengnan Wang, Dr Haokun Wang, Dr Zhou Chen, Zhe Liu, Huazhi Dong, Abdul Hafeez Abdul Bari, Rui Zhang, Jiangnan Xia, Dr Yuan Chen, Yinhao Li, Dr Zhixi Zhang, Wei Han, Hao Yu and Yalei Fu. Thanks for their kind assistance and support in my research and daily life.

Furthermore, the School of Engineering and LeEngSTAR is gratefully acknowledged for providing the studentship to support my PhD study.

Last but most importantly, I am grateful to my family for their selfless love and understanding. Their constant support, encouragement and belief, which help me pass through hard times, have been the cornerstone of my success and have given me the strength to pursue my dreams.
Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Delin Hu
8th June 2023
# Contents

Abstract ii

Lay Summary iv

Acknowledgements v

Declaration vi

List of Figures x

List of Tables xiv

Acronyms and Abbreviations xv

Publication list xvi

1 Introduction 1

1.1 Background and motivation . . . . . . . . . . . . . . . . . . . . . . . 1

1.2 Objectives . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

1.3 Main contributions . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

1.4 Thesis outline . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

2 Literature review 6

2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

2.2 Simulation methods for soft robots . . . . . . . . . . . . . . . . . . 6

2.3 Soft robot proprioception . . . . . . . . . . . . . . . . . . . . . . . . 8

2.3.1 Sensors based on different transducer mechanisms . . . . . . 8

2.3.2 Sensing data interpretation with machine learning tools . . . 10

2.3.3 Proprioceptive geometry resolution . . . . . . . . . . . . . . . . 12

2.4 Tactile sensing for soft robots . . . . . . . . . . . . . . . . . . . . . 14

2.5 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16

3 Coupling field simulation for soft robot perception 18

3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18

3.2 Solid mechanics and electrostatics coupling field simulation . . . 19
# Smart e-skins and machine learning for soft robot perception

4.7.1 Modification on the framework .......................... 78
4.7.2 Implementation .......................................... 79
4.8 Reconstruction results on the real-world dataset .......... 80
4.9 Summary ..................................................... 85

5 Tactile sensing under interference from deformation 86
5.1 Introduction ................................................ 86
5.2 Simplified SCAS ........................................... 87
5.3 Characterisation ............................................ 89
  5.3.1 Inflation ................................................ 89
  5.3.2 Touch under inflation .................................. 90
5.4 Experiment and data acquisition ............................ 94
5.5 Touch recognition ........................................... 96
  5.5.1 Neural network architecture .......................... 96
  5.5.2 Implementation ........................................ 97
  5.5.3 Results .................................................. 97
5.6 Deformation tracking ........................................ 98
  5.6.1 Neural network architecture .......................... 98
  5.6.2 Implementation ........................................ 99
  5.6.3 Results .................................................. 99
5.7 Summary ..................................................... 103

6 Conclusions and future work 104
6.1 Conclusions ................................................ 104
6.2 Future work ................................................ 106
  6.2.1 Automated fabrication ................................ 106
  6.2.2 Multi-modal sensing ................................... 106
  6.2.3 Sim-to-real transfer learning .......................... 107

Reference 108
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Conceptual illustration of the advantage of a high PGR proprioception system over a low PGR (=2) proprioception system [1] in previous research when a segment of the object bends.</td>
</tr>
<tr>
<td>2.2</td>
<td>Conceptual illustration of the advantage of a high PGR proprioception system over a low PGR (=2) proprioception system [1] in previous research when a segment of the object twists.</td>
</tr>
<tr>
<td>2.3</td>
<td>Conceptual illustration of the advantage of a high PGR proprioception system over a low PGR (=2) proprioception system [1] in previous research when the bend direction is not fixed.</td>
</tr>
<tr>
<td>2.4</td>
<td>Conceptual illustration of the advantage of a high PGR proprioception system over a low PGR (=2) proprioception system [1] in previous research when the twisting axis is not fixed.</td>
</tr>
<tr>
<td>3.1</td>
<td>Flowchart of coupling field simulation (CFS) of a 16-electrode soft capacitive sensor array deployed on a square soft arm.</td>
</tr>
<tr>
<td>3.2</td>
<td>Schematic illustration of the soft robot manipulator. Note that the other 8 electrodes are deployed on two hidden surfaces.</td>
</tr>
<tr>
<td>3.3</td>
<td>Examples of deformations (bending) in several selected frames and the calibrated capacitance readouts for all electrode pairs. Lines with different colors indicate different capacitance readouts.</td>
</tr>
<tr>
<td>3.4</td>
<td>Examples of deformations (two-stage twisting and bending) in several selected frames and the calibrated capacitance readouts for all electrode pairs.</td>
</tr>
<tr>
<td>3.5</td>
<td>Examples of deformations (elongation and twisting) in several selected frames and the calibrated capacitance readouts for all electrode pairs.</td>
</tr>
<tr>
<td>3.6</td>
<td>The calibrated capacitance readouts for 4 selected electrode pairs during the whole deformation process (bending).</td>
</tr>
<tr>
<td>3.7</td>
<td>Calibrated capacitance curves over time of bending deformations in two different directions.</td>
</tr>
<tr>
<td>3.8</td>
<td>The calibrated capacitance readouts for 4 selected electrode pairs during the whole deformation process (two-stage twisting and bending).</td>
</tr>
</tbody>
</table>
Smart e-skins and machine learning for soft robot perception

3.9 The calibrated capacitance readouts for 4 selected electrode pairs during the whole deformation process (elongation and twisting). 28

3.10 Absolute error for force estimation on the testing set. Left: absolute error for force component along the x axis. Right: absolute error for force component along the y axis. 30

3.11 Examples of force estimation on the testing set. Top: the ground truth deformations. Bottom: the ground truth applied force (red dot line) and the estimated force (blue line). 31

3.12 Examples of deformation classification on the testing set. Top: the ground truth deformations. Middle: the corresponding measurements (the input of the MLP classifier). Bottom: the output of the MLP classifier. 33

4.1 Schematic illustration of the SCAS. 36

4.2 The CFS model of a 64-electrode SCAS deployed on a soft arm with 64 visual markers. 37

4.3 Electrode pairs (activated, marked in orange) in the first layer to form 28 measurable independent capacitors. 39

4.4 Electrode pairs (activated, marked in orange) between layer 1 and layer 2 form 24 measurable independent capacitors. 40

4.5 A set of examples in one testing episode with a continuous \( L(z,r) \) deformation process. Capacitance readouts, ground truth point clouds and reconstruction of the original C2DT, the C2DT w/o markers and the C2DT w/o Chamfer distance are shown. The colour of each point is the distance from the corresponding ground truth point. 42

4.6 A set of examples in one testing episode with a continuous \( L(x,y) \) deformation process. Capacitance readouts, ground truth point clouds and reconstruction of the original C2DT, the C2DT w/o markers and the C2DT w/o Chamfer distance are shown. The colour of each point is the distance from the corresponding ground truth point. 43

4.7 A set of examples in one testing episode with a continuous \( L_r(x,y) \) deformation process. Capacitance readouts, ground truth point clouds and reconstruction of the original C2DT, the C2DT w/o markers and the C2DT w/o Chamfer distance are shown. The colour of each point is the distance from the corresponding ground truth point. 44
4.8 A set of examples in one testing episode with a continuous $L_{(x,y,r)}$ deformation process. Capacitance readouts, ground truth point clouds and reconstruction of the original C2DT, the C2DT w/o markers and the C2DT w/o Chamfer distance are shown. The colour of each point is the distance from the corresponding ground truth point. ........................................ 45

4.9 Neural architecture of C2DT. .................................................. 48

4.10 A set of examples of reconstructions generated by different C2DTs. .... 53

4.11 Visualisation of position representations for the simulation via t-SNE. ... 56

4.12 The performance of the C2DT under 4 different numbers of markers (mean ± standard deviation). ................................................. 57

4.13 The performance of the C2DT under 4 different electrode layouts (mean ± standard deviation). ............................................... 58

4.14 Four-electrode SCAS modules with EGaIn (centre) and CB (right) wires and the cross-section of the module with EGaIn wires under a 40x digital microscope (left). ......................................................... 59

4.15 Fabrication process of a modular 4-electrode SCAS. ..................... 61

4.16 The soft arm equipped with 8 4-electrode SCAS modules (top) is coated with a white silicone layer and 16 yellow visual markers to improve data acquisition. ................................................................. 63

4.17 The bonding interface between the SCAS and the soft robot platform ... 64

4.18 The SCAS modules are bonded to different sides of a short silicone pillar and cyclically stretched by a stepper motor. ....................... 66

4.19 Three relative capacitance readouts of the SCAS with EGaIn wires during the whole 3 h cycling test (left). The relative capacitance readouts the SCAS with EGaIn wires after 10, 100 and 500 cycles (right). ............. 67

4.20 Three relative capacitance readouts of the SCAS with CB wires during the whole 3 h cycling test (left). The relative capacitance readouts of the SCAS with CB wires after 10, 100 and 500 cycles (right). ............... 68

4.21 Relative capacitance response curves of the two SCAS modules to a 40% periodic linear stretch. The SCAS with EGaIn wires shows better sensitivity, linearity and cycling stability than its CB wires counterpart. ... 69

4.22 Real-world experiment platform. .............................................. 71

4.23 A representative set of complex deformations that appear in the experiment. ................................................................. 73

4.24 The curves of calibrated capacitance readouts of the SCAS during an about 20 s period in the experiment. ............................... 75
4.25 A group of adjacent frames in the real-world dataset is shown, which records a continuous bending process with RGB images, point clouds and capacitance readouts. 76

4.26 A group of adjacent frames under a deformation process that combines elongation and twisting. 77

4.27 The performance of C2DTs, which take different numbers of adjacent frames as inputs (mean±standard deviation). 80

4.28 A representative set of examples of high PGR 3D deformation reconstruction. 81

4.29 Qualitative illustration of ablation studies results. 83

4.30 Two-dimensional t-SNE embedding of the output of the position encoding part in the trained C2DT. 84

5.1 Schematic illustration of the simplified SCAS. 87

5.2 The undeformed (left) and 100% stretched (right) e-skins. 88

5.3 The structure of the 3-chamber pneumatic manipulator. 89

5.4 Twenty-eight calibrated capacitance readouts during the (20,20,20) ml inflation process. 89

5.5 Four selected calibrated capacitance readouts during the (20,20,20) ml inflation process. 91

5.6 Twenty-eight calibrated capacitance readouts during the 2-stage experiment. 92

5.7 Four selected calibrated capacitance readouts during the 2-stage experiment. 93

5.8 Division of the robot body for touch recognition. 94

5.9 Experiment platform. 95

5.10 The framework to estimate the touch point using an MLP. 97

5.11 Confusion map of the classification results. 98

5.12 The structure of the C2DT to estimate coordinates of visual markers based on capacitance signals. 100

5.13 Examples of tracking results. 101

5.14 A tracking trajectory of a selected marker. 102
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Machine learning methods for soft robot proprioception</td>
<td>11</td>
</tr>
<tr>
<td>2.2</td>
<td>Existing soft proprioception systems</td>
<td>12</td>
</tr>
<tr>
<td>3.1</td>
<td>Material parameters</td>
<td>21</td>
</tr>
<tr>
<td>4.1</td>
<td>Dataset specification</td>
<td>46</td>
</tr>
<tr>
<td>4.2</td>
<td>Quantitative evaluation of C2DTs with different network hyperparameters</td>
<td>52</td>
</tr>
<tr>
<td>4.3</td>
<td>Quantitative evaluation of C2DTs with different loss terms on the virtual proprioception dataset</td>
<td>54</td>
</tr>
<tr>
<td>4.4</td>
<td>Quantitative evaluation of C2DTs with different loss terms on the experimental dataset</td>
<td>82</td>
</tr>
</tbody>
</table>
Acronyms and Abbreviations

3D 3-dimensional
SCAS stretchable capacitive e-skin
CFS coupling field simulation
MLP multi-layer perceptron
C2DT capacitance-to-deformation transformer
PGR proprioceptive geometry resolution
FEM finite element method
SOFA simulation open framework architecture
SVM support vector machine
kNN k-nearest neighbours
GP Gaussian process
DT decision trees
LR linear regression
LSTM long short-term memory
RNN recurrent neural network
SVR support vector regression
MVLR multivariate linear regression
PR polynomial regression
PDMS polydimethylsiloxane
ECT electrical capacitance tomography
EIT electrical impedance tomography
ROI region of interest
MSE mean square error
MAE mean absolute error
AD average distance
MD maximal distance
CD Chamfer distance
HD Hausdorff distance
t-SNE t-distributed stochastic neighbour embedding
EGaIn Eutectic Gallium 75.5% Indium 24.5%
CB carbon black
RGB-D red, green, blue plus depth
fps frames per second
Publication list

Journal papers


Conference papers

Chapter 1

Introduction

1.1 Background and motivation

Conventional rigid robots have been successfully applied in a wide variety of industries over the past decades, dramatically improving efficiency and reducing costs. Many people, therefore, get rid of heavy physical labour and harsh working environments. With the development of technology, robots are no longer limited to traditional industrial applications but have gradually entered every aspect of people’s lives [2] such as rehabilitation robots [3, 4], surgical robots [5, 6], social robots [7, 8] and so on. Robotic structures comprised of rigid links and joints can not meet current requirements due to potential risks to the safety of human users. Soft robots inspired by biological organisms and made of flexible materials are able to exploit the softness and compliance of their bodies [9, 10]. This quality enables them to be superior to conventional rigid robots in terms of manoeuvrability, safety to human users and operability in unstructured environments, opening new possibilities to numerous fields, such as biomedicine [11, 12, 13], human-robot interaction [14, 15] and fragile object manipulation [16].

Perception is fundamental for intelligent robots to effectively explore unknown surroundings and safely interact with users, environments and other agents [17, 18]. However, the highly deformable nature makes soft robots incompatible with perception methods developed for their rigid counterparts. New perception systems considering the properties and demands of emerging soft robots need to be explored. Inspired by living organisms that can simultaneously detect internal and external stimuli, perception for soft robots is expected to include proprioception which refers to perceiving active and passive deformation of the robots’ bodies [19], and exteroception, which refers to detecting environmental stimuli such as physical contact [20], temperature [21], humidity [22] and so on.
Smart e-skins and machine learning for soft robot perception

From the perspective of proprioception, the situation of soft robots is much more challenging than that of their rigid counterparts. Within the frame of conventional rigid robots, existing sensing technology already provides viable solutions to implementing proprioception (e.g., body geometry estimation) that meets the requirements of even the most agile and complex robotic platforms. This is due to the inherent predictability of the rigid-body system, whose finite degrees of freedom allow the full geometry to be defined by a bounded set of measurable parameters (such as joint angle and link length) [23]. However, the highly deformable nature of soft robots represents their asset as well as their drawback. The bodily compliance significantly increases the degrees of freedom of a soft body to infinity, making it infeasible to accurately and completely describe the 3D morphology of the soft system with only a limited set of parameters. Although many related studies have been carried out and reported over the past decades, most of them only focus on oversimplified proprioception tasks (e.g., bending angle estimation [1, 24] and tip tracking [25] under prescribed deformations) which is far from the level of an acceptable solution for practical applications.

Tactile sensing is the most important modality of exteroception as it is the prerequisite for soft robots to physically interact with human users and environments. Two challenges make soft robot tactile sensing still an open issue. First, most existing soft tactile sensors are sensitive to both physical contact and body deformations [26]. This implies that the readout signals are susceptible to interference from deformations, dramatically increasing the complexity of decoding the desired tactile information (such as contact location and force) from the signal envelope. Most previous studies only demonstrate the tactile sensing performance on undeformed substrates, where the contribution from structural deformations is not taken into account [27, 28, 29]. Furthermore, considering the limitations of wiring and interfaces with readout electronics, the number of soft tactile sensors deployed in a unit area is limited. Methodologies that rely on sparse sensor layouts should be regarded as favourable.

In summary, the state of the art can hardly meet the requirements for emerging soft robots and their applications. Soft robot perception systems with the following properties are highly desired and need to be investigated.

- The perceptors should be stretchable, lightweight, deployable to irregular surfaces, and have an ignorable impact on the movement of the soft robot investigated.
Smart e-skins and machine learning for soft robot perception

- The perception systems should not be limited to oversimplified proprioception tasks (e.g., bending angle estimation and tip tracking under prescribed deformations) but also be applicable to challenging and important proprioception tasks (e.g. high-resolution morphological reconstruction).
- The perception systems are expected to be able to implement tactile sensing with the interference from deformation produced by internal actuation.

1.2 Objectives

The overall goal of this thesis is to establish soft robot perception frameworks to address current issues and achieve high-resolution morphological reconstruction and tactile sensing with a range of deformations. It can be divided into the following objectives.

1. Design and develop stretchable e-skins (perceptors) capable of detecting geometric variations and physical contacts. Investigate their feasibility for high-resolution morphological reconstruction, and tactile sensing with deformation interference.

2. Establish and implement numerical models to simulate the responses of e-skins to various stimuli, such as different types of deformations. Analyse and discuss the underlying mechanisms that govern the behaviour of the e-skins.

3. Utilize advanced manufacturing techniques, including 3D printing and laser patterning, to fabricate the e-skins. Conduct cycling experiments to quantitatively evaluate their performance.

4. Design and implement simulated and experimental annotated data generation methods to acquire a large volume of labelled data, which will be used for training, validating, and evaluating perception algorithms based on machine learning.

5. Design neural architectures to translate signals collected by the e-skins into target parameters. These parameters include high-resolution representations of soft robot morphologies and the locations of physical contacts. Develop codes and software to implement and evaluate the proposed methods.
1.3 Main contributions

This thesis concentrates on realising challenging but important proprioception and exteroception tasks for soft robots, mainly contributing to related communities in the following aspects.

1. Stretchable capacitive e-skins (SCASs) based on planar electrode arrays were proposed. They are lightweight, sensitive to geometric variations and physical contacts, and have a minute impact on soft robot platforms under investigation.

2. Three-dimensional solid mechanics and electrostatics coupling field simulation (CFS) models were developed and implemented to characterise the e-skins’ responses to a wide range of complex deformations (e.g., bending, twisting, elongation and their compound deformation). The results were further analysed and discussed for a better understanding of the behaviour of the e-skins.

3. Prototypes of SCASs were fabricated using silicone, liquid metal and carbon black with laser patterning. The linearity, sensitivity, repeatability, hysteresis and durability of the prototypes were evaluated through a long-period cycling experiment.

4. Tremendous annotated data were acquired by CFS models and experiments on physical soft robot platforms (a soft arm actuated by external forces and a pneumatic soft manipulator), which would be used for network training, validation and evaluation.

5. A neural architecture based on transformer was proposed to recover high-resolution morphology from SCAS signals. A neural architecture based on multi-layer perceptrons (MLPs) was proposed to detect physical contact with the interference of geometric variations. The proposed networks were implemented via Pytorch with annotated training samples and quantitatively evaluated through several error metrics.

1.4 Thesis outline

The thesis comprises six chapters and the remaining part is organised as follows.

Chapter 2 briefly reviews the literature on existing simulation methods for soft robots, proprioceptors based on different transducer mechanisms, learning-based proprioception methods and tactile sensing for soft robots.
Chapter 3 presents the coupling field simulation method to seamlessly integrate sensing and mechanical parts of a perceptive soft robot. The novel features of the CFS are demonstrated on a 16-electrode capacitive sensor array deployed on a soft robot arm (column). The case studies for two typical perception tasks are implemented to demonstrate their potential to benefit the development of learning-based perception.

Chapter 4 proposes a framework which comprises the stretchable capacitive e-skin (SCAS) and the capacitance-to-deformation transformer (C2DT), to achieve high proprioceptive geometry resolution (PGR) morphological reconstruction. The SCAS has a redundant planar electrode layout that forms a sequence of capacitors sensitive to deformation across the entire soft robot. The C2DT based on transformer architecture can explore the dependency over the SCAS signals and recover deformation from the signals. The superiority of the proposed framework is demonstrated on simulated and experimental datasets.

Chapter 5 investigates the feasibility of identifying physical contact with interference caused by deformation. A simplified version of the SCAS is designed to achieve this goal while reducing fabrication and deployment complexity. The feasibility of applying the simplified SCAS to achieve low PGR proprioception is also validated, which demonstrates its potential to simultaneously detect internal and external stimuli.

Chapter 6 concludes the innovation and major contributions of the thesis and discusses possible future directions with an emphasis on automated fabrication, multimodal sensing and sim-to-real transfer learning, which can further advance the techniques developed in this thesis and accelerate the adoption of soft robots in practical applications.
Chapter 2

Literature review

2.1 Introduction

This chapter reviews the literature on perception frameworks of soft robots from the perspectives of simulation and physical realisation. Existing sensors and signals processing methods for proprioception and exteroception are briefly summarised. Challenges yet not addressed by previous studies are also discussed.

2.2 Simulation methods for soft robots

Simulation is a standard tool for robot design, control and performance analysis. Numerous mature, thoroughly validated methods exist to build fast and reliable simulation models for traditional rigid robotic platforms [30, 31, 32]. However, when it comes to novel soft robotic systems, most existing models concern themselves with the dynamics, morphology and actuation of soft robots, largely neglecting the sensory systems embedded in the mechanical structure [33, 34, 35, 36].

For example, finite element method (FEM) was employed to simulate the deformations of soft robots and their surroundings in [37], which then helped achieve locomotion and manipulation for physical soft robots. Model order reduction [38, 39] could simplify the computation of FEM-based simulations through snapshot-proper orthogonal decomposition, thus allowing a real-time simulation of soft robots with higher complexity. Corucci et al. attempted to optimise geometric configurations of soft locomotion robots in simulated environments via evolutionary algorithms [40]. Shah et al. explored shapes and gaits that adapt to different environments for a soft robot using simulation and successfully transferred the result to a physical platform [41].
**Smart e-skins and machine learning for soft robot perception**

Unified and open-source frameworks can reduce the efforts and prior knowledge required by implementing soft robot simulation, thus accelerating the development of related technology. Simulation Open Framework Architecture (SOFA) [42] is an open-source library for interactive computational simulation and is frequently used in recent studies. For example, a software framework based on the continuum mechanical modelling of soft materials was proposed to simulate soft robots and their surroundings and implemented as a plugin for SOFA in [43]. Shape optimisation for soft robots in a given application scenario was realised on the basis of FEM simulation implemented in SOFA [44]. Combining SOFA with Matlab/Scilab, controllers for soft robots could be designed in a virtual environment, avoiding the inefficient trial-and-error process on physical robotic platforms [45].

Recently, soft sensor design and placement optimisation using simulation methods have been reported. In [46], a neural architecture for co-learning of general robotic tasks (e.g., tactile sensing and proprioception) and sensor placement was developed and demonstrated in a simulated environment. However, instead of simulating sensor behaviour under robot deformation and actuation, the study assumed the sensor has perfect performance (e.g., 100% accuracy, no latency, deployable to every location on the robot body), which is far from the practical conditions. In [47], soft resistive sensors (silicone tubes filled with liquid metal) deployed on a soft arm were modelled. However, the deformations and sensors demonstrated in this work are relatively simple. More complex situations should be considered and further investigation is needed.

Despite the recent advances in soft robot simulation, new simulation approaches that seamlessly integrate mechanical and sensing components are needed (i.e., coupling field simulation, CFS) and are expected to play a critical role in developing perception systems for soft robots. It can benefit the following aspects.

- The feasibility of proposed perception methods can be verified in a fast and cost-effective manner through CFS compared with implementing experiments on physical platforms.
- The response of proposed sensors on various soft robots with different deformations and external stimuli can be characterised using CFS, which assists in understanding sensor behaviour, quantitatively evaluating sensor performance and optimising sensor design.
Deep learning approaches are frequently employed to interpret sensing data to target parameters (e.g. deformation class, applied force and coordinates of the end-effector) due to the complexity of mathematically modelling the behaviour of soft sensors and robots. It normally requires a large number of labelled data as training samples to optimise network parameters. However, data annotation is extremely time- and labour-consuming. Applying data generated by CFS to pre-train the network can significantly reduce the burden of data acquisition. Approaches such as sim-to-real transfer learning can bridge the gap between real and synthetic data [48, 49, 50].

2.3 Soft robot proprioception

This section first reviews existing sensors based on different mechanisms for soft robot proprioception. Due to the complexity of modelling soft bodies and establishing explicit relationships between sensing data and target parameters, machine learning is frequently employed to process sensing data. Related studies are also discussed in this section. Finally, the state of the art is reviewed from the perspective of proprioceptive geometry resolution.

2.3.1 Sensors based on different transducer mechanisms

Soft sensors based on a range of transducer mechanisms have been developed to perform simplified proprioception tasks over the past decades. Resistive [51], capacitive [52] and optical sensors [53] are most frequently used due to their respective advantages. Besides these, other types of sensors such as magnetic [54, 55], piezoelectric [56, 57] and triboelectric [58, 59] sensors have been reported in previous studies as well.

Resistive sensors

The resistance of a conductor is governed by

\[ R = \frac{\rho l}{A} \]  

(2.1)
where $\rho$ is the resistivity of the material; $l$ is the length of the material and $A$ is the cross-sectional area of the material. The geometric variations lead to changes in resistance. Resistive sensors normally quantify target geometric parameters via resistance measurements. For example, soft resistive sensors based on liquid metal were developed to measure strain in [60], the curvature of an underwater soft fish in [61], and the bending angle of a soft hand in [62]. The advantages of liquid metals as materials of soft proprioceptors lie in their flowable nature and high conductivity. However, they are only applicable in a limited temperature range and have a much higher density than most elastomeric wrapping. Ionic liquids are also frequently used in the area of soft sensors [63], which are lightweight, available in a wider temperature range, but poorly conductive.

Piezoresistive sensors [64] belong to the category of resistive sensors. Different from sensors based on conductive liquids which have constant resistivity during deformation, the resistivity of piezoresistive sensors (normally based on elastomeric composites filled with conductive fillers [17, 65]) changes in response to applied pressure. The resistance measurements reflect not only geometric variations but also resistivity changes. Piezoresistive sensors have been deployed on different soft robotic platforms to realise simple proprioception (e.g., curvature measurement for a soft pneumatic actuator in [66] and tip tracking for a soft finger in [25]). Large hysteresis, slow responses and non-linearity are considered as the main drawbacks of this type of sensors [65].

**Capacitive sensors**

A soft capacitive sensor is normally formed by two parallel stretchable electrodes and a soft dielectric inside. The capacitance of the capacitive sensor is governed by

$$C = \varepsilon_0 \varepsilon_r \frac{S}{d}$$  

where $\varepsilon_0$ is the permittivity of vacuum; $\varepsilon_r$ is the relative permittivity of the filling material; $S$ is the area of the electrode and $d$ is the distance between two electrodes. The geometric variations lead to changes in capacitance. Capacitive sensors can quantify target geometric parameters via capacitance measurements. Stretchable electrodes are the foundation of soft capacitive sensors and are usually made of soft conductive materials such as carbon black dispersed silicone [67] and conductive fabric [68]. The applications of this type of sensors as simple proprioceptors include but are not limited
to curvature measurement for a soft robotic joint in [69], discrete pose estimation of a soft prosthetic hand in [70] and inflation state measurement of a pneumatic soft robot in [71]. Capacitive sensors have better linearity and sensitivity, ignorable hysteresis and fast response speed compared with their resistive counterparts. However, they are subject to environmental interference. The proximity of conductive objects can cause significant errors in capacitance readouts.

**Optical sensors**

Soft optical sensors measure morphological variations that occur at light transmission medium through changes of light intensity, frequency, wavelength and/or phase [72, 73, 74, 75]. These kinds of sensors show excellent flexibility and anti-interference but require complex electronics and have limited stretchability. Previous studies have explored methods for the achievement of preliminary proprioception with optical sensors on a range of soft robots. For example, an optical fibre sensor was employed to estimate the pose of a continuum robotic arm in [76]. The bending angle of a bidirectional soft actuator was measured through an optical waveguide sensor in [77]. The curvature of a silicone octopus tentacle was predicted based on fibre bragg gratings technology in [78].

### 2.3.2 Sensing data interpretation with machine learning tools

Establishing explicit relationships between sensing data and target proprioceptive parameters (such as bending angle and deformation category) is not trivial, especially for complex proprioception tasks like high-resolution morphological reconstruction that needs to deploy an abundant number of sensors in the unit area to collect sufficient information. Machine learning [79] is a potential solution to complex sensing data interpretation problem, which have demonstrated remarkable performance in the fields like computer vision [80, 81, 82] and natural language process [83, 84, 85] which involve in tremendous complicated and structural data.

However, the applications of learning tools to soft robot proprioception are in nascent, only involving very basic and simple neural architectures [90]. For example, support vector machine (SVM) [91], k-nearest neighbours (kNN) [92] and decision trees (DT) [93] were employed to classify deformation, and multi-layer perceptron (MLP) [94], Gaussian process regression (GP) [95], linear regression (LR), kNN, DT and SVM were applied to estimate bending and twisting angles in [1]. Scharff et al...
Table 2.1: Machine learning methods for soft robot proprioception

<table>
<thead>
<tr>
<th>Reference</th>
<th>Proprioception task</th>
<th>Machine learning method</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1]</td>
<td>bending &amp; twisting angles estimation</td>
<td>MLP/kNN/SVM/DT/GP/LR</td>
</tr>
<tr>
<td>[86]</td>
<td>surface parameters estimation</td>
<td>MLP/LSTM/SVR/MVLR</td>
</tr>
<tr>
<td>[25]</td>
<td>tip tracking</td>
<td>LSTM</td>
</tr>
<tr>
<td>[87]</td>
<td>shape parameters estimation</td>
<td>MLP/LSTM</td>
</tr>
<tr>
<td>[88]</td>
<td>coordinates estimation for visual markers</td>
<td>MLP</td>
</tr>
<tr>
<td>[89]</td>
<td>coordinates estimation for visual markers</td>
<td>MLP</td>
</tr>
<tr>
<td>[62]</td>
<td>deformation parameters estimation</td>
<td>MLP</td>
</tr>
</tbody>
</table>

compared the performance of different learning methods including long short-term memory networks [96] (LSTM) which is one kind of recurrent neural networks (RNNs) [97], MLP, support vector regression (SVR) [98] and multivariate linear regression (MVLR) [99] on surface parameter estimation of a pneumatic membrane in [86]. Thuruthel et al applied LSTM that can exploit temporal dependency to compensate for hysteresis of piezoresistive sensors and estimate real-time 3D coordinates of the tip on a soft arm [25]. Wall et al utilised the polynomial regression (PR) [100] method to translate sensing data to desired proprioceptive information (e.g., twisting angle). More related work is summarised in Table 2.1.

Existing methods can hardly meet the requirements for many soft robot applications as they only focus on oversimplified proprioception tasks (e.g., bending angle estimation [1, 24] and tip tracking [25] under prescribed deformations) that do not match the practical situations. Innovative applications of advanced machine learning algorithms to complex but important proprioception tasks such as high-resolution morphological reconstruction need further exploration.
Table 2.2: Existing soft proprioception systems

<table>
<thead>
<tr>
<th>Reference</th>
<th>Sensor</th>
<th>Applicable range</th>
<th>PGR</th>
</tr>
</thead>
<tbody>
<tr>
<td>[101]</td>
<td>optical</td>
<td>◇ unidirectional bending, ◇ unidirectional elongation.</td>
<td>2</td>
</tr>
<tr>
<td>[1]</td>
<td>optical fibres</td>
<td>◇ bidirectional bending, ◇ twisting around a fixed axis, ◇ their compound deformations.</td>
<td>2</td>
</tr>
<tr>
<td>[25]</td>
<td>cPDMS</td>
<td>◇ unidirectional bending with contact force.</td>
<td>3</td>
</tr>
<tr>
<td>[87]</td>
<td>piezoresistive silicone sensors</td>
<td>◇ omnidirectional bending.</td>
<td>9</td>
</tr>
<tr>
<td>[102]</td>
<td>capacitive and pneumatic sensing</td>
<td>◇ omnidirectional bending with contact force.</td>
<td>7</td>
</tr>
<tr>
<td>[103]</td>
<td>piezoresistive hydrogel sensors</td>
<td>◇ unidirectional bending, ◇ bidirectional elongation.</td>
<td>2</td>
</tr>
<tr>
<td>[86]</td>
<td>optical</td>
<td>◇ local inflation of a deformable membrane.</td>
<td>147</td>
</tr>
</tbody>
</table>

2.3.3 Proprioceptive geometry resolution

The highly deformable feature gives infinite degrees of freedom to a soft body. It is infeasible to describe completely the 3D morphology of a soft system with only a limited set of parameters. The number of independent parameters used by a soft proprioception system to describe the body geometry determines the smallest size of geometric variations that can theoretically be detected and presented by the system. Generally, the greater the number of independent parameters, the finer and more accurate the geometric variations can be described. Therefore, the number of such independent parameters is defined as the proprioceptive geometry resolution (PGR). Soft proprioception systems with higher PGR are desirable for soft robotics as they can endow soft systems with more comparable bodily awareness to rigid robots, thus
enabling more natural interaction with humans (e.g., real-time 3D geometry of a soft robot can be visually observed without the line of sight, allowing users to operate robots intuitively even in occlusion environments) and underpinning precise closed-loop control.

To the best of my knowledge, there is no off-the-shelf high PGR soft proprioception system (see Table 2.2). Previous studies are focused on low PGR proprioception, limiting their capability to preserve geometric details (e.g., local deformation and surface information) and their usage in practical application scenarios [104, 105]. For example, combined with the mathematical model of the soft robot under investigation, an optical fibre-based proprioception system can successfully reconstruct the 3D geometry based on two parameters [1], i.e., global bending and twisting angles (PGR=2). However, the low PGR fails to describe local geometric variations, e.g., bending of a robot segment (see Fig. 2.1 and 2.2) and is only applicable to a fixed bending direction and twisting axis (see Fig. 2.3 and 2.4).

Some recent studies attempted to build soft proprioception systems with higher PGR by optimising sensor design, introducing advanced machine learning algorithms (e.g., long short-term memory networks [96], LSTM) and employing 3D motion capture devices (e.g., tracking cameras [87]). Redundant polydimethylsiloxane (PDMS) im-

---

**Figure 2.1:** Conceptual illustration of the advantage of a high PGR proprioception system over a low PGR (=2) proprioception system [1] in previous research when a segment of the object bends.
pregnated with conductive carbon nanotubes (cPDMS) sensors with LSTM can estimate 3D coordinates of a soft fingertip (PGR=3) [25]. The simplified 3D geometry (described by 9 parameters) of a trunk-shaped soft robot can be recovered through 12 conductive silicone-based piezoresistive sensors distributed on the robot body (PGR=9) [87]. The 3D deformation of a 4-chamber pneumatic membrane (described by 49 visual markers) was reconstructed using LSTM and integrated optical sensors (PGR=147) [86]. Despite these recent advances, obtaining high PGR across a wide range of complex deformations remains unrealised.

2.4 Tactile sensing for soft robots

Tactile sensing is a prerequisite to enhanced autonomy and trustworthy human-robot interaction. Flexible e-skins are one of the potential solutions that have attracted the most attention. Their highly deformable nature not only improves the safety during human-robot interaction, but also allows tactile sensors to be deployed on irregular rigid robotic surfaces and compatible with soft robotic platforms [106].
Previous studies of tactile e-skin mainly concentrate on conventional rigid robots. For example, Yan et al proposed soft magnetic skin for super-resolution tactile sensing with the functionality of force self-decoupling and demonstrated its excellent performance on a rigid robot claw [107]. Electrical impedance tomography (EIT) [108] can estimate the conductivity distribution within the region of interest by using only a finite set of boundary electrodes. The sparsity of electrodes is a favourable feature for large-area tactile sensing in the field of rigid robots and related investigations have been reported extensively [109, 110, 111].

The above methods can not be directly applied to the area of soft robot tactile sensing as they do not consider the effect of deformation induced by actuation on the sensor readouts. Many studies aiming at soft robots overlook this essential factor as well, only evaluating sensors with static experiments that don’t involve dynamic deformation [112]. This causes the state of the art for soft robot tactile sensing is still far from an acceptable level.
Figure 2.4: Conceptual illustration of the advantage of a high PGR proprioception system over a low PGR (=2) proprioception system [1] in previous research when the twisting axis is not fixed.

Outside the area of soft robotics, flexible tactile sensors insensitive to bending have been demonstrated in skin-interfaced wearable devices through the adoption of porous structures that can accommodate bending-induced deformations [113]. For example, a nanofibre-based resistive sensor was designed to detect only normal pressure and not sensitive to bending deformation in [114]. Ultrawide-range, flexible, bending-insensitive sensors on the basis of a carbon nanotube network-coated elastomer sponge were also reported in [115]. However, soft robots have stronger deformability than human skin and can exhibit a range of more complex deformations, such as twisting, elongation and their compound deformation. Bending-insensitive tactile sensors are insufficient for soft robots and further exploration is still needed.

2.5 Summary

This chapter briefly reviewed the literature related to simulation, proprioception sensors and algorithms, and tactile sensing for emerging soft robots. The advantages and disadvantages of existing methods are discussed. The purpose of this chapter is to provide an introduction to the state of the art and identify open challenges that need to be further investigated. Coupling field simulation seamlessly integrating mechanical
Smart e-skins and machine learning for soft robot perception

and sensing components, high PGR proprioception offering real-time and accurate
global and local geometric feedback and tactile sensing with geometric interference
are three open issues on which the thesis focus. The innovative work and scientific
contributions will be subsequently presented in the following chapters.
Chapter 3

Coupling field simulation for soft robot perception

3.1 Introduction

Simulation plays a critical role in designing, analysing and controlling intelligent robots. A wealth of thoroughly validated approaches exist to establish reliable models for rigid robots [116], but these hardly lend themselves to their soft robotics counterparts. This is due to the highly deformable nature of soft materials that dramatically increases the degrees of freedom of the robot body, thus exacerbating the complexity of the simulation [33]. Recently, many attempts related to soft robot simulation have been reported [117, 118]. However, most of them only concern the dynamics, morphology and actuation of a soft robot, neglecting the sensory systems embedded in the mechanical structure. While sensing in traditional robots already provides reliable and accurate measurements of essential parameters for the system characterisation and control, sensing technology for soft systems lags behind, hindered by yet unsolved challenges in sensor design, signal processing and interpretation. Simultaneous simulation of soft sensors and actuators can help to understand the response and performance of sensor systems during actuation, optimise sensor design, fast and cost-effectively verify the feasibility of designed sensors and generate abundant annotated data as training materials for learning-based perception methods, thus accelerating the development of soft perception systems.

In this chapter, coupling field simulation (CFS) models for a 16-electrode capacitive sensor array deployed on a soft manipulator are presented and implemented. The solid mechanics field simulation is used to compute the deformation caused by an external force load, while the sensor readouts corresponding to the deformation are
obtained through the electrostatics field simulation. The proposed CFS models can generate annotated data less costly and effectively, benefiting the development of learning algorithms for soft robot perception. The proposed CFS framework is shown in Fig. 3.1.

3.2 Solid mechanics and electrostatics coupling field simulation

3.2.1 Principle of capacitive sensor array

A planar capacitive sensor array consisting of several electrodes (e.g., 8 or 12) deployed on the outer surface of the region of interest (ROI) can record the capacitance formed by different electrode combinations. In the non-deformable setup, the permittivity distribution within the ROI can be inferred through the capacitance measurements by solving a dedicated inverse problem [119]. This technique is known as electrical capacitance tomography (ECT) and is frequently applied in industrial processes to non-invasively monitor the dynamic behaviours of multiphase flows. The relationship between the capacitance and its influencing factors can be described as...
\[ C = \frac{Q}{V} = -\frac{1}{V} \int \varepsilon(x,y,z) \nabla \phi(x,y,z) d\Gamma \quad (3.1) \]

where \( V \) is the potential difference between two electrodes that constitute the capacitor; \( \varepsilon(x,y,z) \) denotes the permittivity distribution in the ROI; \( \phi(x,y,z) \) represents the potential distribution and \( \Gamma \) is the area of the electrode surface.

Conventional ECT utilizes rigid electrodes and aims to reconstruct the permittivity distribution based on a series of capacitance measurements. Here we consider capacitive sensor arrays made of soft materials. In this case, the capacitance is not only determined by the electrical properties of the internal medium but also affected by the geometries of the objects under investigation. Notably, the target manipulator has a constant permittivity while its 3D domain deforms. Therefore, the capacitance variation primarily reflects the geometric variation in the proximity of the electrode pair, enabling infer the boundary deformation through capacitance readouts.

In the case of parallel plate capacitors with homogeneous material inside, the determining equation for capacitance can be simplified to Eq. 2.2. The capacitance is then positively correlated with the area of electrodes \( S \) and negatively correlated with the distance between two electrodes \( d \). This is an intuitive observation that can assist qualitative understanding of the sensor response to deformation in the latter discussion.

### 3.2.2 Dynamics of the testbed

The solid mechanics and electrostatics fields are coupled to simultaneously simulate the dynamic deformation and capacitive sensor response of a square soft manipulator. To simulate the deformation, an external force is exerted at the endpoint of the manipulator. The relationship between stress and strain of an elastomer can be characterised by Hooke’s law [121]

\[ \xi = \frac{1 + \nu}{E} \sigma - \frac{\nu}{E} \text{tr}(\sigma) I \quad (3.2) \]

where \( \xi \) is the strain tensor; \( \sigma \) is the stress tensor; \( \nu \) denotes the Poisson’s ration; \( E \) is the Young’s modulus; \( \text{tr}(\cdot) \) represents trace operator and \( I \) represents second-order identity tensor.
Geometric structure of the mock-up robot arm and the 16-electrode capacitive sensor

- Silicone-based robot arm
- Planar electrode array to form the capacitive sensor
- Force applied on the bottom to cause bending deformation

**Figure 3.2:** Schematic illustration of the soft robot manipulator. Note that the other 8 electrodes are deployed on two hidden surfaces.

The deformation of the manipulator is inferred through Eq. (3.2), and then Eq. (3.1) is applied to calculate capacitances formed by a series of electrode pairs following the conventional 3D ECT sensing protocol under the deformation.

### 3.2.3 Coupling field simulation setup

The soft robot manipulator is made of silicone (see Fig. 3.2 for the geometric structure). For computational simplicity, the robot arm is set as a square cylinder with the size of $100 \times 100 \times 1000$ mm. It is actuated by external forces, which allows to generate complex deformations such as the compound deformation of elongation and twisting. Sixteen planar electrodes are uniformly distributed on the surface of the manipulator to form the capacitive sensor array (4 layers and each consists of 4 electrodes). Each electrode is a $105 \times 30$ mm surface. The material parameters are set as Table 3.1.

**Table 3.1:** Material parameters

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>$1.28 \times 10^3$ kg m$^{-3}$</td>
</tr>
<tr>
<td>Relative permittivity</td>
<td>3</td>
</tr>
<tr>
<td>Poisson’s ratio</td>
<td>0.022</td>
</tr>
<tr>
<td>Young’s modulus</td>
<td>4.15 MPa</td>
</tr>
</tbody>
</table>
Figure 3.3: Examples of deformations (bending) in several selected frames and the calibrated capacitance readouts for all electrode pairs. Lines with different colors indicate different capacitance readouts.

Three episodes of pseudo-continuous deformations with different external loads are simulated using solid mechanics and electrostatics CFS in COMSOL Multiphysics®. Each episode contains \( n = 40 \) discrete time frames. The stationary solutions in each time frame are computed to mimic the continuous deformation by gradually increasing the magnitude of the external loads applied to the manipulator. The simulation focuses on the relationships between sensor signals and deformations in quasi-static conditions rather than in actual dynamics. This significantly simplifies the construction of simulation models and minimises computational effort without affecting sensor characterisation.

The first episode models a bending deformation (see Fig. 3.3 top half), which is achieved by applying a force load in the x-y plane to the bottom of the structure, regarded here as the end-effector of the manipulator. The top is computationally treated as a fixed boundary and physically regarded as the base frame of the robotic artefact (the same settings are also adopted in the following episodes). The increase rate for the force is \((-0.613, -3.953, 0)\), i.e., the force in time frame \( t \) is \((-0.613t, -3.953t, 0)\) N. Note that the increase rate is generated randomly and the value can be modified to produce different bending deformations.
Figure 3.4: Examples of deformations (two-stage twisting and bending) in several selected frames and the calibrated capacitance readouts for all electrode pairs.

The second episode includes 2 stages (see Fig. 3.4 top half). During the first stage (the first 13 time frames), the manipulator is twisted with the axis of \((-0.194, 0.004, -1)\). The increase rate for the twisting angle is \(-6.846\), i.e., the twisting in time frame \(t\) is \(-6.846t^\circ\). During the second stage, a force load is added with the increase rate of \((-1.667, 3.636, 0)\), i.e., the force in time frame \(t\) is \((-1.667(t - 13), 3.636(t - 13), 0)\) N, while the twisting component remains constant.

The third episode models the compound deformation of twisting and elongation (see Fig. 3.5 top half), which is achieved by simultaneously applying a rotation and a displacement along the z-axis to the bottom of the manipulator. The rotation axis is \((0.048, -0.069, -1)\). The increase rates for the rotation angle and the displacement are \(-2\) and \((0, 0, -6)\), i.e., the rotation angle and displacement in time frame \(t\) are \(-2t^\circ\) and \((0, 0, -6t)\) mm respectively.

From the measurement perspective, any two electrodes can form a capacitor and generate a capacitance readout. The 16-electrode capacitive sensor can theoretically produce 120 readouts per measurement frame. However, many of them are too small to be robustly measured in the physical world. To better approximate the practical
**Figure 3.5**: Examples of deformations (elongation and twisting) in several selected frames and the calibrated capacitance readouts for all electrode pairs.

In all conditions, only the capacitance readouts generated by two electrodes in the same layer are recorded. Each layer includes 4 electrodes and can form 6 valid independent capacitance readouts per frame. In total, the 16-electrode sensor has 24 readouts per frame.
3.3 Simulation results

3.3.1 Sensor response under bending

![Figure 3.6: The calibrated capacitance readouts for 4 selected electrode pairs during the whole deformation process (bending).](image)

The 24 capacitance readouts of the soft arm under the bending deformation over time are shown in Fig. 3.3 bottom half. The calibrated capacitance is obtained using:

\[ c = \frac{c_{\text{raw}} - c_{\text{nol}}}{c_{\text{nol}}} \]  

(3.3)

where \( c \) is the calibrated capacitance; \( c_{\text{raw}} \) is the raw measurement and \( c_{\text{nol}} \) is the measurement without any loads. The signals from the capacitive sensor array change with the deformation, demonstrating the feasibility of applying it to soft robot perception.
**Figure 3.7:** Calibrated capacitance curves over time of bending deformations in two different directions.

Fig. 3.6 illustrates calibrated capacitance readouts of 4 selected electrode pairs during the whole bending process. It can be observed from the figure that some capacitance readouts increase as the amplitude of the bending grows, while others have the opposite trend. This is because bending deformation causes one part of the soft body to expand (electrodes deployed in the area will also expand) while the other part contracts (electrodes deployed in the area will also contract). The locations of expansion and contraction depend on the bending direction. Fig. 3.7 shows the calibrated capacitance readout of the same electrode pair exhibits distinct responses to bending deformations in different directions. This property enables the proposed capacitive sensor array to measure bending direction.
3.3.2 Sensor response under two-stage twisting and bending

Capacitance readouts of the two-stage twisting and bending deformation are illustrated in Fig. 3.4 bottom half and Fig. 3.8 illustrates calibrated capacitance readouts of 4 selected electrode pairs during the whole deformation process. In the first stage (the first 13 time frames), twisting is implemented. The magnitude of the sensor response to twisting is smaller compared with that of bending, which results in the bending deformation dominating the capacitance readouts at the second stage. Several capacitance signals flip at the 14th time frame due to the introduction of bending (e.g. Fig. 3.8 left half).

**Figure 3.8:** The calibrated capacitance readouts for 4 selected electrode pairs during the whole deformation process (two-stage twisting and bending).
3.3.3 Sensor response under elongation and twisting

**Figure 3.9:** The calibrated capacitance readouts for 4 selected electrode pairs during the whole deformation process (elongation and twisting).

Capacitance readouts of the compound deformation of twisting and elongation are shown in Fig. 3.5 bottom half and Fig. 3.9 illustrates calibrated capacitance readouts of 4 selected electrode pairs during the whole deformation process. The capacitance readouts monotonically increase with the degree of deformation. This is readily justified by the elongation deformation dominating the response of the capacitive sensor. The area of individual planar electrodes grows during elongation, leading to the increase in capacitance readouts.
3.4 Applications of CFS in soft robot perception

The proposed CFS models that seamlessly integrate robot mechanical component and sensor response can generate many annotated data, providing samples to train neural networks for various perception tasks including proprioception (e.g., deformation classification) and exteroception (e.g. applied force estimation). The networks trained with simulated data have the potential to be transferred to practical applications using sim-to-real transfer learning approaches, reducing the burden of data acquisition in the physical world. This subsection implements two typical perception tasks (i.e., applied force estimation and deformation classification) as case studies to demonstrate the potential of CFS.

3.4.1 Applied force estimation

Deformation induced by applied force can lead to the change in capacitance measurements, thus making it feasible to estimate the magnitude of force through capacitance readouts. The CFS model for the soft manipulator subject to bending is implemented to generate annotated data. The bending deformation is induced by a force load \( f = (v_x t, v_y t, 0) \) applied to the tip of the manipulator. The parameters \( v_x \) and \( v_y \) are varied in each episode to ensure the diversity of the synthesis dataset. In total, we generate 300 episodes of data (12,000 time frames; with each episode containing 40 frames).

The goal of this subsection is not to develop a novel algorithm with superiority for applied force estimation. Instead, it aims to verify the potential of CFS as a tool to analyse sensor and algorithm performance at the design stage and generate annotated data to benefit the development of learning-based soft robot perception. Therefore, a simple multi-layer perceptron (MLP) [122] is employed as the force estimator.

The MLP has one hidden layer with 128 neurons. The input of the MLP is the 24-dimensional measurement vector from the capacitive sensor. The output is the estimation of the 2-dimensional force vector, i.e., \( (\hat{f}_x, \hat{f}_y) \) (the component in the z direction is 0). The activation functions for the hidden and output layers are ReLU and Linear, respectively. The data are randomly divided into three exclusive groups, i.e., the training set (172 episodes, 6,880 frames), the validation set (74 episodes, 2,960 frames) and the testing set (54 episodes, 2,160 frames).
The network is implemented in the PyTorch platform. The mean square error (MSE) between prediction and ground truth is selected as the loss of the force estimator. The Adam[123] optimiser is employed to update the learnable parameters and minimise the loss. The initial learning rate is set as 0.0005, which decays every 15 epochs by a factor of 1.2. The gradient is clipped with the threshold of 0.5. The network is trained with the training set for 200 epochs and the batch size is set to 256. After training, the network with the least validation loss is retained as the final model. Two error metrics are employed here to quantitatively evaluate the performance of the trained network, i.e., MSE and mean absolute error (MAE), in this case, defined as:

\[
\text{MSE} = \frac{1}{2N} \sum_{i=1}^{N} \left[ (\hat{f}_x^i - f_x^i)^2 + (\hat{f}_y^i - f_y^i)^2 \right] \tag{3.4}
\]

\[
\text{MAE} = \frac{1}{2N} \sum_{i=1}^{N} \left( |\hat{f}_x^i - f_x^i| + |\hat{f}_y^i - f_y^i| \right) \tag{3.5}
\]

where \( N \) is the number of samples and superscript \( i \) represents the \( i^{th} \) sample.

**Figure 3.10:** Absolute error for force estimation on the testing set. Left: absolute error for force component along the x axis. Right: absolute error for force component along the y axis.
The trained MLP can achieve 1.373 for MAE and 3.589 for MSE on the testing set. The estimation results for $f_x$ and $f_y$ on the testing set are shown in Fig.3.10. It shows that samples distributed in the middle have smaller errors. A likely explanation for this lies in the unbalanced distribution of training samples and the limited expressive power of the MLP when associated with only one hidden layer. Fig.3.11 shows two examples of the ground truth deformations, the corresponding applied force and the estimation of the MLP.

Figure 3.11: Examples of force estimation on the testing set. Top: the ground truth deformations. Bottom: the ground truth applied force (red dot line) and the estimated force (blue line).
3.4.2 Deformation classification

The capacitance signals induced by different types of deformations have distinct patterns. This property enables the proposed sensor array to be applied to classify deformations. A binary classifier is trained to tell the difference between bending (the first type of deformations described in 3.2.3) and the two-stage twisting and bending deformation (the second type of deformations described in 3.2.3). Three hundred episodes of bending data are already obtained in the applied force estimation task. The CFS model for the soft manipulator subject to the two-stage deformation is implemented to generate annotated data. The diversity of the synthesis dataset is ensured by varying the values of external loads, including the axis of twisting, the increase rate for twisting angle and the increase rate for the applied force. Complementary with the dataset produced for the force estimation task, 598 episodes of data are acquired.

The class label of bending is set as 0. The class label of two-stage twisting and bending deformation is set as 1. Only the data after the 14th frame in each episode are retained to ensure the difference between deformations is sufficiently large to be detected. Finally, a dataset is constructed with 15,548 frames of samples, among which 7,748 frames belong to deformation 1 (the combination of bending and twisting) and the remaining 7,800 frames are deformation 0 (bending).

An MLP with one hidden layer is employed as the classifier to demonstrate the potential of CFS methods. Its structure is the same as the force estimator except for the dimension (1) and the activation function (Sigmoid) of the output layer. The capacitance vector is fed to the MLP classifier and expected to output the prediction of the class label (i.e., the probability that the deformation contains a twisting). The dataset is randomly divided into three exclusive groups, i.e., the training set (343 episodes, 8,918 frames), the validation set (147 episodes, 3,822 frames) and the testing set (108 episodes, 2,808 frames). The binary cross entropy loss function is selected and the classifier is trained using the same training procedure as the applied force estimation task (the only difference is that the initial learning rate is set as 0.0001).

After training, the classifier can achieve 100% classification accuracy on the testing set, demonstrating the feasibility of using capacitance signals to distinguish these two types of deformations. Two examples of deformation classification are shown in Fig.3.12. The top line of the figure shows a bending deformation (left) and a two-stage twisting and bending deformation (right). The middle line illustrates the capacitance readouts corresponding to the two deformations. The bottom line gives the classifier output, i.e., the probability that the deformation contains a twisting. For the bending
Figure 3.12: Examples of deformation classification on the testing set. Top: the ground truth deformations. Middle: the corresponding measurements (the input of the MLP classifier). Bottom: the output of the MLP classifier.

deformation (left), the classifier predicts that it has a 0.5% probability of containing a twisting component. For the two-stage deformation (left), the classifier predicts that it has a 99.3% probability of containing a twisting component. The results of the deformation classification for the two samples are correct.

Note that the 100% accuracy is only achievable under ideal circumstances. In this study, the classification task is very simple, which only includes 2 deformation classes, and the signals are assumed to be noise free. The accuracy is expected to drop if the task involves more types of deformations and/or the noise is taken into account.
3.5 Summary

In this chapter, the CFS model for a 16-electrode capacitive sensor array on a soft robot manipulator was developed to seamlessly integrate the sensing and mechanical components. The responses of the capacitive sensor array to various deformations (including bending, two-stage twisting and bending and the compound deformation of twisting and elongation) were characterised through the CFS model, which helps to establish a better understanding of sensor behaviours. The case studies for two perception tasks (i.e., applied force estimation and deformation classification) based on the annotated dataset produced through CFS demonstrate the potential of CFS to benefit learning-based perception. Combined with simple machine learning frameworks, the capacitance sensor array performed well in both the applied force estimation (1.373 MAE) and deformation classification tasks (100% accuracy). The models trained with simulated data could be transferred to practical applications using sim-to-real transfer learning, significantly reducing time and costs for data acquisition through physical experiment platforms. The proposed CFS will be used in the next chapter, which helps to quantitatively evaluate the proposed proprioception methods in a virtual environment, thus optimising sensor and learning algorithms design.
Chapter 4

High PGR morphological reconstruction

4.1 Introduction

Many robotic tasks require knowledge of the exact 3D robot geometry. However, this remains extremely challenging in soft robotics because of the infinite degrees of freedom of soft bodies deriving from their continuum characteristics. The number of independent parameters used by a soft proprioception system to describe the body geometry determines the smallest size of geometric variations that can theoretically be detected and presented by the system. Generally, the greater the number of independent parameters, the finer and more accurate the geometric variations can be described. Therefore, the number of such independent parameters is defined as the proprioceptive geometry resolution (PGR). Previous studies have achieved only low PGR proprioception, thus suffering from loss of geometric details (e.g., local deformation and surface information) and limited applicability (e.g., only applicable to prescribed simple deformations). Soft proprioception systems with higher PGR are desirable as they can endow soft systems with bodily awareness more comparable to that of rigid robots, paving the way towards the realisation of their autonomy.

In this chapter, a high PGR (3,900) proprioception system is proposed to confer full-geometry, millimetre-level bodily awareness to soft robots. The proprioception system encapsulates an intrinsically stretchable capacitive e-skin (SCAS) and a purpose-designed neural architecture (i.e., the capacitance-to-deformation transformer, C2DT). The SCAS has four different functional layers and employs a redundant planar skin electrode layout that forms a sequence of capacitors sensitive to deformations across distal and proximal locations, allowing it to detect geometric variations across the entire soft body (see Fig. 4.1). The C2DT based on self-attention mechanism [83]
explores the dependency over the e-skin signals and directly translates the measurements to the point cloud of the morphology. The synergistic combination of the SCAS and C2DT can achieve accurate and high PGR 3D shape reconstruction under complex deformations, which is one or two orders of magnitude improvement over previous methods (for comparison, see Table 2.2). The proposed system does not require mathematical modelling of the robot under investigation. Therefore, it theoretically should be agnostic to the shape of the soft body, and has the potential to be extended to soft robotic platforms with unprescribed morphology. This high PGR proprioception capability can assist in solving the most fundamental challenges in soft robotics, such as precise closed-loop control in complex tasks, thereby facilitating their widespread adoption.
4.2 E-skin design based on CFS

4.2.1 CFS model of the e-skin

![CFS model of a 64-electrode SCAS deployed on a soft arm with 64 visual markers.](image)

**Figure 4.2:** The CFS model of a 64-electrode SCAS deployed on a soft arm with 64 visual markers.

Different from conventional parallel capacitive sensors frequently used in many previous studies [52], the design of SCAS is inspired by 3D electrical capacitance tomography (ECT) sensor and its sensing strategy [124]. 3D ECT has demonstrated that the capacitance readout of a boundary electrode pair is related to the permittivity of the medium within the sensitive region, and its geometry. In soft robot proprioception, the permittivity remains constant. The change of capacitance primarily reflects geometric variations and, therefore, can be used to infer local and global deformations.

The coupling field simulation is implemented in COMSOL Multiphysics to simultaneously generate virtual SCAS sensing data and deformation data to demonstrate the effectiveness of the proposed method. The object of study is a square soft robot arm made of silicone (length: 100 mm, width: 100 mm, height: 1000 mm, see Fig. 4.2). An array of 64 electrodes (8 × 8) is placed on the surface of the robot arm to form a 64-electrode SCAS. For simplicity, each electrode is set as a 105 × 30 mm flat surface without thickness. The distance between two adjacent electrodes on the same side is 20 mm both horizontally and vertically. The distance between each edge and the nearest electrode is 10 mm. Relevant material properties are set as follows: Young’s modulus $E = 4.15$ MPa, Poisson’s ratio $\nu = 0.022$, density $\rho = 1.28 \times 10^3$ kg m$^{-3}$, relative permittivity $\varepsilon_r = 3$. 

37
**Smart e-skins and machine learning for soft robot perception**

Note that the 64-electrode layout has multiple subsets of sparse electrode layouts (e.g., the 16-electrode layout in 3.2). The simulation results of all the subsets are known once the CFS results of the dense layout are computed. The choice of the 64-electrode layout is made to facilitate the investigation of the impact of the density of electrodes on the proprioception performance.

In addition, it is impractical to ascertain the exact point-to-point correspondences of all points between two different deformations in the physical world. A scheme that can be realistically implemented is resorted. Sixty-four points are selected as visual markers whose correspondences are available during network training and the correspondences of the remaining points are only used in testing for evaluation.

### 4.2.2 Sensing strategy of the e-skin

Theoretically, any two electrodes can form a capacitor. The SCAS with 64 electrodes can produce 2,016 independent capacitance readouts in each measurement frame (select 2 electrodes to form a capacitor, i.e., $C_{64}^2 = 2,016$). However, many of them are extremely small and cannot be reliably measured in the real world. Therefore, only capacitances of electrode pairs in the same layer and capacitances of certain electrode pairs between two adjacent layers are recorded. Fig. 4.3 shows all 28 electrode pairs in the first layer that form measurable independent capacitors. Fig. 4.4 shows all 24 electrode pairs between the first and second layers that form measurable independent capacitors. Following this sensing scheme, the SCAS can generate 392 independent capacitance readouts per measurement frame. Each readout is calibrated using Eq. 3.3. The capacitances formed by these non-redundant combinations of SCAS electrodes are considered to contain sufficient information to portray full-geometry deformations as their receptive fields cover the entire soft body.
Electrode pairs in the same layer to form 28 independent capacitors

**Figure 4.3:** Electrode pairs (activated, marked in orange) in the first layer to form 28 measurable independent capacitors.
Electrode pairs between two adjacent layers to form 24 independent capacitors

**Figure 4.4:** Electrode pairs (activated, marked in orange) between layer 1 and layer 2 form 24 measurable independent capacitors.
4.2.3 Complex deformation generation

Four different types of loads are applied to generate various complex deformations:

1. The compound deformation of elongation and twisting $L_{(c,r)}$: A torsion force and a pulling force along the z-axis are simultaneously applied to the tip of the robot arm;

2. Pure bending $L_{(x,y)}$: A pulling force in the x-y plane is applied on the tip of the arm;

3. Two phase twisting and bending $L_{r,(x,y)}$: A torsion force is applied on the tip of the arm in the first $r$ frames ($r$ ranging from 6 to 16), and then a pulling force in the x-y plane is applied on the tip while maintaining the twisting state;

4. The compound deformation of twisting and bending $L_{(x,y,r)}$: A torsion force and a pulling force in the x-y plane are applied to the tip at the same time.

Each deformation is represented by a 3D point cloud with 1,716 points including 64 visual markers and corresponds to 392 capacitance readouts. Examples of deformations induced by different types of force loads and corresponding capacitance readouts are shown in Fig. 4.5-4.8 (the first row gives capacitance readouts and the second row shows point clouds depicting deformations).
Figure 4.5: A set of examples in one testing episode with a continuous $L_{(c,r)}$ deformation process. Capacitance readouts, ground truth point clouds and reconstruction of the original C2DT, the C2DT w/o markers and the C2DT w/o Chamfer distance are shown. The colour of each point is the distance from the corresponding ground truth point.
Figure 4.6: A set of examples in one testing episode with a continuous $L_{(x,y)}$ deformation process. Capacitance readouts, ground truth point clouds and reconstruction of the original C2DT, the C2DT w/o markers and the C2DT w/o Chamfer distance are shown. The colour of each point is the distance from the corresponding ground truth point.
Figure 4.7: A set of examples in one testing episode with a continuous $L_{r,(x,y)}$ deformation process. Capacitance readouts, ground truth point clouds and reconstruction of the original C2DT, the C2DT w/o markers and the C2DT w/o Chamfer distance are shown. The colour of each point is the distance from the corresponding ground truth point.
Figure 4.8: A set of examples in one testing episode with a continuous $L_{(x,y,r)}$ deformation process. Capacitance readouts, ground truth point clouds and reconstruction of the original C2DT, the C2DT w/o markers and the C2DT w/o Chamfer distance are shown. The colour of each point is the distance from the corresponding ground truth point.
4.2.4 Virtual proprioception dataset

956 different episodes of deformations are implemented through CFS to produce a virtual soft robot proprioception dataset. Each episode mimics a time-continuous deformation process and is discretised into about 40 frames. In each frame, the deformation and the corresponding capacitance readouts of the SCAS are recorded.

The dataset includes a total of 39,334 frames (956 episodes) of deformations and capacitance readouts, of which 2,319 frames (53 episodes) are with $L(\alpha, \beta)$; 12,552 frames (300 episodes) are with $L(\delta, \theta)$; 12,269 frames (303 episodes) are with $L(r, \delta, \theta)$ and 12,194 frames (300 episodes) are with $L(\delta, \theta, \phi)$.

The dataset is divided into three exclusive parts, i.e., training, validation and testing sets. The training set includes 22,517 frames (548 episodes), of which 1,334 frames (31 episodes) are with $L(\alpha, \beta)$, 7,204 frames (172 episodes) are with $L(\delta, \theta)$, 6,980 frames (173 episodes) are with $L(r, \delta, \theta)$ and 6,999 frames (172 episodes) are with $L(\delta, \theta, \phi)$. The validation set includes 9,721 frames (236 episodes), of which 550 frames (12 episodes) are with $L(\alpha, \beta)$, 3,093 frames (74 episodes) are with $L(\delta, \theta)$, 3,098 frames (76 episodes) are with $L(r, \delta, \theta)$ and 2,980 frames (74 episodes) are with $L(\delta, \theta, \phi)$. The testing set includes 7,096 frames (172 episodes), of which 435 frames (10 episodes) are with $L(\alpha, \beta)$, 2,255 frames (54 episodes) are with $L(\delta, \theta)$, 2,191 frames (54 episodes) are with $L(r, \delta, \theta)$ and 2,215 frames (54 episodes) are with $L(\delta, \theta, \phi)$. The specification of the dataset is shown in Table 4.1.

<table>
<thead>
<tr>
<th>Load type</th>
<th>$E_{\text{total}}$</th>
<th>$F_{\text{total}}$</th>
<th>$E_{\text{train}}$</th>
<th>$F_{\text{train}}$</th>
<th>$E_{\text{val}}$</th>
<th>$F_{\text{val}}$</th>
<th>$E_{\text{test}}$</th>
<th>$F_{\text{test}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L(\delta, \theta)$</td>
<td>300</td>
<td>12552</td>
<td>172</td>
<td>7204</td>
<td>74</td>
<td>3093</td>
<td>54</td>
<td>2255</td>
</tr>
<tr>
<td>$L(\alpha, \beta)$</td>
<td>53</td>
<td>2319</td>
<td>31</td>
<td>1334</td>
<td>12</td>
<td>550</td>
<td>10</td>
<td>435</td>
</tr>
<tr>
<td>$L(r, \delta, \theta)$</td>
<td>303</td>
<td>12269</td>
<td>173</td>
<td>6980</td>
<td>76</td>
<td>3098</td>
<td>54</td>
<td>2191</td>
</tr>
<tr>
<td>$L(\delta, \theta, \phi)$</td>
<td>300</td>
<td>12194</td>
<td>172</td>
<td>6999</td>
<td>74</td>
<td>2980</td>
<td>54</td>
<td>2215</td>
</tr>
<tr>
<td>Overall</td>
<td>956</td>
<td>39334</td>
<td>548</td>
<td>22517</td>
<td>236</td>
<td>9721</td>
<td>172</td>
<td>7096</td>
</tr>
</tbody>
</table>

$E$ denotes episode and $F$ denotes frame.
Smart e-skins and machine learning for soft robot perception

4.3 Capacitance-to-deformation transformer (C2DT)

4.3.1 Neural architecture

In general, the C2DT is a deep model (Fig. 4.9) that is able to deform the source point cloud \( P_s \) to approximate the target point cloud \( P \) based on the measurement characteristic tensor \((c, Q_{e1}, Q_{e2})\). Here \( P_s \in \mathbb{R}^{N_p \times 3} \) is the point cloud without deformation; \( N_p \) is the number of points in \( P_s \), which is 1,716 in this case; \( P \in \mathbb{R}^{N_p \times 3} \) and \( \hat{P} \in \mathbb{R}^{N_p \times 3} \) are the ground truth and reconstructed point clouds with a specific deformation respectively; \( c \in \mathbb{R}^{N_m \times 1} \) is the corresponding calibrated capacitance readouts vector; \( N_m \) is the number of readouts in \( c \) with the value of 392 in this case; \( Q_{e1} \in \mathbb{R}^{N_m \times 3} \) and \( Q_{e2} \in \mathbb{R}^{N_m \times 3} \) are the coordinates of electrodes to generate \( c \).
Figure 4.9: Neural architecture of C2DT.
Smart e-skins and machine learning for soft robot perception

The C2DT architecture consists of two parts, i.e., encoding and decoding. The input of the encoding part is \( c, Q_{e1} \) and \( Q_{e2} \). \( Q_{e1} \) and \( Q_{e2} \) are considered as positional signals that can help distinguish different elements in \( c \). They pass through the multilayer perceptron (MLP) \( f_q(\cdot) \) to obtain the geometrical representations of individual electrodes. An element-wise max function is selected to integrate the two electrode representations into the final geometrical representations for electrode pairs as the capacitance is independent of the order of electrodes according to the reciprocal theorem. The MLP \( f_c(\cdot) \) maps \( c \) to high-dimensional representations, and the sum of capacitive and geometrical representations is the input of the transformer encoder \( E(\cdot) \) with the length of \( N_m \). For the decoding part, \( P_s \) is first fed to the MLP \( f_s(\cdot) \), and then multi-head attention is implemented over the outputs of \( f_s(\cdot) \) and \( E(\cdot) \) through the transformer decoder \( D(\cdot) \). The MLP \( f_d(\cdot) \) is used to map the output sequence of \( D(\cdot) \) to the displacement of each point, and the reconstruction \( \hat{P} \) is obtained by adding it to \( P_s \).

\( \hat{P} \) is expected to be as close as possible to the target point cloud \( P \). This goal is achieved by minimising the following loss function consisting of the squared distance term of visual markers (of which point-to-point correspondences are known) and the Chamfer distance term of the remaining points (of which point-to-point correspondences are unknown).

\[
\mathcal{L} = \mathbb{E}_{P \sim \mathcal{P}} \left[ \lambda_1 \sum_{i=1}^{N_v} |p^i_v - \hat{p}^i_v|^2 + \lambda_2 \sum_{j=1}^{N_r} \left( \min_{p_r \in P_r} |p_r - \hat{p}^j_r|^2 + \min_{\hat{p}_r \in \hat{P}_r} |p^j_r - \hat{p}_r|^2 \right) \right] \tag{4.1}
\]

where \( P_r \in \mathbb{R}^{N_r \times 3} \) represents the remaining points; \( p^j_r \in \mathbb{R}^3 \) is the coordinates of the \( j^{th} \) remaining point; \( p^i_v \in \mathbb{R}^3 \) is the coordinates of the \( i^{th} \) visual marker; \( N_v \) and \( N_r \) are the numbers of the visual markers and the remaining points respectively; \( \mathcal{P} \) is the distribution of \( P \); \( \lambda_1 \) and \( \lambda_2 \) are the weights of the squared distance term of the visual markers and the Chamfer distance term of the remaining points, respectively.

The structures of subnetworks of the C2DT are as follows:

- \( f_s \): Linear(3, \( h_{em} \)) \( \rightarrow \) ReLU \( \rightarrow \) LayerNorm(\( h_{em} \)) \( \rightarrow \) Linear(\( h_{em} \), \( d_{model} \)) \( \rightarrow \) ReLU \( \rightarrow \) LayerNorm(\( d_{model} \))
- \( f_q \): Linear(3, \( h_{em} \)) \( \rightarrow \) ReLU \( \rightarrow \) LayerNorm(\( h_{em} \)) \( \rightarrow \) Linear(\( h_{em} \), \( d_{model} \))
- \( f_c \): Linear(1, \( h_{em} \)) \( \rightarrow \) ReLU \( \rightarrow \) LayerNorm(\( h_{em} \)) \( \rightarrow \) Linear(\( h_{em} \), \( d_{model} \))
\textbf{Smart e-skins and machine learning for soft robot perception}

- $f_d$: \text{Linear}(d_{\text{model}}, 3) \rightarrow a \ast \text{Tanh}
- $E$: \text{LayerNorm}(d_{\text{model}}) \rightarrow \text{Transformer.EncoderLayer}(d_{\text{model}}, d_{\text{ff}}, h, P_{\text{drop}}) \otimes n_{\text{e-layer}}$
- $D$: \text{Transformer.MutualLayer}(d_{\text{model}}, d_{\text{ff}}, h, P_{\text{drop}}) \otimes n_{\text{m-layer}} \rightarrow \text{Transformer.DecoderLayer}(d_{\text{model}}, d_{\text{ff}}, h, P_{\text{drop}}) \otimes n_{\text{d-layer}}$

where $h_{\text{em}}=32$, $d_{\text{model}}=128$, $a=1.2$, $d_{\text{ff}}=256$, $h=8$, $P_{\text{drop}}=0.1$, $n_{\text{e-layer}}=3$, $n_{\text{m-layer}}=1$ and $n_{\text{d-layer}}=2$. Linear layers in $f_d$ and $f_c$ do not have learnable biases while others have. The LayerNorm in $E$ takes the sum of capacitive and geometrical representations as input. \text{Transformer.EncoderLayer} and \text{Transformer.DecoderLayer} are exactly the same as the original transformer [83]. The first self-attention cell of \text{Transformer.DecoderLayer} is removed and the remaining part is used as \text{Transformer.MutualLayer} because $P_s$ remains constant. \text{Transformer.EncoderLayer} \otimes n_{\text{e-layer}} represents a stack of $n_{\text{e-layer}}$ \text{Transformer.EncoderLayer}.

4.3.2 Implementation

The C2DT is implemented in Python and PyTorch packages [125]. The Adam [123] optimiser ($\beta_1=0.9$, $\beta_2=0.98$, $\varepsilon=10^{-9}$) is employed to update learnable parameters and minimise $\mathcal{L}$. The initial learning rate is set to 0.001, which is decayed by a factor of 1.2 every 15 epochs. $\lambda_1$ and $\lambda_2$ are computed as follows: $\lambda_1 = \lambda / 3(\lambda N_v + 2N_r)$, $\lambda_2 = 1/3(\lambda N_v + 2N_r)$, where $\lambda = \max(1, 300 - 2 \ast (\text{epoch} - 1))$. The gradient is clipped with the threshold of 0.5. The C2DT is trained using the training set for 300 epochs with a batch size of 24. Each epoch takes about 9 min on 3 Nvidia Quadro P5000. The network with the least validation loss is saved as the final model.

4.4 Morphological reconstruction results on the virtual dataset

4.4.1 Error metrics

The performance of the C2DT is quantitatively evaluated through 4 error metrics, i.e., the average distance (AD), the maximal distance (MD), the Chamfer distance (CD) and the Hausdorff distance (HD):

$$AD = \frac{1}{N_p} \sum_{i=1}^{N_p} |p^i - \hat{p}^i|_2$$  \hspace{1cm} (4.2)
4.4.2 Results with different network hyperparameters

Several C2DTs with different hyperparameters are trained and their performance is compared in Table 4.2. The C2DT with 6 transformer layers outperforms the other candidates. The AD error achieved with this setup is as low as 1.379±1.048 mm, comparable to the accuracy achieved with RGB-D cameras frequently used as ground truth in the relevant research [126].

4.4.3 Ablation studies

Ablation studies of the C2DT with 6 transformer layers are implemented to better understand the role of each loss term and position encoding. The results are shown in Fig. 4.5-4.8, 4.10 and Table 4.3. The C2DT cannot learn correct point-to-point correspondences without including visual markers in training. This phenomenon is illustrated in Fig. 4.10, where the points in the region of interest of the source point cloud are not mapped into the correct corresponding region using the C2DT w/o markers. Although reconstructions show similarities with the ground truth by minimising the Chamfer distance term, point-to-point errors remain large. Local distortions arise in a set of frames of reconstructions if retaining only the squared distance term of the visual markers during training. This indicates that the Chamfer distance term can benefit the geometrical quality of the reconstructions.
Table 4.2: Quantitative evaluation of C2DTs with different network hyperparameters

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Hyperparameters</th>
<th>o/a</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n_e-layer</td>
<td>n_m-layer</td>
</tr>
<tr>
<td>AD</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>MD</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>CD</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>HD</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

The results are based on the simulation testing set and presented in mean ± standard deviation format. The unit of all metrics in the table is mm. The least errors among different models are marked in bold.
Figure 4.10: A set of examples of reconstructions generated by different C2DTs.
Table 4.3: Quantitative evaluation of C2DTs with different loss terms on the virtual proprioception dataset

<table>
<thead>
<tr>
<th>Models</th>
<th>AD</th>
<th>MD</th>
<th>CD</th>
<th>HD</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2DT</td>
<td>1.379 ± 1.048</td>
<td>4.609 ± 3.775</td>
<td>1.356 ± 0.865</td>
<td>4.579 ± 3.532</td>
</tr>
<tr>
<td>w/o markers</td>
<td>119.5 ± 3.1</td>
<td>187.7 ± 4.2</td>
<td>8.402 ± 2.217</td>
<td>29.190 ± 8.039</td>
</tr>
<tr>
<td>w/o Chamfer distance</td>
<td>2.573 ± 1.250</td>
<td>18.285 ± 10.006</td>
<td>2.329 ± 0.995</td>
<td>17.254 ± 9.048</td>
</tr>
</tbody>
</table>

The results are based on the simulation testing set and presented in mean ± standard deviation format. The unit of all metrics in the table is mm. The least errors among different models are marked in bold.
4.4.4 Position representations visualisation

It is observed that the network does not converge in training if the position encoding part is removed. The position representations of the trained C2DT are visualised through t-distributed stochastic neighbour embedding (t-SNE) [127] in Fig. 4.11. After position encoding, the electrode pairs with high geometrical correlation tend to cluster together, and the electrode pairs geometrically far apart are also far apart in the feature space. It suggests that the position encoder can generate distinctive geometric representations based on the locations of the input electrode pairs.

4.4.5 Effect of numbers of visual markers and electrodes

The redundant SCAS design validates its feasibility in the virtual environment. However, the high density of markers and electrodes poses practical challenges to the fabrication and experimentation of the physical system. In order to reduce complexity while maintaining proprioception performance, the impact of the number of markers and electrode layout on the performance of the C2DT is investigated. The results of this analysis, shown in Fig. 4.12, prove the accuracy improvement from increasing the number of markers plateaus at 16. It provides evidence that a small set of markers is sufficient for the C2DT to establish correct point-to-point correspondences. Similarly, the reconstruction performance improves with the density of electrodes, but the improvement is minimal after the number of electrodes exceeds a certain value (e.g., 32), as illustrated in Fig. 4.13. These results highlight a favourable trade-off between reconstruction accuracy and electrode/markers units, confirming that it is safe to sacrifice a minute part of the performance to simplify the fabrication and deployment of the SCAS.
Figure 4.11: Visualisation of position representations for the simulation via t-SNE.
Figure 4.12: The performance of the C2DT under 4 different numbers of markers (mean ± standard deviation).
Figure 4.13: The performance of the C2DT under 4 different electrode layouts (mean ± standard deviation).
4.5 E-skin fabrication and characterisation

4.5.1 E-skin fabrication

Based on the conclusions from the above investigation, a physical SCAS with 32 electrodes consisting of 8 4-electrode SCAS modules is selected. This balances the full-geometry reconstruction performance with fabrication complexity. Multiple 4-electrode SCAS modules are manufactured in parallel using established elastomer processing technologies [67]. The electrodes are made of carbon black (CB) dispersed elastomers. However, this material is unsuitable for wires and interfaces due to its high resistivity and non-linear, irreversible conductivity response under deformation[128, 129]. Therefore, Eutectic Gallium 75.5% Indium 24.5% (EGaIn) is employed to fabricate the wires and interfaces due to its high conductivity \((3.4 \times 10^7 \text{ S m}^{-1})\) and stable response to deformation.

Figure 4.14: Four-electrode SCAS modules with EGaIn (centre) and CB (right) wires and the cross-section of the module with EGaIn wires under a 40x digital microscope (left).

The 4-electrode SCAS module \((20 \times 120 \text{ mm})\) consists of 4 different functional layers (Fig. 4.14), i.e., the protective substrate (thickness: 0.39 mm), the electrode layer (0.08 mm), the isolation layer (0.24 mm), and the sealing layer (0.3 mm). Each SCAS module is fabricated layer by layer. The steps are shown in Fig. 4.15:
**Smart e-skins and machine learning for soft robot perception**

i) Smooth-on Ecoflex 00-30 part A (1.0) and part B (1.0) are mixed and poured on a glass plate. Then a TQC Sheen micrometer film applicator is used to flatten the mixture which then is cured for 3 min at 100°C.

ii) Imerys Enasco 250P conductive carbon black (0.2) is first mixed with isopropyl alcohol (2.0), after which the uncured silicone mixture (2.0) is added. The whole mixture is stirred for 3 min. A layer of uncured conductive silicone mixture is coated on the protective substrate and is cured for 3 min in a 100°C oven.

iii) A 40 W Aeon MIRA 5 laser machine is employed to pattern CB electrodes. The parameters are set as follows: 28% Power, 300 mm s\(^{-1}\) Speed and 0.05 mm Interval. The planar size of each electrode is 21 × 6 mm, which is one-fifth of the one studied in the simulation.

iv) We use the same method as in step i) to fabricate a silicone membrane for the isolation layer on the top of the electrode layer.

v) Two rounds of engraving are performed with 20.5% Power, 300 mm s\(^{-1}\) Speed and 0.05 mm Interval to generate micro channels of liquid metal wires and connections to readout electronics. Four rounds of engraving are conducted with the same parameters to generate vertical interconnect holes. The planar size of readout connections and vertical interconnect holes is 3 × 2 mm, and the width of wires is 0.5 mm. Then we cut the rectangular area of the modular SCAS with 19.5% Power and 25 mm s\(^{-1}\) Speed and remove the remaining part.

vi) We fabricate a new silicone membrane following step i, and we uniformly coat a very thin layer of uncured silicone mixture on its surface as the adhesive. Then we bond the SCAS cut in step v with the membrane. The curing takes about 4 h under room temperature to ensure high-quality bonding.

vii) We inject Eutectic Gallium 75.5% Indium 24.5% (EGaIn, Sigma Aldrich) ink from readout connections, and meanwhile exhaust the air in microchannels through the vertical interconnect holes.

viii) We obtain the final modular 4-electrode SCAS. The planar size of the SCAS module is 120 × 20 mm, of which 100 × 20 mm is the area of the electrodes, and 20 × 20 mm is the interface to readout electronics.

The layer thicknesses are 0.39 mm, 0.08 mm, 0.24 mm and 0.3 mm, respectively. Since the fabrication is easy to scale up, 5 SCAS modules are manufactured in parallel.
Figure 4.15: Fabrication process of a modular 4-electrode SCAS.
4.5.2 E-skin deployment

A square cylinder robot arm (made of Ecoflex 00-30 silicone) with the size of $20 \times 20 \times 240$ mm is selected as the testbed, which is one-fifth of the one in the simulation. The extra 40 mm in height is the interface area for driving the deformation, connecting to electronics and bonding with the fixed ceiling. Eight 4-electrode SCAS modules are bonded to the robot's surface to form the 32-electrode SCAS (see Fig. 4.16 top line). The soft robot and SCAS are fabricated with the same material (Ecoflex 00-30), which allows them to be firmly merged, with no modulus mismatch, by using uncured Ecoflex 00-30 silicone as the adhesive. The unity of the material enables the robot and SCAS to be considered as a whole system during experiments, thus minimising the effect of SCAS on the original robot motion and deformation. Fig. 4.17 shows the adhesion between the SCAS and the robot under various deformations. No separation or dislocation was observed in all cases.

The transparency of silicone adversely impacts the quality of the point clouds collected by RGB-D cameras based on the time-of-flight principle. Therefore, a silicone layer with white Smooth-on Silc Pig Silicone Pigments is coated on the surface for better reflection. Sixteen yellow dots are attached to the robot arm as visual markers, assisting network training with correspondence information. The interface to readout electronics is covered with black acrylic tape to reduce its interference in point cloud collection (Fig. 4.16 bottom line). Individual electrodes are indexed and accessible from the readout electronics.
Figure 4.16: The soft arm equipped with 8 4-electrode SCAS modules (top) is coated with a white silicone layer and 16 yellow visual markers to improve data acquisition.
Figure 4.17: The bonding interface between the SCAS and the soft robot platform
4.5.3 E-skin characterisation

To characterise the response of the SCAS module and verify the superior performance of EGaIn wires compared with CB wires, a 4-electrode SCAS with CB wires and a 4-electrode SCAS with EGaIn wires are attached on the front and back sides of a segment of the square cylinder silicone structure (20×20×140 mm) and cyclically stretch them using a Nema23 stepper motor with an SFU1605 ball screw (see Fig. 4.18). Each cycle takes 20 s, and the SCASs are strained by up to 40%. The entire test takes about 3 h (more than 500 cycles). Relative capacitance readouts of each SCAS are illustrated in Fig. 4.19-4.21. The results show that the SCAS with EGaIn wires has better sensitivity (larger response under the same deformation), linearity (no distortions in response curves) and cycling stability (no drift after 500 cycles).
Figure 4.18: The SCAS modules are bonded to different sides of a short silicone pillar and cyclically stretched by a stepper motor.
Figure 4.19: Three relative capacitance readouts of the SCAS with EGaIn wires during the whole 3 h cycling test (left). The relative capacitance readouts the SCAS with EGaIn wires after 10, 100 and 500 cycles (right).
Figure 4.20: Three relative capacitance readouts of the SCAS with CB wires during the whole 3 h cycling test (left). The relative capacitance readouts of the SCAS with CB wires after 10, 100 and 500 cycles (right).
Figure 4.21: Relative capacitance response curves of the two SCAS modules to a 40% periodic linear stretch. The SCAS with EGaIn wires shows better sensitivity, linearity and cycling stability than its CB wires counterpart.
4.6 Experimental data acquisition

4.6.1 Experimental setup

The experiment platform consists of the soft robot arm equipped with the 32-electrode SCAS, the readout electronics, two Microsoft Azure Kinect RGB-D cameras [130] and a laptop installed with a customised software to control the readout electronics and record data from the cameras and the SCAS (see Fig. 4.22). The readout electronics is based on a 32-electrode ECT system that supports arbitrary switching schemes [131]. Its capacitance measurement resolution is 3 fF, and the signal-to-noise ratio of all 32 channels is above 60 dB.

The two cameras are placed directly opposite and in a straight line with the robot arm to capture its 3D deformations from two complementary views in real time. The deformations are saved and represented via the colour point cloud format. The data recording of the cameras and readout electronics is synchronised. The frame rate can reach about 30 fps if only record the point cloud and capacitance data. It will decrease to around 20 fps if RGB images are also recorded.
Figure 4.22: Real-world experiment platform.
4.6.2 Data acquisition and pre-processing

The reliability of the SCAS allows recording capacitance readouts frames (each frame comprises 76 independent readouts) when the robot arm is subject to arbitrary external loading applied via the bottom holder over a long period (about 1,220 s of deformation data are intermittently collected during a 10 h experiment). To demonstrate the superiority of the proposed approach, a random sequence of complex deformations is implemented, including omnidirectional bending, omnidirectional elongation, twisting around an arbitrary axis and their compound deformations (examples are shown in Fig. 4.23), during the experiment. The SCAS and camera data (i.e., capacitance readouts, colour point clouds, and sometimes RGB images) are synchronously recorded. In total, 36,465 frames (about 1,220 s) of experimental data are collected. In this real-world dataset, the first 36,013 frames (about 1,200 s) record only the capacitance readouts and colour point clouds; the last 452 frames (about 20 s) also save the RGB images with a reduced frame rate.

The 32-electrode SCAS can produce 76 capacitance readouts in a single frame, which are calibrated using the same method as in the simulation. The point clouds from the two cameras are fused in one coordinate system using the chessboard calibration method [132, 133]. The raw data is noisy and contains many meaningless background points, making it unusable for direct training. Matlab is employed to selectively retain only the points on the surface of the robot arm. The points on the black acrylic tape and red holders are eliminated via colour filtering. In order to further reduce the negative impact of noise and outliers, data in the regions whose local point densities are lower than a preset threshold are filtered.

Due to inevitable visual occlusion, in many frames the cleaned point clouds cannot completely represent 3D deformations. To alleviate this issue, further preprocessing is required prior to training. Average grid downsampling with a 4 mm box gird filter is implemented at first. Then $\alpha$ shape reconstruction [134] on the basis of the downsampled point clouds is used to alleviate the issue of incomplete representation. The triangular meshes of the alpha shapes are subdivided three times, and vertexes are extracted as new point clouds with supplementary points. In the C2DT framework, the numbers of points in the source and target point clouds are expected to be the same. In order to meet this requirement, average grid downsampling with a 4 mm box gird filter is first implemented, after which farthest point sampling [135] is applied to eventually select 1,300 points in each point cloud.
Figure 4.23: A representative set of complex deformations that appear in the experiment.
Yellow visual markers are extracted from cleaned point clouds before downsampling and \( \alpha \) shape reconstruction based on the RGB information of each point. A graph is created according to one frame of marker points. The connection of each two points in the graph is determined by their distance. The threshold of the connected distance is 6 mm. Each connected subgraph with more than 10 points is considered as a visual marker, and the average of the coordinates of all points in a subgraph is used to represent the marker position. The number of extracted visual markers is not always 16 due to camera occlusion.

It is almost impossible to automatically obtain point-to-point correspondences of visual markers under the current experimental setup. Therefore visual markers are aligned layer-to-layer. The 16 visual markers can be divided into 4 layers, and each layer includes 4 markers. A graph is created based on one frame of coordinates of extracted markers with the connected distance threshold of 26 mm. Each subgraph is a layer of markers. The permutation of the layer is determined by the relative position in the y-axis of the fused coordinate system among all 4 layers. All abnormal frames, for which the number of extracted markers is larger than 16 and/or the number of layers is not equal to 4, are deleted. The layers for which the number of markers is less than 4 are filled with (0,0,0) to ensure all layers have the same number of points, which can improve computational efficiency during training. Furthermore, the frames with critically missing points issues are removed because of the low quality of their reconstructed \( \alpha \) shapes. The number of markers in individual layers indicates the severity of missing points. The frames with at least 2 markers in all layers are retained while others are dismissed.

Upon the above filtering process, a total of 30,973 frames of data remain available for analysis. A sample of 500 frames out of the dataset is inspected and no serious missing points issues are found. A set of examples in this dataset is shown in Fig. 4.24-4.26. The real-world dataset is divided into three exclusive parts. The first 26,771 frames (about 1,020 s) are used for training (20,693 frames) and validation (6,018 frames), and the last 4,262 frames (about 200 s) are used for testing.
Figure 4.24: The curves of calibrated capacitance readouts of the SCAS during an about 20 s period in the experiment.
Figure 4.25: A group of adjacent frames in the real-world dataset is shown, which records a continuous bending process with RGB images, point clouds and capacitance readouts.
Figure 4.26: A group of adjacent frames under a deformation process that combines elongation and twisting.
4.7 C2DT for the real-world dataset

4.7.1 Modification on the framework

The challenge of real-world high PGR proprioception is exacerbated by the relatively poor quality of point clouds (restricted by the accuracy of cameras, occlusion, and light conditions), noise in the SCAS signals, and imperfect synchronization between different devices. In order to compensate for these added sources of inaccuracy, the C2DT framework is enhanced by increasing the number of input frames ($N_i$ adjacent frames of SCAS readouts) and introducing a regularisation term in its loss function to limit the distance change between neighbouring points before and after deformation.

The basic framework of C2DT in the real-world experiment is analogous to that in the simulation. However, the loss function in simulation is no longer applicable, as in experiments the point-to-point correspondences of visual markers are not available. Instead, a modified loss function is proposed as follows.

$$
\mathcal{L}^* = E_{P \sim G} \left\{ \lambda_1 \sum_{k=1}^{N_l} \sum_{i=1}^{N_{lv}} \left[ d(\hat{p}_{i,k}^j, P_{l,k}) \cdot S_{r2g}^{k,i} + d(p_{i,k}^j, \hat{P}_{l,k}) \cdot S_{r2g}^{k,i} \right] 
+ \lambda_2 \sum_{j=1}^{N_r} \left[ d(\hat{p}_r^j, P_r) + d(p_r^j, \hat{P}_r) \right] 
+ \lambda_3 \sum_{j=1}^{N_r} \sum_{l=1}^{N_l} \left[ \left( |\hat{p}_r^j - \hat{p}_r^j| - |\hat{p}_r^j - p_r^j| - \delta_u \cdot S_{s}^{j,l} \right)^2 \cdot S_{d}^{j,l} + \left( |\hat{p}_r^j - \hat{p}_r^j| - |\hat{p}_r^j - p_r^j| - \delta_u \cdot S_{s}^{j,l} \right)^2 \cdot S_{u}^{j,l} \right] \right\} (4.6)
$$

The first term of $\mathcal{L}^*$ counts the Chamfer distance between the reconstruction and the ground truth of markers layer-by-layer, where $P_{l,k} \in \mathbb{R}^{N_{lv} \times 3}$ is the coordinates of the visual markers in the $l_k$ layer; $p_{i,k}^j \in \mathbb{R}^3$ is the coordinates of the $i^{th}$ point in $P_{l,k}$; $d(\hat{p}_{i,k}^j, P_{l,k})$ is the squared distance between $\hat{p}_{i,k}^j$ and the nearest point in $P_{l,k}$; $N_l$ is the number of layers; $N_{lv}$ is the number of markers in each layer and the values of $N_l$ and $N_{lv}$ are 4 in this case. When computing the loss, we only need to consider the marker points extracted in the data preprocessing and ignore the padding points. Note that all points in $\hat{P}_{l,k}$ are marker points as they are generated by the network based on the corresponding capacitance readouts and the source point, which does not include padding points. In order to eliminate the effect of padding points during training, we synthesize masks $S_{r2g}^{k,i}$ and $S_{r2g}^{k,i}$ as follows.

- $S_{r2g}^{k,i}$ is set to 1 if $p_{i,k}^j$ is a marker point. $S_{r2g}^{k,i}$ is set to 0 if $p_{i,k}^j$ is a padding point.
Smart e-skins and machine learning for soft robot perception

- \( S^{k,i}_{r2g} \) is set to 1 if \( P_k \) does not include any padding points, otherwise \( S^{k,i}_{r2g} \) is set to 0.

The second term in \( \mathcal{L}^* \) is exactly the same as its simulation counterpart that counts the Chamfer distance between the reconstruction and ground truth of the remaining points. The third term is a regularizer that encourages the distance between neighbouring points to not change significantly before and after deformations, where \( \hat{p}^j_r \) is the \( i^{th} \) neighbour of \( \hat{p}^j_r \); \( s^{j,l} \) is the distance between the corresponding two points in the source point cloud; \( \delta_d \) and \( \delta_u \) are coefficients of thresholds. We count the loss only if the neighbour distance in the reconstruction falls outside the preset range. We achieve this with masks \( S^{j,l}_d \) and \( S^{j,l}_u \) as follows.

- \( S^{j,l}_d \) is set to 1 if \( |\hat{p}^j_r - \hat{p}^{j,l}_r|_2 - \delta_d \cdot s^{j,l} < 0 \), otherwise \( S^{j,l}_d \) is set to 0.
- \( S^{j,l}_u \) is set to 1 if \( |\hat{p}^j_r - \hat{p}^{j,l}_r|_2 - \delta_u \cdot s^{j,l} > 0 \), otherwise \( S^{j,l}_u \) is set to 0.

4.7.2 Implementation

The number of input frames in the physical world is not constant to 1. In contrast, the C2DT takes several \((N)\) adjacent frames as its input. The first linear cell in \( f_c \) is therefore modified to Linear\((N, h_{em})\). The hyper-parameters of the C2DT are set as: \( h_{em}=32, d_{model}=64, d_{ll}=128, h=4, P_{drop}=0.1, n_{e-layer}=2, n_{m-layer}=1 \) and \( n_{d-layer}=1 \). The network is trained and evaluated using almost the same procedure as presented earlier.

The other learning parameters are set as \( \delta_d=0.5, \delta_u=2, \lambda_1 = \lambda / [\lambda \sum_{j=1}^{N_j} \sum_{l=1}^{N_{e}} (S^{k,i}_{r2g} + S^{k,i}_{g2r}) + 2N_j], \lambda_2 = 1 / [\lambda \sum_{j=1}^{N_j} \sum_{l=1}^{N_{e}} (S^{k,i}_{r2g} + S^{k,i}_{g2r}) + 2N_j], \lambda_3 = 1 / [\lambda \sum_{j=1}^{N_j} \sum_{l=1}^{N_{e}} (S^{k,i}_d + S^{k,i}_u)], \) where \( \lambda = \max(1, 300 - 10 \times (epoch - 1)) \). In total, the network is trained 200 epochs with a batch size of 39. The one with the least validation loss is saved as the final model.
4.8 Reconstruction results on the real-world dataset

Several C2DTs with different input frame numbers are trained using the filtered real-world dataset. The full-geometry reconstruction performance improves as the input frame number increases and achieves the minimum error at 3 adjacent input frames (Fig. 4.27). This improvement indicates that increasing the number of input frames can reduce the negative impacts of noise in SCAS signals and asynchronisation between devices. The temporal correlation among adjacent frames can also be considered favourable for deformation reconstruction.

![Figure 4.27: The performance of C2DTs, which take different numbers of adjacent frames as inputs (mean±standard deviation).](image)

A representative set of reconstructions of the C2DT with 3 input frames is shown in Fig. 4.28. The results achieve $2.322 \pm 0.687$ mm for the CD metric (Table 4.4) with a PGR of 3,900 (i.e., 1,300 points in each point cloud). Compared to simulation results, some elaborate geometrical features of reconstructed deformations in certain frames are less obvious, especially for those related to twisting. This is mainly because the ground truth point clouds acquired by RGB-D cameras cannot reach the quality of point clouds synthesised by the CFS method.
Figure 4.28: A representative set of examples of high PGR 3D deformation reconstruction.
Smart e-skins and machine learning for soft robot perception

Ablation studies are implemented to evaluate the effect of individual loss terms. Fig. 4.29 shows the results and demonstrates that visual markers play a similar role in physical and virtual environments, facilitating the network to learn correct point-to-point correspondences. The introduction of the neighbour regularisation term can slightly improve the reconstruction quality. Table 4.4 provides the quantitative evaluation of different C2DTs using CD and HD metrics, which do not require point-to-point correspondences.

Table 4.4: Quantitative evaluation of C2DTs with different loss terms on the experimental dataset

<table>
<thead>
<tr>
<th>Models</th>
<th>CD</th>
<th>HD</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2DT</td>
<td>2.322 ± 0.687</td>
<td>8.498 ± 3.798</td>
</tr>
<tr>
<td>w/o neighbour regularizer</td>
<td>2.292 ± 0.760</td>
<td>8.528 ± 4.072</td>
</tr>
<tr>
<td>w/o markers</td>
<td>2.326 ± 0.647</td>
<td>8.683 ± 3.855</td>
</tr>
<tr>
<td>w/ Chamfer distance only</td>
<td>2.382 ± 0.617</td>
<td>8.900 ± 3.727</td>
</tr>
</tbody>
</table>

The results are based on the real-world testing set and presented in mean ± standard deviation format. The unit of all metrics in the table is mm.

The position encoding part is crucial to extract useful proprioceptive information from physical SCAS signals. Similar to its contribution in training with the simulation dataset, position encoding can assign discriminative high-dimensional representations to different electrode pairs based on their geometrical structures (see Fig. 4.30 for the t-SNE embedding of position representations).
Figure 4.29: Qualitative illustration of ablation studies results.
Figure 4.30: Two-dimensional t-SNE embedding of the output of the position encoding part in the trained C2DT.
4.9 Summary

This chapter proposes a proprioception system that could visualise high PGR 3D full-geometry deformations of soft robots. It is empowered by an intrinsically-stretchable capacitive e-skin that leverages capacitances formed by the combinations of planar boundary electrodes, and an end-to-end neural architecture to translate SCAS signals directly into point clouds. The proposed proprioception system can achieve real-time (30 fps), high PGR (3,900) and full-geometry deformation reconstruction with high accuracy (2.322 ± 0.687 mm CD error) under complex deformations. This level of proprioception represents a step change over previous attempts and is beyond existing proprioception systems. Notably, the system has the potential to be extended to different types of soft bodies through a straightforward learning process without requiring a priori knowledge. Implementing such high PGR, full-geometry proprioception is essential for perceiving full-body status and achieving precise closed-loop control of soft robots, the key to breakthroughs in performing complex tasks.
Chapter 5

Tactile sensing under interference from deformation

5.1 Introduction

Tactile sensing in soft robots remains particularly challenging because of the coupling between contact and deformation information which the sensor is subject to during actuation and interaction with the environment. This often results in severe interference and makes disentangling tactile sensing and geometric deformation difficult. Most existing approaches implement tactile sensing without considering the impact of deformation on sensing data, preventing them from real-world adoption. The SCAS developed in the previous chapter is sensitive to variations in the permittivity distribution typically induced by the physical proximity and contact with external objects, allowing it to be applied to tactile sensing. Although the SCAS signals can be affected by the deformation of the soft robot investigated, the permittivity and geometric variations trigger distinct sensor responses, offering the opportunity to decouple the signals to obtain tactile and morphological information at the same time.

This chapter proposes a simplified SCAS to achieve physical contact detection under deformation interference. The simplified SCAS has a concise structure and sparse layout reducing complexity in fabrication, deployment and wiring, making it favourable to soft robot tactile sensing. Furthermore, the C2DT is also employed to track 5 visual markers deployed on the soft robot platform investigated on the basis of the e-skin signals, which demonstrates the potential of the SCAS to simultaneously detect internal and external stimuli.
5.2 Simplified SCAS

The simplified SCAS is schematically illustrated in Fig. 5.1. It consists of 3 layers, i.e., the protective substrate, the liquid metal wires and interfaces, and the sealing layer. The size of the e-skin is $40 \times 110$ mm. The CB electrodes are removed from the original design and the EGaIn liquid metal wires act as the electrodes of which the combinations form capacitors. The different lengths of individual wires are designed to magnify the difference in capacitance signals when touching different areas of the e-skin. The wires are 1 mm wide and respectively 20, 45, 70 and 95 mm long. The size of the liquid metal interfaces is $5 \times 5$ mm.

Figure 5.1: Schematic illustration of the simplified SCAS.

The protective substrate with micro-channels of liquid metal wires and interfaces are fabricated using Ecoflex 00-30 silicone and a 3D printed casting. A silicone membrane manufactured through film coating is bonded to the substrate as the sealing layer using uncured silicone as the adhesive. EGaIn liquid metal ink is injected from the interfaces and the air is exhausted through the ends of the wires. Fig. 5.2 shows the snapshots of the e-skin without deformation and at 100% elongation.
In the previous chapter, the performance of the SCAS has been well validated under various deformations such as bending, twisting and elongation. In this chapter, a 3-chamber pneumatic manipulator (50×120 mm) is selected as the testbed. It can generate inflation deformations which are not included in the previous chapter, thus further demonstrating the wide application range of the SCAS.

The structure of the robot is shown in Fig. 5.3. The size of each chamber is 40×30 mm. The width of each inlet is 1.5 mm. Two e-skin modules are bonded to the robot surfaces (one on the front and one on the back) using uncured silicone as the adhesive. They form an 8-electrode capacitive sensor (each module has 4 electrodes) which generates 28 capacitance readouts per measurement frame.
5.3 Characterisation

5.3.1 Inflation

A 3-dimensional vector $\mathbf{p} = (p_1, p_2, p_3)$ is used to describe the inflation state of the robot, where $p_1$, $p_2$ and $p_3$ are the volumes of air injected into the 1\textsuperscript{st}, 2\textsuperscript{nd} and 3\textsuperscript{rd} chambers, respectively. Twenty ml of air is injected into each chamber of the robot simultaneously, i.e., $\mathbf{p} = (20, 20, 20)$ ml, and record the e-skin response signals. Figure 5.4 shows all 28 capacitance readouts during the inflation process.

![Figure 5.4: Twenty-eight calibrated capacitance readouts during the (20,20,20) ml inflation process.](image-url)
Smart e-skins and machine learning for soft robot perception

The $p = (20,20,20)$ ml inflation can stimulate a maximum variation of around 40% relative capacitance change. The capacitance readouts of capacitors formed by electrodes in the same surface typically increase as the pneumatic robot is inflated (examples see Fig. 5.5.a and c). The capacitance readouts of capacitors formed by electrodes in different surfaces show the opposite trend (examples see Fig. 5.5.b and d). The inflation of the robot body makes the area of electrodes (positively correlated to capacitance) larger and the distance between electrodes (negatively correlated to capacitance) longer. For capacitors formed by electrodes on the same surface, the increase in the area of electrodes dominates the change in capacitance. For capacitors formed by electrodes on different surfaces, the variation in the distance between electrodes is more significant, which dominates the capacitance variation.

5.3.2 Touch under inflation

A 2-stage experiment is implemented to show the difference between capacitance change induced by touch (permittivity variation) and deformation (geometric variation). In the first stage, the manipulator is inflated to (0,20,20) ml without touch. Then the front surface of the robot is divided into 9 sub-regions (see Fig. 5.6). Each sub-region is touched sequentially. Fig. 5.6 shows the overall response of the e-skin. The capacitance response of the (0,20,20) ml inflation is similar to that of (20,20,20) ml inflation shown in Fig. 5.4. Inflation can trigger variations in a large part of capacitance readouts simultaneously, while touch only stimulates changes in several specific readouts based on the location of the touch point. Examples of capacitance readouts from 4 different electrode pairs are shown in Fig. 5.7. It illustrates that different electrode pairs have different perceptive fields. A capacitance readout only reflects the touch within its own perceptive field and is not sensitive to touches outside the area. For example, in the first graph, the capacitance readout fluctuates from 15 s to 25 s and keeps constant during other periods. It demonstrates that the capacitance readout is sensitive to the contacts on sub-regions 7 and 8, and is not affected by the contacts on the other sub-regions. Small fluctuations in capacitance readouts are also observed during touch. This is induced by deformations of the manipulator caused by touch (e.g., bending).
Figure 5.5: Four selected calibrated capacitance readouts during the (20,20,20) ml inflation process.
Figure 5.6: Twenty-eight calibrated capacitance readouts during the 2-stage experiment.
Figure 5.7: Four selected calibrated capacitance readouts during the 2-stage experiment.
5.4 Experiment and data acquisition

Experiments are conducted to collect data to verify the feasibility of recognising touch during deformation (e.g., inflation) using the proposed simplified SCAS. Ideally, it requires us to acquire e-skin signals and the location of the contact point (as labels) simultaneously. However, accurate touch locations are hard to pinpoint in dynamic experiments where the e-skin is not fixed on a stationary platform but installed on a manipulator that can move and deform. Therefore, the surface of the manipulator is divided into 18 sub-regions (see Fig. 5.8). A sub-region is touched randomly while recording the e-skin signals. In this way, the index of the sub-region is identified as a low spatial resolution location of the touch point.

![Division of the robot body for touch recognition.](image)

Furthermore, the possibility of extracting deformation information from the coupling e-skin signals also needs to be investigated, as deformation tracking is a critical topic, and sensing devices with multiple functions are desired in soft robotics. For the pneumatic robot platform, the inflation information is usually known, as the volume of air injected into the chamber represents the control input. Therefore, this chapter focuses on deformations caused by user interaction, such as bending induced by touch. To acquire deformation labels during the experiment, 5 reflective visual markers...
are bonded to the sides of the robot. Three OptiTrack Flex 13 cameras are deployed around the robot to capture real-time 3D coordinates of the markers. The coordinates provide a brief description of the deformation and are used as deformation labels in the latter.

![Experiment platform](image)

**Figure 5.9:** Experiment platform.

Fig. 5.9 shows the experimental platform. It includes the pneumatic robot equipped with the 8-electrode capacitive e-skin and 5 reflective visual markers, OptiTrack Flex cameras to capture the ground truth coordinates of visual markers and readout electronics to collect the e-skin signals. The data recording speed for the cameras and readout electronics is set to 30 fps.

The experiment includes 2 stages. In the first stage, the soft manipulator is inflated to a steady inflation state according to a preset $p$. Each element in $p$ can be 0 ml, 10 ml, and 20 ml, thus combining a total of 27 different inflation states. A period of 30 s capacitance data for each inflation process is recorded and is annotated as no contact. In the second stage, we randomly touch a sub-region of the robot and induce different deformations by changing the contact force. The contact and deformation process lasts 30 s. The tracking cameras and readout electronics synchronously record data during this period. We implement the same process for 6 different touch sub-regions (3 on the front and 3 on the back) in the same inflation state.
Smart e-skins and machine learning for soft robot perception

The 2-stage experiment is repeated 27 times, corresponding to 27 different inflation states. Although only 6 of 18 sub-regions are touched in each inflation state, each sub-region is touched at least once during the experiment of the 27 different inflation states. The proposed method is expected to have a generalisation ability that enables it to infer tactile information of unseen samples. Therefore, collecting contact data of all sub-regions in one inflation state is unnecessary.

For touch recognition, 189 groups of data are recorded (27 different inflation states, 6 touch points and 1 no contact data in each state, $27 \times 7 = 189$). Each group of data includes 30-second capacitance signals (i.e., 900 frames at a sampling speed of 30 fps) and the index of the corresponding touch point. The data is exclusively divided into training ($125 \times 30 = 3750$ seconds, $3750 \times 30 = 112500$ frames), validation ($32 \times 30 = 960$ seconds, $960 \times 30 = 28800$ frames) and testing ($32 \times 30 = 960$ seconds, $960 \times 30 = 28800$ frames) sets.

For deformation tracking, a group of data consists of capacitance signals of 30 s and the trajectory of visual markers recorded by cameras. The data without touch and suffering from visual occlusion issues is manually filtered. Eventually, 146 groups of data are acquired, which are exclusively divided into training ($108 \times 30 = 3240$ seconds, $3240 \times 30 = 97200$ frames), validation ($18 \times 30 = 540$ seconds, $540 \times 30 = 16200$ frames) and testing ($20 \times 30 = 600$ seconds, $600 \times 30 = 18000$ frames) sets.

5.5 Touch recognition

5.5.1 Neural network architecture

A multi-layer perceptron (MLP) is employed to achieve touch point classification (see Fig. 5.10). There are 19 different classes in this study, 18 different touch points and one case without touch. The input of the MLP is 28 calibrated capacitance readouts in one frame. The MLP outputs a vector with a size of 19, indicating the class probability. Cross-entropy is selected as the loss function. The MLP has one hidden layer with 128 neurons. The activation function for the hidden layer is ReLU. Dropout ($p = 0.1$) is used to prevent overfitting.
28 capacitance readouts

Multi-layer perceptron

After training, the MLP can achieve 99.88% classification accuracy on the testing set. It demonstrates that the proposed simplified SCAS can estimate coarse touch point location using a simple deep learning model even when the signals are seriously interfered with by the inflation of the manipulator’s body. The confusion map of the classification results is shown in Fig. 5.11. It suggests that only 34 out of 28800 frames of testing samples are misclassified. All misclassifications occur between adjacent sub-regions. For example, 34 touches on sub-region 2 are incorrectly classified as sub-region 3. This is because signals induced by touches on adjacent sub-regions have relatively high similarity, which probably confuses the network.
5.6 Deformation tracking

5.6.1 Neural network architecture

Estimating coordinates of visual markers based on capacitance signals can be treated as a set-to-set issue. To this end, the C2DT developed in the last chapter can be employed. The structure of the C2DT to this application is shown in Fig. 5.12. The cameras and readout electronics are synchronised by an auto-click script, which leads to a slight delay between data recorded by different devices. To alleviate this problem, 10 frames of calibrated capacitance signals are fed into the C2DT at the same time. The position signals ($P$ and $Q$) consist of the location information of the electrode pair to form the capacitor, which can help the C2DT distinguish capacitance readouts generated by different electrode pairs. The squared error between estimation and ground truth is selected as the loss function.
Smart e-skins and machine learning for soft robot perception

This study focuses on deformations caused by user interactions (e.g., touch) rather than inflation (which, in most scenarios, is a known parameter, given that air volume injected into the chamber is a control parameter). However, the signals induced by these deformations are much smaller compared with signals triggered by inflation and touch. This makes it extremely challenging and in some cases altogether unfeasible to directly estimate the coordinates of the visual markers. In order to address this issue, the first frame data in each trajectory are used as prior knowledge, i.e., the capacitance readouts are used as the reference to calibrate the capacitance input, and the coordinates of the markers are used as the source sequence (initial coordinates of markers in Fig. 5.12) of the transformer decoder.

5.6.2 Implementation

The training of the C2DT is implemented in Pytorch. The Adam optimiser is employed to update the learnable parameters to minimise the squared loss between predicted and ground truth coordinates. The initial learning rate is set to 0.001 and decayed by a factor of 1.2 every 15 epochs. 150 epochs of training are run with a batch size of 255 using the training set and the network with the smallest loss on the validation set is saved as the final model. The whole training process takes 2.5 hours on 3 Nvidia Quadro P5000 GPU cards.

5.6.3 Results

After training, the C2DT can achieve $2.905 \pm 2.207$ mm AD error. This demonstrates that, with prior knowledge (the capacitance signals and coordinates of markers in the first frame of each trajectory), the proposed simplified SCAS can be applied to track the deformation using the C2DT even in operative conditions subject to severe interference (e.g. inflation and permittivity variations caused by touch).

The examples of several tracking tests are shown in Fig. 5.13, where the location of the estimated visual markers (red) and the ground truth visual markers (blue) are presented. For all cases shown, the estimated locations of the visual markers are close to the ground truth ones, indicating the high level of accuracy of the deformation tracking (maximum AD error among the 6 examples is 3.754 mm). Fig. 5.14 illustrates the tracking performance of a selected marker over a 30-s period. No significant tracking errors are observed during the whole trajectory, which further demonstrates the remarkable performance of the proposed method.
<table>
<thead>
<tr>
<th>Load Types</th>
<th>Models</th>
<th>FC-64-392</th>
<th>SC64-64-392</th>
<th>NC-64-392</th>
<th>MK64-64-392</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L_o/a (x, y)</td>
<td>1.8429 ± 0.579</td>
<td>1.209 ± 0.711</td>
<td>1.600 ± 0.843</td>
<td>1.423 ± 0.573</td>
</tr>
<tr>
<td></td>
<td>r (x, y)</td>
<td>1.819 ± 0.951</td>
<td>1.364 ± 0.826</td>
<td>1.282 ± 0.837</td>
<td>1.343 ± 0.912</td>
</tr>
<tr>
<td></td>
<td>(x, y, r)</td>
<td>1.819 ± 0.951</td>
<td>1.364 ± 0.826</td>
<td>1.282 ± 0.837</td>
<td>1.343 ± 0.912</td>
</tr>
</tbody>
</table>

**Table 1:** Modified Chamfer distance of different models under different load types.

<table>
<thead>
<tr>
<th>Load Types</th>
<th>Models</th>
<th>L_o/a (x, y)</th>
<th>r (x, y)</th>
<th>(x, y, r)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L_o/a (x, y)</td>
<td>0.0439 ± 0.0949</td>
<td>0.0045 ± 0.0388</td>
<td>0.0049 ± 0.0228</td>
</tr>
<tr>
<td></td>
<td>r (x, y)</td>
<td>0.0160 ± 0.0356</td>
<td>0.0036 ± 0.0333</td>
<td>0.0049 ± 0.0209</td>
</tr>
<tr>
<td></td>
<td>(x, y, r)</td>
<td>0.0160 ± 0.0356</td>
<td>0.0036 ± 0.0333</td>
<td>0.0049 ± 0.0209</td>
</tr>
</tbody>
</table>

**Table 2:** Mis-corresponding rate of different models under different load types.

**Figure 5.12:** The structure of the C2DT to estimate coordinates of visual markers based on capacitance signals.
Figure 5.13: Examples of tracking results.
Figure 5.14: A tracking trajectory of a selected marker.
5.7 Summary

This chapter reported a simplified SCAS that can reduce fabrication, deployment and wiring complexity. The feasibility of the simplified SCAS to perform touch recognition and deformation tracking under environmental interference was validated on a pneumatic robotic platform. Prediction of contact location under a range of inflation conditions is successfully demonstrated at a low level of spatial granularity. In addition, feeding a transformer-based architecture (C2DT) with the information of the first frame in a trajectory as prior knowledge enables us to successfully estimate the deformation of the actuator in any following frames. These achievements are evidence of the potential of our proposed method to be used in challenging applications, such as human-robot interaction, which typically involve various stimuli sources (e.g., deformation induced by actuation and physical contacts).
6.1 Conclusions

The lack of perception systems capable of providing high PGR proprioceptive feedback and tactile information under interference induced by deformation is one of the key challenges that prevents the realisation of the autonomy of soft robots. This thesis aimed to address the issue and conducted a comprehensive study of soft robot perception (including proprioception and exteroception) from the perspectives of coupling field simulation, sensor design and fabrication and learning-based perception algorithms. The contributions of this thesis are concluded as follows.

The literature on simulation, proprioception and tactile sensing in soft robots was reviewed. It introduced the scientific background and the state-of-the-art of the research topic. The challenges that have not been previously addressed and motivate the study in this thesis were discussed.

Existing soft robot simulation approaches mainly concentrate on the actuation of a soft robot, neglecting its sensory systems. The CFS model for a 16-electrode capacitive sensor array deployed on a soft manipulator was developed to seamlessly integrate the sensing and mechanical components. The capacitive sensor array was characterised under various deformations, including bending, two-stage twisting and bending and the compound deformation of twisting and elongation via the CFS model. The results help to understand sensor behaviours better. The case studies for two typical perception tasks, i.e., deformation classification (proprioception) and applied force estimation (exteroception) were implemented on the basis of the annotated data generated using CFS to demonstrate the benefit of the proposed CFS method to the development of learning-based perception techniques. Combined with simple ma-
Smart e-skins and machine learning for soft robot perception

Machine learning frameworks, the capacitance sensor array performed well in both tasks. The models trained with simulation data could be transferred to practical applications using sim-to-real transfer learning approaches, significantly reducing time and costs for data acquisition through physical experiment platforms.

Soft robot closed-loop control relies on real-time high PGR morphological feedback. This thesis developed a high PGR proprioception system empowered by a SCAS that leverages capacitances formed by the combinations of planar boundary electrodes, and an end-to-end neural architecture to translate SCAS signals directly into point clouds. The proposed proprioception system can achieve real-time (30 fps), high PGR (3,900) full-geometry deformation reconstruction with high accuracy (2.322 ± 0.687 mm CD error) under complex deformations including omnidirectional bending, twisting around an arbitrary axis, omnidirectional elongation and their compound deformations. This level of proprioception has not been demonstrated previously and represents a step change over previous attempts. Although the system was demonstrated only on a squared soft manipulator, it does not require prior knowledge. Therefore, it can be extended to different types of soft bodies. Implementing such high PGR, full-geometry proprioception is essential to accurately and comprehensively capture morphological information and go beyond open-loop control for soft robots.

Finally, the thesis performed a preliminary exploration of tactile sensing and/or deformation tracking with severe interference which is inevitable in some practical applications. A simplified SCAS was proposed to achieve coarse touch recognition and deformation tracking with interference while reducing the complexity of fabrication, deployment and wiring. The feasibility of the proposed method was validated on a pneumatic robotic platform. Prediction of contact location under a range of inflation conditions was successfully demonstrated at a low level of spatial granularity. In addition, feeding a transformer-based architecture (C2DT) with the information of the first frame in a trajectory as prior knowledge enables us to successfully estimate the deformation of the actuator in any following frames. The achievement in these two tasks brings evidence of the potential of the proposed method to be adopted in challenging applications that are typical with severe interference (e.g. human-robot interaction).
6.2 Future work

This thesis advanced the field of soft robot perception by developing novel perception methods and demonstrating their superiority in different soft robotic platforms and on various complex deformations. However, several issues of the proposed approaches remain unsolved and need to be further explored.

6.2.1 Automated fabrication

The SCAS fabrication involves manual operation (e.g., liquid metal injection, sealing layer attachment, interface to sensing electronics), leaving room for performance improvement. Although calibration of sensor readouts can, to a certain extent, mitigate this issue, a desirable solution would require automated manufacturing technologies, such as direct writing of liquid metal and 3D printing of soft materials. Furthermore, the thickness of the SCAS is about 1 mm, which is suitable for demonstrating high PGR proprioception in our soft robot testbed (20 × 20 × 200 mm) and other similar proprioception scenarios. However, more advanced fabrication approaches [136] could be adopted to fabricate e-skins with smaller sizes and more complex structures which could be deployed on micro soft robots and extended to other application domains such as skin-interfaced wearable devices.

6.2.2 Multi-modal sensing

Due to the capacitive nature of the SCAS, it is sensitive to geometric and permittivity variations. It is feasible to apply learning-based methods to extract deformation and tactile information from the coupled sensing data as the two different stimuli trigger different signal patterns. However, it is not the optimal solution due to the drawbacks, such as the requirement of prior knowledge during deformation tracking and performance degradation. Multi-modal sensing is preferable as it not only has the potential to address current issues by introducing auxiliary sensing data but also can endow the capability to detect other types of stimuli, such as temperature and humidity.
6.2.3 Sim-to-real transfer learning

This thesis concentrated on the paradigm of supervised learning that requires abundant labelled data for training. A notorious problem is that the acquisition of labelled data is expensive, time-consuming and in some cases even impossible. For instance, point clouds of compression-induced deformations cannot be easily collected through vision-based methods due to inevitable occlusion. The proposed coupling field simulation can generate a large number of high-quality labelled training samples. However, the gap between virtual and physical environments leads to performance deterioration if the network is trained only on the simulation dataset. Sim-to-real transfer methods are considered as the potential solution. The development and application of sim-to-real approaches suitable for soft robot proprioception can significantly increase the value of the simulation data and reduce the cost of real-world data acquisition.
Reference


Smart e-skins and machine learning for soft robot perception


Smart e-skins and machine learning for soft robot perception


Smart e-skins and machine learning for soft robot perception


Smart e-skins and machine learning for soft robot perception


Smart e-skins and machine learning for soft robot perception


Smart e-skins and machine learning for soft robot perception


116
Smart e-skins and machine learning for soft robot perception


Smart e-skins and machine learning for soft robot perception


Smart e-skins and machine learning for soft robot perception


Smart e-skins and machine learning for soft robot perception


Smart e-skins and machine learning for soft robot perception


Smart e-skins and machine learning for soft robot perception


Smart e-skins and machine learning for soft robot perception


